UNIVERSITY OF ROME ‘LA SAPIENZA’
PH.D. IN COGNITIVE PSYCHOLOGY
2001/2004

PH.D. THESIS

Multimodal Investigation on Spatial Attention Mechanisms: A Model of Shared Attention Resources (ShAR)

Valerio Santangelo

ECONA

INDEX

Introduction

The primary objective of this work is to provide an extensive investigation on multimodal mechanisms of attention. Several aspects of this crucial domain will be deeply analyzed by means of behavioural and electrophysiological approaches. Relevant issues in this investigation will be involved with reflexive (exogenous) orienting of attention in case of multimodal or unimodal objects (e.g., are functioning modalities of spatial attention differently triggered by multimodal events or by unimodal events? Which are the neural basis of attentional orienting toward unimodal vs. multimodal events?); selective spatial attention in case of exogenous and endogenous stimuli (e.g., does it exist a difference between reflexive and voluntary orienting toward multimodal events?); crossmodal divided attention (e.g., how does attention manage to allocate resources on spatial information incoming from different sensorial modalities?); influence of multimodal integration on orienting mechanisms (e.g., does audio-visual integration affect the magnitude of cuing effect?); temporal intercourse of audio-visual integration (e.g., does it exists a temporal window in which audio-visual information can be integrated together?); architectural description of attention (e.g., can attention be depicted as: a supramodal system, modality-specific systems, separated-but-linked systems, or what else?). Collected data will be bound together in order to provide a common theoretical framework as a departure point for further research.

This work is organized into six chapters. In Chapter 1 a review of existing models of crossmodal attention is accomplished, from which it clearly emerges that an exhaustive architectural model of multimodal attention mechanisms is still laking. The unimodal approach to the study of attention will be deeply illustrated, reviewing crucial studies concerning auditory attention and visual attention processes. Then on the basis of recent findings in crossmodal research providing evidence in favour of a multimodal-toward-unimodal elaboration of complex information (McDonald et al., 2003; Fort et al., 2002; Vroomen & de Gelder, 2000; Macaluso et al., 2000), a theoretical bridge between unimodal and multimodal research is launched. The aim of this chapter is to trace the state of the art of multimodal attention research in order to provide a basic framework for the empirical investigations reported in the following chapters, and the model of "Shared Attention Resources" (ShAR) proposed in Chapter 6.

In order to contribute to the debate concerning the supramodal versus modality-specific nature of attention, Chapter 2 is based on a crossmodal divided attention task described to asses the real independence of the visual and the auditory attentional systems. Participants were required to make an elevation discrimination task of visual targets preceded by visual cues, auditory cues, or both. They were instructed to pay attention to both visual and auditory cues, as the first or the second one, or both, indicated in the 80% of trials the correct side in which visual targets were delivered, facilitating their elevation discrimination task. The rationale is that if each attention modality-specific system is quite independent from the others, then no difference in participants' performance would exist between a condition wherein the attentional load engages only one modality (single visual or auditory cue condition) and another wherein the attentional load engages two modalities (bimodal cue condition). On the other hand, if the attention system is supramodal, then the splitting of attentional resources over two modalities would dramatically decrease performance in bimodal compared with single cue conditions. Finally, if modality-specific attention system are separated-but-linked, then performance differences would be attenuated with respect to the previous case. In other words, it is reasonable to attend a significant decrease of performance in case of bimodal cues with respect to single cues, but also enough independence to avoid a sharp decrease of performance as though the system was supramodal.

Chapter 3 deals with the attempt to establish whether a temporal window in which audio-visual information can be integrated exists. To this purpose we used again the bimodal condition (Chapter 2), but using fixed intervals of milliseconds in which the auditory cue occurred. More specifically we delivered the second (auditory) cue at different Stimulus Onset Asynchronies (SOA) in order to establish the trend of cuing effect through time. If the information provided by the visual cue and the auditory one can be bound together and used for the spatial discrimination task in a narrow time window, then we should find that the to progressively increasing in the temporal distance between the visual and the auditory cue correspond increasing of RTs, with a consequent worsening in performance. Otherwise, if the process of audio-visual integration can be accomplished without substantial differences through time, then the cuing effect should be stable for each level of SOA.

Chapter 4 is concerned with the attempt to verify whether spatial attention triggered by multimodal events acts differently in comparison to unimodal (either visual or auditory) triggers. The literature concerning reflexive mechanisms of spatial attention has shown that reaction time (RT) decreases when a spatial cue (e.g., a flash of light or a burst of sound), being unpredictive with regard to the location of a forthcoming target, is delivered on the target side (i.e. a validly cued trial) as compared to when it occurs on the non-target side (i.e. an invalidly cued trial; Posner, 1980). This 'cuing effect' (the difference between validly and invalidly cued trials) was found in both unimodal and crossmodal settings. However, it has been obtained typically for single cue situations, while what happens when two cues from different modalities are presented at the same time was only seldom investigated. This issue appears to have relevant implication from an ecological point of view, as in everyday life multimodal stimuli from different modalities occur very frequently, while unimodal cues are rare. In order to investigate the attentional orienting mechanisms in case of multimodal or unimodal events, we compared a condition wherein a visual target was preceded by both visual and auditory exogenous cues delivered together at the same side (bimodal cue), with conditions wherein a visual target was preceded by either a visual or an auditory cue (single cue) in a spatial attention discrimination task. If orienting attentional mechanisms take advantage when the trigger is a multimodal event, RTs will be shortened in case of double cue presentation; otherwise, no difference will be found between single and double cue conditions.

A similar rationale underlies the study reported in Chapter 5, wherein it was investigated the influence of multimodal integration on exogenous orienting and the neural basis of reflexive orienting toward unimodal and multimodal (audio-visual) events. It may be hypothesized that the speeding up of responses due to exogenous orienting effects of bimodal cues exceeds the sum of single cues. Namely, the amplitude of event related potentials (ERPs) elicited by a combined audiovisual stimulus is larger than the sum of a single auditory and visual stimulus, which may reflect crossmodal integration and could affect the magnitude of orienting effects. Behavioral data, however, revealed no increased orienting effect (Chapter 4), although this null effect may be due to a failure of crossmodal integration. To control for this, we measured both ERPs elicited by task-irrelevant bimodal, crossmodal, and unimodal cues, and determined their exogenous orienting effects in a visual discrimination task. ERP components elicited by bimodal cues were indeed larger than the sum of a single auditory and visual stimulus, but no enlarged orienting effect was found. Thus, crossmodal integration seems to fulfils no special role for exogenous orienting.

Data collected and reported in chapters 2 to 5 are integrated in a theoretical model proposed in Chapter 6. This model, based on *shared attention resources* (ShAR), suggests a hierarchical elaboration of sensorial information, which imply different levels of processing, from multimodal to unimodal analysis. As it will be pointed out, this model is able to take into account many data deriving from literature. Relevant implications will be discussed in relation to the previously disserted hypotheses of crossmodal attention (supramodal, modality-specific, separated-but-linked) and empirical ways to test the model's predictions is suggested, substantially contributing to the crucial domain of selective spatial attention.

CHAPTER 1 - Crossmodal Attention: Shifting the Perspective from Unimodal to Multimodal Approach

In the last decade development of crossmodal research was so rapid and impressive to determine a revolutionary shifting of perspective from unimodal (auditory, visual, tactile, etc…) to multimodal approach. Nowadays, it appears no more possible to understand how attention works leaving out of consideration mutual connections and relationships between the different sensorial modalities.

This chapter is primarily focused on the attempt to build a theoretical bridge between unimodal and multimodal approach. This would be useful in order to catch the whole complexity of this crucial domain, and provide a basic framework for further investigations. Findings derived from unimodal investigations should be integrated with crossmodal results, in order to obtain a coherent and structured model of the attentional domain.

Traditional Approach to the Study of Attention

Auditory Attention Research

In the following sections, we are going to review studies specifically focused on auditory attention. Our review will take into account the two main lines of research existing into the literature. The first one is involved with the attentional modulation of hemispheric asymmetries reported by dichotic listening-based studies. The second line tries to determine the spatial distribution of auditory attention, and the features of sound (e.g., frequency, location, etc.) able to facilitate (or to inhibit) the allocation of auditory attention resources.

Dichotic Listening and Auditory Asymmetries of Attention

In the history of experimental psychology, dichotic listening represents the method of research which produced the deepest impact on the study of attentional processes, providing data for the first well-known models and theories of attention mechanisms (Broadbent, 1957; Treisman, 1960; Deutsch & Deutsch, 1963; Norman, 1968). Dichotic listening is an experimental paradigm based on a simultaneous presentation (to both right and left ears) of auditory stimuli. When the stimulus material is linguistic (digit names, consonant-vowel syllables, consonant-vowel-consonant words, and so on), items presented to the right ear are more likely reported than items presented to the left ear. The right-ear advantage (REA) is normally interpreted as representative of hemispheric functional asymmetries and, more specifically, a left-hemispheric advantage for verbal and linguistic tasks. According to Kimura's (1967) structural model, the REA reflects the superior conductivity of the crossed auditory pathways as well as a central occlusion mechanism that inhibits the ipsilateral pathways during dichotic competition. Information from the right ear reaches the left hemisphere directly, whereas information from the left ear reaches the left hemisphere indirectly via the right hemisphere and corpus callosum (Sparks, Goodglass, & Nickel, 1970). However, Kimura's theory fails to account for several subsequent findings (for a review of such data see Henry, 1979; 1983). As numerous studies have shown that binaural rivalry is not a necessary condition for obtaining a REA, an occlusion of ipsilateral signals need not to be invoked as an explanatory principle. Even though the crossed auditory pathways are better conductors than the uncrossed pathways, pathway differences alone cannot account for REA. Other findings showed on the inadequacy of structural explanation of REA: ear differences tend to be unstable from one testing session to the next (Pizzamiglio, De Pascalis, & Vignati, 1974); they may be influenced by the participant's familiarity with the stimulus material (Johnson, 1977), and, above all, by *attentional factors* (Bloch & Hellige, 1989; Hiscock, Inch, & Kinsbourne, 1999a, 1999b; Hiscock & Stewart, 1984; Hugdahl & Andersson, 1986; Mondor & Bryden, 1991, 1992a, 1992b; Murray, Allard, & Bryden, 1988; Wood, Hiscock, & Widrig, 2000). As argued by Bryden (1978):

> If one is concerned primarily with the investigation of cerebral asymmetries in the normal subjects, then we must take every precaution to minimize the subject's control over the situation. In particular, the subject cannot be left free to deploy attention as he chooses, but some formal control must be employed. Otherwise, we run the danger of measuring not cerebral lateralization, but how the subject responds to the situation" (p. 143).

The shifting from *divided-attention procedure* to *forced-attention procedure*, advocated by Bryden (1978), proved to be an optimal way to verify the contribution of attentional processes on perceptual asymmetries. Whereas in divided-attention procedure listeners were instructed to distribute their attention across both ears (without the possibility for researchers to find out any strategies or to check attentional mechanisms involved in dichotic asymmetries), in forced-attention procedure listeners were required to pay attention and to report only from one ear at a time (block of trials).

Hugdahl and Andersson (1986) reported a study in which each listener participated in a non-forced (divided attention) condition, in a forced-right condition, and in forced-left condition. The stimuli were six stop-consonants paired with the vowel *a*. The result showed a significant REA in all groups of listeners during the non-forced condition. During the forced-right condition, significantly more correct recalls were obtained form the right compared to the left ear. During the forced-left condition, significantly more correct recalls were obtained form left the compared to the right ear. Finally, comparing correct recalls from the unattended right ear (during the forced-left condition) with the unattended left ear (during the forced-right condition) revealed a significant REA in most of listeners. Thus, the typical REA effect for identification of verbal stimuli can be largely suppressed be means of attentional processes. A similar conclusion was reported by many other studies (Bloch & Hellige, 1989; Hiscock et al., 1999a, 1999b; Hiscock & Stewart, 1984; Mondor & Bryden, 1991, 1992a, 1992b).

More recently, Hiscock et al. (1999a, 1999b) managed to abolish REA in localization but not in detection tasks. Participants listened for specified targets and reported the ear of entry. The material in Hiscock et al. (1999a) consisted of pairs of consonant-vowel syllables (Experiment 1) and pairs of rhyming consonat-vowel-consonat words (Experiment 2); whereas in Hiscock et al. (1999b) consisted of lists of digit words (Experiment 1) and lists of words (Experiment 2). Regardless of this variety of experimental stimuli, results showed an alteration of ear asymmetry following shifts of attention for localizing but not for detecting targets (except for words, wherein attention shifts altered both detection asymmetry and localization asymmetry, but the effect on detection seemed to reflect differential retrieval from short-term memory rather than differential perception). According to the authors, the dissociation between detection and localization indicates that volitional shifts of attention influence late (response selection) process rather than early (stimulus identification) processes. They proposed a model of selective-listening that involves two stages of processing: the first one, automatic and characterized by a relative fixed asymmetry favouring the right ear for linguistic stimuli, and the second one, controlled and characterized by volitional attention shifts.

Spatial Distribution and Selection of Auditory Information

Which is the exact nature of the spatial representation subserving auditory attention system? Several authors have tried to determine whether auditory attention shifts were subserving from an analogical or non-analogical (discrete) system. For instance, Rhodes

(1987) studied the spatial features of auditory attention in a series of experiments in which participants had to discriminate between several (9, 8, or 5) loudspeakers (located either around the circumference of a semicircle in front of, or in a complete circle surrounding the participant), which of them was played a tone, by naming its location as quickly and accurately as possible. The rationale was that if auditory spatial information is represented analogically, then shifts of auditory attention should be constrained by the spatial structure of the representation. More specifically, the time taken to shift auditory attention should increase with the distance moved, given that the distance of attentional shift was defined as the physical separation between tones on successive trials. As predicted, the time taken to localize a sound increased linearly as a function of the angular distance from the listener's point of focal attention, up to a distance of 90°, but it is worth noting that there was little change in RTs as distance increased above these asymptotic distances. For distances greater than 90°, however, this is no longer the case. Besides, the results showed that the rate at which attention was shifted is a function of the task difficulty (Experiment 1 vs. Experiment 2) and the memory load (Experiment 3). These results provided the first evidence that auditory attention exhibits spatial constraints, even though it is not possible to rule out the presence of priming effect in these data (see section 'Early studies'). The author interpreted this data as demonstrating that "a shift [in the source of sound] of up to 90° in either direction from the attended location would be followed by an analogical shift of attention, whereas for larger movements attention would just pop-up in the new place" (p. 13).

This model is not consistent with Mondor and Zatorre's (1995) findings. They argue that evidence in favour of analogical shifts reported by Rhodes (1987) may have been affected by two methodological problems. First, listeners were required to identify verbally the loudspeaker from which a target had been presented. Because loudspeakers were numbered consecutively, listeners may have developed an ordinal arrangement of loudspeakers name and counted their way through this arrangement to derive the correct response: this strategy (and not the spatial separation between consecutive targets) could be responsible for the linear increase of RTs. Second, the correction procedure used to eliminate the effect of azimuthal position from the assessment of attentional shift distance appears to underestimate the azimuthal effect contribution. Critically, if the underestimation is much more severe for peripheral locations than for central location, and peripheral locations are more often involved in long attentional shifts than are central locations, then the correction procedure itself can contribute to an apparent increase in RT with attentional shift distance. Mondor and Zatorre (1995) used an identification paradigm able to eliminate these possible biases. Indeed, identification response does not allow listeners to form an ordinal arrangement of response that could be related to stimulus location, and it has not been shown to be related in any way to azimuthal location. Their results showed that performance improved as time available to shift attention to a cued spatial position increased; accurate spatial cues facilitated performance more that inaccurate cues; performance declined as the distance of an unexpected target from a cued spatial location increased; and, above all, performance was virtually identical for shifts of attention ranging from 0° and 180°. These findings provided evidence that auditory attention may be allocated to a specific location in response to an auditory spatial cue, and that the time required to shift attention does not depend on the distance of the shift, as required by an analogical model. Authors concluded that "auditory attention appears to be distributed as a gradient, with the density of resources declining gradually with the distance from an attentional focal point" (p. 407).

The gradient model is also supported by Mondor and Bregman's (1994) observations. They reported a series of experiments conducted to determine whether judgments about targets may be substantially influenced by the spatial distribution of auditory attention. On each trial, a frequency cue was presented and was followed by a target tone. The cue indicated that most likely frequency of the forthcoming target about which participants were required to make a duration judgment. The rationale was that if participants were able to allocate attention to the cued frequency region, then judgments of any features of a tone of the cued frequency should be facilitated with respect to tones of different frequency. As predicted, results showed that duration judgments were more quickly and accurate when the cue provided accurate frequency information than when it did not. Moreover, performance generally declined as the frequency separation between cue and target increased. The latter result appears to confirm the gradient model hypotheses (see also Rorden & Driver, 2001), wherein attention can be distributed in a graded fashion, with maximal processing at the focus of attention, which gradually falls off with increasing distance from this focus. Indeed, any model that assumes sharp transitions from attended to unattended regions is incapable of accounting for this result.

The relevance of the gradient model hypothesis is somehow undetermined by the results reported by Ferlazzo, Couyoumdjian, Padovani, and Olivetti Belardinelli (2002, Experiment 5) about the head-centered meridian effect. They found slower RTs when auditory cues and auditory targets were at the opposite side of the head-centered meridian than when they were separated by the visual meridian or not separated by any meridian. In this case the decrease of performance could not be explained in terms of gradual decline of attentional resources between cue and target locations, as the speakers were placed at the same distance between them. Thus, the gradient model cannot account for decrease of performance in the case of crossing of head-centered meridian. More recently, the validity of the head-centered meridian effect was extended on schizophrenics and blind subjects (Olivetti Belardinelli & Santangelo, in press). However, it is worth noting that the head-centered meridian might be the unique case in which the gradient model failed to account for the spatial distribution of auditory information, likely for a specific role played by such a meridian in the perception and organization of space (Pillow & Rubin, 2002; Kuyk & Niculescu, 2001).

Other fascinating issues concerning the auditory attention domain are: is it possible to allocate auditory attention in a specific region of space, and, if this is the case, how is it accomplished, or in other words, which are the sound features able to influence the selection of auditory information? Mondor and Bryden (1991, 1992a, 1992b) demonstrated that phoneme identification could be substantially influenced by the allocation of auditory attention. In their experiments, participants heard two verbal stimuli simultaneously delivered at both ears. Before each trial, a cue tone was delivered at either right or left ear. Participants were instructed to identify the syllable presented to the cued ear. In all experiments, identification performance for the cued ear improved as a function of the time between the spatial cue and the syllable. Several control experiments confirmed that this effect was due to an allocation of attention to the cued ear as superior performance was obtained when participants were provided with an accurate rather than an inaccurate spatial cue. These investigations provided strong evidence that the instruction to focus attention on spatial cues can facilitate or inhibit auditory perception. Other pieces of evidence were found again by Mondor and colleagues in auditory target identification tasks with advance cues regarding probable target location (Mondor & Zatorre, 1995) or target frequency (Mondor & Bregman, 1994). In both studies (see previous section for their extensive review), the speed and accuracy of target identification was improved by accurate advance information (valid trials) relative to performance in trials in which the cue provided inaccurate information (invalid trials). These effects established that auditory attention can facilitate stimulus identification through selection by either spatial location or frequency region. All these evidences taken together provided a clear demonstration that it is possible to allocate auditory attention on specific region of space (Mondor & Bryden 1991, 1992a, 1992b; see also Quinlan & Bailey, 1995) and that at least two sound features are able two influenced selection of information in auditory processes: frequency and location (Mondor & Bregman, 1994; Mondor & Zatorre, 1995).

Frequency and location are the most extensively studied features relative to selection of auditory information. Several works investigated how they facilitate or inhibit target identification process, and which is the relation between them (Woods, Alain, Diaz, Rhodes, & Ogawa, 2001; Mondor, Zatorre, & Terrio, 1998; Jones, Jagacinski, Yee, Floyd, & Klapp, 1995; Woods, Alho, & Algazi, 1994). For instance, Woods et al. (1994) reported some evidence consistent with independent selection dimensions for frequency and location. In an event-related potentials study they found that the processing of frequency and location information during a selective listening task was associated with different scalp distributions, at least within the first 80-110 ms after stimulus onset. As argued by Mondor et al. (1998), although these results suggest the possibility that selection may be guided either by location or frequency channels, they do not allow to determinate if the selection occurs independently for either location or frequency dimension. Alternatively, the conjunction of frequency and location could occur before the independent analysis of each of these features is complete, and because of this, selectively attending to a location may entail also directing attention to frequency information and vice versa (Woods et al., 1994). Deutsch (1974), with her “scale illusion”, provided some evidence of independent frequency and location analysis. The scale illusion was created by presenting a succession of pure tones dichotically. In one ear, an 800 Hz tone was presented three times, followed by two presentations of a 400 Hz tone. In the other ear, a 400 Hz tone was presented three times (simultaneous with the 800 Hz tones), followed by two presentations of an 800 Hz tone (simultaneous with the 400 Hz tones). If the rate of presentation was rapid enough, listeners reported hearing only one of the two frequencies presented at each instant: frequency heard and the location from which that frequency apparently originated seemed to be governed by separate process. Specifically, the frequency heard appeared to be that presented to the listener’s preferred ear (defined as the right ear for right-handed listeners and the left ear for left-handed listeners), whereas the location from which the sound seemed to originate depended on the location of the higher frequency tone (i.e., whether it was consciously perceived or not). These data may be interpreted as indicating that frequency and location information are processed separately. It stands to reason that if this is the case, then it should be possible not only to allocate attention to each of these dimensions separately but also to guide selection independently by each of these channels.

It is possible to find some indirect evidence of this assumption in recent works wherein resulted a significant effect of frequency separation in the listener’s ability to attend competitive tone sequences (Woods et al., 2001) or competitive pattern structures (e.g.

frequency and time relation; Jones et al., 1995). These results are somehow undetermined by the evidence reported by Mondor et al. (1998) in a series of experiments wherein participants had to categorize tones by location and frequency. Crucially, they found that the tone-classification by one dimension (frequency or location) was interfered when a variation on a second, irrelevant dimension was introduced. This effect was obtained both for location classification with frequency variation and for frequency classification with location variation. Therefore, it does not seem possible for listeners attending independently to any of these features, but attending to one of them necessarily entails attending to the other. The authors suggested that frequency, location, loudness, and timbre are all integral perceptual attributes, in the sense that listeners are unable to attend to one of these dimensions without attending to the others. It is noteworthy that selection appears to be partly constrained by results of an 'early perceptual organization', as the time required to detect a pre-specified target increases as its similarity both in frequency and location to the other items in a tone sequence (distractor) increases (Mondor et al., 1998).

Taken together, the evidence of a preattentive perceptual grouping stage and the effect of target-distractor similarity may provide an insight into the intriguing issue of selection of auditory information. For instance, Alain and Woods (1993) reported a study in which participants were required to detect occasional target tones of longer duration (Experiment 1) or increased loudness (Experiment 2) in an evenly spaced condition, in which attended and distractor frequencies differed by 6 and 12 semitones, or in a clustered condition, in which the distractor frequencies were grouped (the middle tone was paired with a distractor that was either 1 semitone lower or 1 semitone higher in frequency) and the attended tones differed from the distractors by 6 and 7 semitones. In both experiments, results showed that clustering of the unattended frequencies improved the detectability of targets and speeded target RTs. According to the authors, "distractor clustering enhanced performance mainly by decreasing interference of irrelevant frequencies" (p. 513). We argue that Alain and Woods's (1993) results are consistent with the existence of an early stage of perceptual organization which may be accomplished by a preattentive grouping mechanism. The rationale is that if clustering of unattended frequencies occurs at a preattentive level, then it is clear the reduction of attentional load with the following improvement of target detection in both accuracy and RTs. In other word, a preattentive grouping stage may operate as an integrator of acoustical features highly compatible one with each other (Mondor & Terrio, 1998; Woods at al., 2001) in order to constitute a solid base for the subsequent allocation of attentional resources and selection of auditory information. Within this framework, the *decreasing interference of* *irrelevant frequencies* postulated by Alain and Woods's (1993) to explain the clustering effect over unattended frequencies, may be only the consequence of a previous (preattentive) analysis and integration operated by the grouping mechanism.

Integration of location and frequency does not necessarily imply that these features might not be independently registered at an early stage of processing. Woods et al. (1994) reported different ERP distributions for attending to location versus frequency, consistent with evidence that neurons within distinct cortical fields are sensitive to specific acoustic features. Although these data suggest that different neural mechanisms are involved in different stimulus dimensions, Woods et al. (1994) also showed that features appear to be conjoined rapidly (within 130 ms). More recently, in a PET study, Zatorre, Mondor, and Evans (1999) tested the hypothesis that similar neural systems are involved in attending to spectral and to spatial features of sounds. In each of four conditions participants heard tones varying randomly in frequency and location and responded to either the low- or the high-frequency stimuli, ignoring location, or to stimuli on the left or right, ignoring frequency. In comparison to a silent baseline, a CBF increase was observed in auditory cortex bilaterally and in the right superior parietal, right dorsolateral frontal, and right premotor regions, with no modulation as a function of attentional condition. The data imply that auditory attention engages a network of right-hemisphere cortical regions for both spatial location and tonal frequency and support a model whereby auditory attention operates at a level wherein separate features have been integrated into a unitary representation.

Visual Attention Research

In the previous chapter, we have seen that attention can improve performance in auditory tasks. Specifically, to focus attentional resources over a spatial location, or a sound frequency improves RTs in detection and discrimination of a specific target following a similar (for location or frequency) precue, whereas RTs are impaired as much as cue-target dissimilarity increases. We have also seen the strength of selective attention in balancing the normal REA by means of a simple instruction to attend the left ear. However, auditory attention research is faraway to understand which are the real mechanisms responsible for performance improvement, yet. Visual research also found that focused attention increase performance in several tasks, such as visual search and detection of luminance and vernier targets (Baltz & Hock, 1997; Carrasco & Yeshurun, 1998; Morgan, Ward, & Castet 1998; Posner, 1980; Yeshurun & Carrasco, 1998), and today's research is just oriented towards the discovery of involved mechanisms. More specifically, researchers are wondering if visual performance improvement is due to a *reduction in internal or external noise* (Morgan et al., 1998), a *change in decisional criteria* (Kinchla, 1992), or *signal enhancement* (Lu & Dosher, 1998; Posner, 1980).

Enhancement of spatial resolution

Recently, Carrasco and colleagues reported new evidences in favour of a signal enhancement model of attention (Carrasco, Penpeci-Talgar, & Eckstein, 2000; Yeshurun & Carrasco, 2000, 1999, 1998). For instance, Yeshurun and Carrasco (1999) and Carrasco et al. (2000) reported a series of experiments in which they observed a reduction of the eccentricity effect by using a typical cuing paradigm (Posner, 1980). In the eccentricity effect, observers' performance is slower and less accurate as target eccentricity increases. However, this effect diminished when the target location was precued, suggesting that to pay attention to the location of a stimulus could improve its sensory representation. Given that in all these experiments the target was always presented *alone* in a large number of different locations, it is possible to rule out the contribution of noise reduction in decreasing of the eccentricity effect. Thus, the sensory representation improvement would be due to the enhancing of spatial resolution at the cue location. Attention may enhance spatial resolution by increasing the sensitivity of small, foveal receptive fields which are responsible for detection of small details (Baltz & Hock, 1997). Other evidence for a signal enhancement model of attention comes from the counterintuitive impairment of performance that Yeshurun and Carrasco (1998) found at central retinal locations. In a texture segregation task, indeed, they observed an improvement of performance at the peripheral locations where the spatial resolution was low, and an impairment of performance at the central location where the spatial resolution was high (see also Yeshurun & Carrasco 2000). This effect could result from an attentional mechanisms that enhances resolution by effectively decreasing the average of cellular filters at the attended location: given that performance is worse at the fovea because its spatial filters are too small and have already too high resolution for the scale of texture, further increasing resolution at foveal locations led to a more pronounced drop in performance. The same mechanism manages to improve performance at regions where spatial filters are too large and have too low resolution for the texture scale.

Physiological studies support the hypothesis that attention can enhance spatial resolution by reducing the size of spatial filters at the attended area. When attended and unattended stimuli are both within a cell's receptive field, the neuronal response is primarily determined by the attended stimuli; responses to the unattended stimuli are attenuated, as if the cell's receptive field shrinks around the attended stimulus (Desimone and Duncan, 1995; Luck, Chelazzi, Hillyard, & Desimone, 1997). These authors proposed that attentional modulation of sensory processing is accomplished in two stages: first, top-down signals bias activity in favour of the neurons representing the relevant location, and second, these favoured neurons compete with other neurons, ultimately suppressing their response. This competition may result from mutual inhibition between cells or between the inputs to the cells, and its outcome could effectively reduce the cell's receptive field, allowing finer spatial resolution. Alternatively, enhanced resolution at the attended location could result from increased sensitivity of the neurons with the smallest receptive fields at the attended area (Baltz & Hock, 1997; Yeshurun & Carrasco, 1999), which in turn may inhibit neurons with larger receptive fields there.

Temporal performance fields

As illustrated in the previous section, discriminability decreases with eccentricity in many visual tasks, but covert attention has the inverse effect. Covertly allocating attention to a location enhances discriminability, and the magnitude of this effect increases with eccentricity (Carrasco, Williams, & Yeshurun, 2002; Carrasco & Yeshurun, 1998; Yeshurun & Carrasco, 1998, 1999, 2000). These findings suggest that covert attention provides the most help at locations where it is most needed. However, this does not appear to be a general principle. Discriminability also varies at different locations at a fixed eccentricity, but covert attention affects all isoeccentric locations to a similar degree (Cameron, Tai, & Carrasco, 2002; Carrasco, Talgar, & Cameron, 2001; Talgar & Carrasco, 2002). The beneficial effects of covert attention are not limited to merely enhancing discriminability. Covert attention also speeds information accrual (Carrasco & McElree, 2001). Consequently, to fully characterize how attention affects information processing across the visual field, it is essential to examine both spatial and *temporal* dimensions of performance.

Carrasco, Giordano and McElree (2004) first documented the existence of a "temporal performance field" (determined by the fact that discriminability is not homogeneous at isoeccentric locations) by showing that the rate of information accrual varies for different locations at a fixed eccentricity. More specifically, they showed that covertly attending to different isoeccentric locations speeded information accrual in a manner that effectively eliminated inherent differences in visual information uptake among the locations. The authors used the speed accuracy tradeoff procedure, which allows the construction of a full time-course function describing how discrimination varies with processing time. Time-course functions are derived by requiring observers to respond immediately to a response signal tone presented at one of several times after the onset of the display; by sampling a suitable range of times, the full time course of processing can be measured, which includes periods when discrimination is at chance levels, periods during which discrimination performance increases, and periods when discrimination has reached its maximum or asymptotic level. With this procedure, they derived conjoint measures of how isoeccentric locations and precueing targets location affect speed and accuracy in a search task. These findings suggest that attention has the same compensatory effect on temporal dynamics at a fixed eccentricity as it does on discriminability across different eccentricities.

Thus, the study of temporal dynamics appears to be a relevant issue for attentional investigations. Seldomly, temporal differences and timing processing are taken into account by researchers, reducing the ecological validity of most of the results obtained in the attentional domain.

Recent Advances in Crossmodal Research

In the following sections we are going to closer examine studies specifically involved with multisensory integration processes. This will be useful in order to integrate 'unimodal' results in a theoretical background which takes into account mutual interactions among the different sensory modalities and attentional mechanisms involved in the information processing coordination.

Crossmodal Integration: a Rule, not an Exception

In recent years, development of crossmodal approach obtained a lot of relevant and fascinating results, showing that crossmodal analysis is likely "the rule rather than the exception'' (Driver & Spence, 2000; p. R731) in the information processing accomplished by our cognitive system. For instance, McDonald, Teder-Sälejärvi, Di Russo and Hillyard (2003) reported that an involuntary shift of attention to a sound preceding a visual target first modulates visual event-related potentials (ERPs) in a multimodal brain region and only subsequently in the extrastriate (modality-specific) visual cortex. Likewise, Fort, Delpuech, Pernier and Giard (2002) found early crossmodal neural activities during the recognition of objects characterized by non-redundant auditory and visual features, demonstrating that specific unimodal (auditory and visual) analysis was preceded by a crossmodal one. Similarly, Vroomen and de Gelder (2000) reported that judgements for one modality are influenced by a second modality, even though the latter one provide no relevant information for the task, as the cognitive system would try to understand whether stimuli should be integrated between them or not.

Other pieces of evidence derive from the McGurk effect (McGurk & McDonald, 1976; Saldaña & Rosenblum, 1993) wherein to watch lip-movements can alter the auditory perception of a phonema (as well as non-speech sound) and from the ventriloquism effect (Bertelson, 1999; Spence & Driver, 2000) in which the apparent location of a speech or non-speech sound is altered by the synchronized presentation of a transient visual event. In both cases, integration derived from more senses simultaneously, appears to be dominant with respect to specific unimodal analysis.

We are in front of a basic change in the perspective of attention research, which appears oriented to catch all the complexity of real-life situations where we constantly interact with multimodal objects, providing at the same time different modality-specific information. To deeply understand how attention manages this complex pattern of information appears to be the intriguing task of the immediate future, while to look at the attentional processes leaving out of consideration multisensory connections appears to be just a prosecution of a reductionist approach.

Unimodal Analysis: just a Subsequent Step

The activation of unimodal areas seems to follow the activation of multimodal ones. Macaluso, Frith and Driver (2000) reported an fMRI study wherein a spatially congruent tactile cue following to a visual target enhance neural activity in unimodal visual cortex (lingual gyrus), even though the lingual gyrus showed no response to single tactile stimulation. Similar evidence was found in an ERP study by Kennett, Eimer, Spence and Driver (2000) and by Calvert, Bullmore, Brammer, Campbell, Williams, McGuire, Woodruff, Iversen and David (1997) in a non-spatial study wherein visual lip-movements were able to activate primary auditory cortex.

Taken together such evidence showed that:

> Brain areas traditionally considered as ‘unimodal’, may only be so in terms of their afferent projections. Back projections from multimodal convergence areas could result in response to the primary modality in ‘unimodal’ brain areas being modulated by stimulation in a second modality […] subjective experience within one modality can be dramatically affected by stimulation within another […] again suggesting a feedback influence from multimodal levels of representation (Driver & Spence, 2000; p. R734).

It is worth noting that to consider the importance of crossmodal approach does not mean to deny the relevance of unimodal investigation, but just to point out that unimodal findings should be integrated with multimodal ones, in order to provide a coherent and, likely, hierarchical (multimodal towards unimodal analysis) model of attention system. Thus, unimodal research carry on to maintain all its intriguing interest also advocating a crossmodal point of view.

Assessing the Integration between Sensorial Modalities

The following sections are specifically focused on studies that tried to underline how it is possible to allocate attentional resources on complex pattern of data, which involve at the same time auditory, visual, and tactile spatial information.

Early Studies

Typically, target detectability increases, as far as RTs are concerned, when a spatial cue (e.g., a flash of light or a burst of sound) is delivered on the target side (valid trial), and decreases in the opposite case (invalid trial; Posner, Davidson, & Nissen, 1976). However, some studies initially failed to show any 'cuing effect' (defined as performance difference between valid and invalid trials) for auditory targets. Michael Posner (1978), summarizing results of numerous cuing studies in vision, audition and touch, reported that, although reliable spatial cuing effects were obtained for visual and tactile targets, no such effects were found for auditory targets preceded by informative central cues when a simple detection response was required. Scharf, Quigley, Aoki, Peachey, and Reeves (1987) also reported similar null effects from a preliminary study: they found no facilitation for detection of acoustic targets coming from expected versus unexpected directions. Butchel and Butter (1988) found that both informative auditory and visual peripheral cues had no effects on detection latency of auditory targets, even though visual detection was influenced by cues in either modality. In quite similar studies (except for the use of uninformative cues), Klein, Brennan, and Gilani (1987) found that auditory detection was uninfluenced by visual cues that nevertheless affected visual detection.

Why should the cuing paradigm, so successful in visual attention studies, be so ineffective in measuring auditory attention? One possible explanation is the use of simple detection tasks, which are inherently spatial tasks. Indeed, it must be consider an important difference between vision and audition: the former is intrinsically spatial, whereas the latter (at least at early stages) is not. In vision, as early as the peripheral receptor level, stimuli from different locations are represented separately, that is, spatiotopically. In audition, by contrast, information from different loci is initially represented tonotopically, regardless of the spatial position of the input source. To discriminate auditory stimuli by means of their source location is a computationally complex operation, relying on temporal phase and intensity differences between the ears, plus contributions from head movements. Such a complexity in sound localization is primarily the result of differences in physic features of light and sound. Whereas vision is spatial even at the peripheral receptor level, the initial stages of hearing are intrinsically non-spatial. This implies that likely is more appropriate to use measures based on auditory frequency instead of source spatial location in order to study this topic (Spence & Driver 1994; Ward, 1994). In fact, as we will illustrate in the following sections, frequency discrimination tasks produced relevant effects on auditory covert orienting.

Other studies found some evidence of auditory cuing effect (Bedard, Massioui, Pillon, & Nandrino, 1993; Farah, Wong, Monheit, & Morrow, 1989; Rhodes, 1987; Ward, 1994) but it remains possible no attentional interpretation. Indeed, none of these studies assessed the spatial relationship between cues and targets. Ward (1994) reported two experiments in which visual or auditory targets were preceded by both visual or auditory cues. Participants had to discriminate the target location by pressing a left button for left side targets and a right button for right side ones, in both cases with the correspondent hand. Ward found that visual cues affected response time to localize both visual and auditory targets, while auditory cues affected only the time to localize auditory targets. However, the experimental procedure used by Ward (1994) does not allow to rule out the existence of non-attentional disturbing factors: 1) facilitation produced by lateralized response left-hand left-side, right-hand right-side, and, above all, 2) targets delivered at the same or overlapped location of cues. Thus, a *priming effect* could be actual responsible for auditory cuing facilitation (Spence & Driver, 1994), in this as in the other above-mentioned works.

Orthogonal Cuing Method

How is possible to study crossmodal links between audition, vision and touch avoiding that any methodological bias spoils the experimental procedure? Spence and Driver (1994, 1997a, 1997b, 1996; see also Spence, Nicholls, Gillespie, & Driver 1998) found a simple but efficient way to assure the correct examination of the experimental hypothesis.

Their central idea was to test participants in a task requiring speeded discriminations concerning the elevation along a vertical axis (up versus down) of each target, which can be visual, auditory or tactile; while the presentation of the cue (visual, auditory or tactile as well) remain associated to an horizontal axis (left versus right). The main point was that "since the direction of the required up/down judgment is orthogonal to the left/right direction of the non-predictive cue, the lateral position of the cue cannot bias the decision". Indeed in this picture participants have to take an "up/down decision regardless of the side on which the target appears, and regardless of whether this is the same as or different from the side of the immediately preceding cue event (which is equiprobable)" (Driver & Spence, 1998a; p.255).

In all experiments, fixation must be maintained straight ahead upon a central fixation point, and this was generally monitored with infra-red trackers. In studies concerning audio-visual links, at the beginning of each trial an auditory or visual cue, equally likely to be on the left or right, was presented for about 100 ms. After a further delay (depending on SOA, the Stimulus Onset Asynchrony), the auditory or the visual target was presented. Participants were required to press the farthest button with respect to them for an auditory or visual target from either of the upper target positions (regardless of side) and the nearest button for a target from either of the lower positions, responding as rapidly and accurately as possible, generally with the index finger of either hand.

In studies involving tactile cues and visual or auditory targets, participants gripped a small cube with each hand. Each cube contained two vibrotactile stimulators, one at the index finger-pad, and one at the pad of the thumb. Each trial began with a short vibrotactile event at both stimulators of either hand. Shortly afterwards, a target light or a target loudspeaker became briefly active, and participants made a speeded discrimination whether the light or the sound came from above or below (by lifting the toes of their right foot for upper targets, and their right heel for lower targets), regardless of its side. Obviously, in studies with visual or auditory cues and tactile targets, the roles of the modality are reversed. In some case, such schema was slightly modified: participants were required to discriminate between continuous or pulsed tactile targets presented randomly to the index finger of either hand (Spence et al., 1998).

Spatial 'Advantages': Exogenous versus Endogenous Links between Audition, Vision, and Touch

The study of 'covert' attention (i.e., without eyes movements, opposite to 'overt' attention, wherein eyes movements are allowed during trials) revealed a distinction between two different component: a reflexive or *exogenous* covert orienting, which is considered to take place automatically under pure stimulus control, and a voluntary or *endogenous* covert orienting, which is under strategic control. The former is induced by uninformative peripheral cues which do not predict the target locus, but may appear directly at it: exogenous attention is 'pulled' to the location of a salient event. The latter is induced by informative symbolic cues, as a central arrow that indirectly predicts the likely target side: endogenous attention is 'pushed' to the expected target location.

As far as endogenous audiovisual links are concerned, two central arrows, which were placed immediately to the left or to the right of the central fixation point, indicated where participants had to direct either auditory or visual attention on each trial (Rorden & Driver, 2001; Spence & Driver 1994, 1996, 1997b). Generally, both left and right arrows were illuminated in neutral condition (Spence & Driver, 1996). In other works, endogenous resources were pushed to the expected target location by means of two dots immediately to the left or to the right of a central fixation point. They were used above all to study eccentricity effects: the nearest dot (relative to central fixation point) indicated the nearest target, and vice versa. Dots could be either red, indicating a following visual target, or green, indicating a following auditory target. The illumination of both red and green cue dots served as a neutral cue to signal divided attention (Rorden & Driver, 2001; Spence & Driver 1997b).

Audio-visual studies of endogenous covert attention showed that when participants voluntary attend to a particular location in one modality (because they expect a target in that modality), their attention also shifts in the same direction for the other modality (although in an attenuated manner), even when this cannot be to their strategic advantage (indeed, even when it is somewhat disadvantageous). In other words, when people strongly expect a sound on one side, there is not only an increase in of the auditory localization (with respect to the other side), but also the visual one, even if visual events are actually twice as probability on the other side. Similarly, focusing endogenous visual attention on one side, because of a strong spatial expectancy concerning visual targets, leads to an increasing of the auditory discrimination performance on that side. This suggests a symmetrical tendency for common shifts in endogenous spatial attention across the two modalities. The endogenous spatial attention effect is invariably smaller by about 50% in the secondary modality, than in the primary modality for which the strong spatial expectancy holds (Driver & Spence, 1998a). This implies that the spatial distribution of endogenous attention in the primary modality somehow spreads into other modality, but at an attenuated level. Analogous behavioural results were found for tactile modality, as concerns both visual-tactile (Driver and Spence, 1998b) and audio-tactile (Lloyd, Merat, McGlone, & Spence, 2003; Eimer, van Velzen, & Driver, 2002) links.

As concerns exogenous links between audition, vision and touch, the orthogonal cuing method showed that a spatially non-predictive cue in one modality can attract covert attention towards its location in the same or different modalities (Rorden & Driver, 1999, 2001; Spence & Driver 1994; 1997a; Spence et al., 1998). For instance, Spence et al. (1998) reported a series of crossmodal experiments, wherein participants had to make speeded discrimination response (continuous vs. pulsed) for tactile targets presented randomly to the index finger of either hand. Targets were preceded (with 150, 200, or 300 ms of SOA) by a spatially uninformative cue that was either auditory (Experiment 1) or visual (Experiment 2) on the same or opposite side as the tactile target. In Experiment 3, spatially uninformative tactile cues were presented prior to randomly intermingled auditory or visual targets requiring an elevation discrimination response (up vs. down). In Experiment 1 and 2, they found that the speeded discrimination of tactile targets was affected by the prior presentation of a spatial uninformative auditory peripheral cue or visual peripheral cue on the same versus opposite side as the subsequent tactile target. In Experiment 3, elevation discriminations for auditory and visual targets were similarly influenced by the ipsilateral versus controlateral presentation of a tactile cue, with better performance on the side of the cue shortly after it. Such results reveal without doubts crossmodal links in attention. Abrupt sounds attract visual and tactile attention to their location, not merely auditory attention. Likewise, sudden touches attract visual and auditory attention toward them, not merely tactile attention. Finally, sudden flashes attract tactile as well as visual attention.

Audiovisual links in exogenous covert orienting were already reported by Spence and Driver (1997a). They also pointed out a substantial asymmetry between them. Indeed, whereas performance was better for visual or auditory targets when preceded by an uninformative auditory cue on the side of the target (Experiment 1-4), Experiments 5 and 6 showed that no visual-to-auditory cuing effects were found unless saccades between visual cues and auditory targets were allowed. These results suggest a one-way crossmodal dependence in exogenous covert attention whereby audition influences vision, but not vice versa. Then, visual cues seems to have no effect upon auditory elevation discriminations when eye-movements towards the flash are prevented. Intriguingly, when a saccade is permitted towards the visual cue, auditory localization in the orthogonal cuing method improves in the vicinity of the cue before the eye physically moves (Rorden & Driver, 1999). This suggests that it is the release of a saccade program, rather than the shift of eye-position itself, which is responsible for the effect of visual cue upon hearing (see section ‘Inhibition of Return’).

The question of the ‘missing link’

This pattern of asymmetric audiovisual links in exogenous orienting could make sense if one considers the deep differences between audition and vision. In addition to the above-mentioned spatiotopic-versus-tonotopic arrangement of these modalities, it could be useful to take into account a more ecological perspective, according to which it is clearly more economic for an attention system does not focus too much resources on the same source. It could be clearly useful for tactile attention to be directed toward sudden visual events, in order to facilitate manipulation of perceived objects in the surrounding environment. Likewise, it could be useful for visual attention to be directed toward sudden acoustic events, in order to focus own gaze over the source of sound and to recognize possible dangers or other. By contrast, it is not immediately clear what behavioural advantage would occur from orienting auditory attention toward the location of a visual cue. Many researchers have argued that one of the primary purposes of auditory localization is the reverse influence: to direct attention toward interesting events that may be outside the current field or focus of view (Aitkin, 1986; Heffner & Heffner, 1992). According to this explanation, there is no adaptive advantage in directing auditory attention to a sudden visual event, given that any such event in the field of view will automatically result in an exogenous shift of visual attention in its direction. However, we reckon that the existence of tactile-to-auditory cuing effects showed by Spence at al. (1998) limits the validity of this fascinating theory: why exogenous auditory attention should be attracted by tactile cues and not by visual cues? Same arguments above-mentioned, indeed, should be valid to prevent the influence of touch as well as vision on audition.

Some insight might arrive from neurophysiological research. A recent study showed that although behavioural RT effects are present, spatially non-predictive visual cues which precede auditory targets modulate ERPs at short SOAs (between 100 and 300 ms; McDonald et al, 2001). Thus, such a ‘missing link’ might be simple harder to be showed in RTs performance (despite its absence remains a fact!), notwithstanding it actual operates in the spatial allocation of crossmodal visual-to-auditory resources. In fact, several neural structures (e.g., superior colliculus, parietal lobe, superior temporal sulcus, putamen, plus regions in the frontal lobe) seem to contain units that have been shown to respond to multiple modalities, with these cells coding the different modalities in approximate spatial register (Wallace, Meredith, & Stein, 1998; Andersen, Snyder, Bradley, & King, 1997; Graziano & Gross, 1996, 1998). These findings induced Driver and Spence (1998a) to propose a cellular explanation for exogenous crossmodal links, which assume that crossmodal cuing effects may be related to the activation of above multimodal neurons. They suggested:

> Activation of such multimodal cells by a cue in one modality may lower their threshold for responding to a concurrent (or closely following) event in a different modality, when at the same versus different location. This could explain the crossmodal cuing effects observed behaviourally in people. […] Cellular responses can indeed be enhanced by multimodal stimulation at a common location, and suppressed for discrepant locations, as compared with unimodal baselines (p. 256).

Thus, asymmetric pattern of spikes at multimodal neuron level in response to visual or auditory events following visual or auditory precues may be the reason of the absence of visual-to-auditory cuing effect from a behavioural point of view. Obviously, future neurophysiological studies have to solve this basic issue.

Spatial 'Disadvantages': the Inhibition of Return

Attention system seems to share between the different sensorial modalities not only synergetic behaviours like crossmodal links, but also some kind of temporal-based spatial inhibition. The inhibition of return (IOR) is a specific effect defined by those who have investigated it within the visual modality (Posner & Cohen, 1984) as a delay subsequent to uninformative cues in returning attention to a previously attended location. In the initial demonstration of IOR by Posner and Cohen (1980), participants responded to the onset of a bright probe dot appearing either in a centrally fixated box or in one of two boxes to the right and left of fixation. At the beginning of the trial, either the left or the right box was briefly increased of brightness. The target appeared at different SOAs (0, 50, 100, 200, 300, or 500 ms) after the peripheral box was brightened. In most of the trials, the target appeared in the central box; on 20% of trials, however, the target appeared in either the cued or the uncued peripheral box. At the shortest delays, RT was faster at the cue location than at the opposite location. By 300 ms, however, this effect reversed: RT was slower at the cue location. Authors originally proposed that the effect arises because covert attention is first drawn to the cued location but then moves on with an inhibitory bias against returning to the briefly attended cued location; hence the term 'inhibition of return'.

Notwithstanding of IOR is a well-documented effect in the visual domain, only in recent years it was fully demonstrated in the auditory one. Indeed, some author initially failed to find any form of auditory IOR (Mondor & Bregman, 1994), as well as several crossmodal studies in both healthy (Klein et al., 1987; Schmitt, Postma, & De Haan, 2000; Spence & Driver, 1997a;) and in unilateral parietal-lesioned patients (Farah et al., 1989) failed to observe any IOR effect from uninformative auditory cues preceding visual targets and vice versa (see also Reuter-Lorenz & Rosenquist, 1996). As argued by Spence and Driver (1998), there are several reasons why these studies may have failed to show both auditory and crossmodal IOR effects. First, none of them used a central reorienting event between cue and target onset. In the unimodal visual literature, several researchers reported that IOR is more readily apparent when another event, which is generally presented at central fixation, intervenes between cue and target (e.g., Posner & Cohen, 1984; Terry, Valdes, & Neill, 1994). The rationale is that central cues direct participants' attention back to fixation from the cue location, prior to target onset. If this is not implemented, any persistent facilitatory effects caused by exogenous covert orienting to the cue may oppose inhibitory influences as they start to emerge, and so mask the IOR. Second, the use of a limited range of cue-target SOAs means that any IOR effect emerging only at very long SOAs would have gone undetected. The maximum cue-target SOA was 500 ms in Klein et al. (1987); 700 ms in Spence and Driver (1997a); and 1000 ms in Farah et al. (1989). Third, the specific discrimination task (up vs. down) used by Spence and Driver (1997a) may be insensitive to IOR, as visual research revealed that IOR affects response latencies only in certain discrimination tasks (Christ, McRae, & Abrams, 2002; Terry et al., 1994).

Utilizing these shrewdness, Spence and Driver (1998) provided the first unequivocal evidence for crossmodal IOR. In particular, they showed the importance of target modality unpredictability: when unpredictably auditory or visual targets followed auditory or visual cues, the typical pattern of inhibition became evident (Experiment 3). The authors discussed these results in relation with saccade programming and activation of the oculomotor system. Tassinari and colleagues (Tassinari, Aglioti, Chelazzi, Marzi, & Berlucchi 1987; Tassinari, Biscaldi, Marzi, & Berlucchi 1989) were the first researchers to suggest the importance of saccade programming in IOR: they proposed that visual IOR is caused by an ‘oculomotor suppression’ of the natural tendency to make a saccadic response toward the peripheral cue event, as required by the usual central fixation instructions. The authors argue that a peripheral visual cue results in the generation of an automatic reflex to move the eyes in its direction. Participants must then prevent such an overt movement by generating a motor instruction against reacting to stimuli in the cued direction. Thus, they respond more slowly to targets appearing on the side ipsilateral to the cue: the motor set that prevents saccades to the cue also impairs simple manual responses for event in that direction. Unfortunately, already Posner et al. (1985) found crucial evidence against this fascinating theory. They reported an experiment in which participants made a saccade in the direction indicated by a central arrow in order to identify a digit presented at the peripheral location. Participants then had to re-fixate the centre, prior to target onset. Crucially, detection latencies were slower for targets appearing at the location to which a peripheral saccade had just been made. Therefore, no saccadic suppression is necessary to obtain the IOR effect, as predicted by the oculomotor suppression theory (see also Rafal et al., 1994).

The findings that IOR can occur even when saccades are actually executed to the cue location led Rafal and colleagues (Rafal, Calabresi, Brenna, & Sciolto 1989; Rafal, Egly, & Rhodes 1994) to propose an alternative account for IOR in terms of *oculomotor priming*. They suggested that it is the preparation of an eye movement itself that is both necessary and sufficient for the subsequent generation of IOR in the direction of the prepared saccade, and that if IOR occur in response to peripheral cues without any explicit instructions that invoke the eye movement system, then peripheral cues *automatically* engage or prime the oculomotor system in a manner that is sufficient to induce IOR. This hypotheses is consistent with Rafal et al.'s (1989) finding that endogenous saccade preparation produces IOR for the prepared side, even when the planned saccade is subsequently not executed. Incidentally, it is worth noting that this seems to be also the unique condition in which IOR can affect endogenous attentional mechanisms. Indeed, Posner and Cohen (1984) failed to show any inhibitory effect when attention is directed to a location in response to a central (endogenous) cue without an oculomotor activation.

Evaluating the actual involvement of the oculomotor system

The oculomotor priming theory found strong evidence in several interesting works. In particular, Schmidt (1996) reported IOR in two experiments in which participants were explicitly instructed to prepare an eye movement in response to the cue but to refrain from executing it. These results are consistent with Reuter-Lorenz and Rosenquist's (1996) finding in which they found no evidence for IOR following auditory cues (Experiment 1-3), but they did with an explicit oculomotor activation: participants had to move their eyes to the location of the auditory cue and the to return their gaze rapidly to the central fixation point (Experiment 4). These findings seem to provide a clear demonstration that automatic covert orienting alone is not sufficient to generate IOR to auditory cue, whereas they underline the decisive impact of saccade programming in the generation of inhibitory effects. Moreover, these data provide a way to analyse those works in which no auditory or crossmodal IOR appears: it could be a question of insufficient oculomotor activation. If the auditory cues were insufficient to prime the oculomotor system, then there would be no response tendency to suppress. In other words, the oculomotor priming theory suggests to consider the possibility that auditory cues used in these kinds of experiments were not able to automatically prime the oculomotor system. Otherwise, it must not be surprising the absence of IOR effect. Nevertheless, the existence of several studies in which auditory and crossmodal IOR effect is obtained without any direct implication of the oculomotor system (Röder, Spence, & Rösler, 2000; Mondor, 1999; Mondor & Lacey, 2001; Spence & Driver, 1998; Tassinari & Berlucchi, 1995), impose us to be prudent. This advice is somehow strengthen by results reported by Röder et al. (2000) who showed IOR in absence of oculomotor control using a tactile variant of the orthogonal cuing method in congenitally blind and in individual who had no eyes. Moreover, Tassinari and Berlucchi (1995) reported IOR using auditory cues and auditory targets in a simple detection task. As the main difference with the Reuter-Lorenz and

Rosenquist's (1996) experimental set seems to be the bigger eccentricity for cue and target presentation with respect to the central fixation point, it would be interesting to verify if some interaction exists between eccentricity and oculomotor system activation. It could happen that only for certain degrees of eccentricity the activation of oculomotor system is necessary to produce IOR (e.g. small eccentricity as in Reuter-Lorenz and Rosenquist, 1996), whereas at other degrees of eccentricity (e.g. large eccentricity as in Tassinari & Berlucchi, 1995; see also Ward, 1994) automatic covert orienting alone may be sufficient to generate IOR to auditory cue.

Mondor and colleagues also made many investigations about IOR (Mondor, 1999; Mondor, Breau, and Milliken, 1998; Mondor and Breau, 1999; Mondor and Lacey, 2001). In a series of interesting studies they enlarged knowledge about auditory IOR and suggested some fascinating explanations of this phenomenon. For instance, Mondor, Breau, and Milliken (1998) reported an experiment in which listeners were required to detect, localize, or identify the frequency of an auditory target following an auditory cue. At short SOA (150 ms), performance was best when stimuli sounded from the same location or were the same frequency. However, at long SOA (750 ms), performance was best when stimuli differed in location or were different in frequency. These results reveal the existence of both *location-based* and *frequency-based* auditory IOR. Mondor and Breau (1999) ruled out the possibility that auditory IOR could result because of a modulation in perceptual sensitivity, utilizing an auditory target that was presented in a wide-band noise background such that the required discrimination was made relatively difficult. Even in this condition, a facilitative effect was found at the 150 ms SOA and an inhibitory effect was found at the 750 ms SOA. Mondor and Lacey (2001) extended this pattern of results in a study in which participants had to judge the identity (Experiment 1) or the presence of a brief silent gap (Experiment 2) between auditory cues and targets with the same or different duration, intensity, or timbre. Once again, in both experiment, at short SOA (150 ms) performance was best when cue and target were identical, whereas at large SOA (750 ms) performance was best when cue and target differed. Thus, it seems to emerge the existence of duration-, intensity-, and timbre-based auditory IOR. Moreover, Mondor (1999) found that the time-course of auditory IOR depends on the predictability of the temporal and spatial relations between cue and target. More specifically, whereas an advantage for change trials emerges at about 700 ms SOA when there is no predictable relation between cue and target, it emerges at much longer SOAs when the cue is likely to provide accurate information (i.e., when there is a greater proportion of repeated than change trials) and at briefer SOAs when the cue is likely to provide incorrect information (i.e., when there is a greater proportion of change than repeat trials).

The 'dual-process' model

To explain such data, Mondor (1999) suggested a *dual-process model*, which is based on the interactions between two separate processes, one that controls the transmission of acoustic information to auditory short-term memory and the other involved in response assignment. More specifically, Mondor suggested that an attentional template or filter acts to govern the speed with which auditory information is passed on to the memory. This template is normally defined by basic auditory features, such as frequency, location, duration, intensity, and timbre (Mondor & Lacey, 2001). The precise definition of the template varies from one moment to the next as a function of both incoming acoustic information and the listener's current goal. The degree to which the perceptual features of an incoming sound matches the template determines the speed with which it is transmitted to auditory short-term memory. Mondor suggested that in the typical cue-target paradigm, the template may be set in accordance with the properties of the advance cue. Then, if the subsequent target matched these settings (as would occur on a repeat trial), selection would be accomplished quickly. In contrast, if the subsequent target did not match the template, selection would be accomplished more slowly. Irrespective of whether a target is selected quickly or slowly, a representation of it is passed on to memory, at which point a response assignment process may begin.

According to the model, accurate response assignment requires that the memory representation of the target be distinguished from that of the cue. This differentiation process depends on the similarity of the cue and the target representations, with the result that the speed of differentiation increases as the similarity of the two representation decreases. Then, whereas the selection processes accomplished at the template engenders an advantage for repeat trials (i.e., a facilitative effect), the memory representation differentiation process involved in response assignment produces an advantage for change trials (i.e., an inhibitory effect). Mondor (1999) suggests that the time-course of auditory IOR is determined by a trade-off in the importance these opposing influences. As a facilitative effect normally is apparent only for quite brief SOAs of a few hundred milliseconds, he suggests that specific settings of the parameters of the template dissipate over the course of about 300-400 ms. Once the specific template definition has dissipated, the action of the memory process is revealed in a performance advantage for change trials. In essence, Mondor suggested that inhibitory effect apparent a long SOAs is indicative of inhibition at all but, rather, reflects the fact that the time required to differentiate two different memory representations depends on similarity.

This model is consistent not only with the expression of location-, frequency-, duration-, intensity-, and timbre-based IOR (Mondor & Breau, 1999; Mondor et al., 1998; Mondor & Lacey, 2001), but also with experiment demonstrating that the time-course of auditory IOR changes significantly as a function of the predictability of the relation between cue and target (Mondor, 1999).

Hypotheses and Models

The main attempt to build an exhaustive model of spatial attention orienting mechanisms can be trace within the well-know (and quite alive) debate on supramodal versus modality-specific nature of attention system. The lack of definitive evidence in favour of one of these alternative assumptions will be object of discussion in the following sections.

A Supramodal or Modality-Specific System?

One of the main purposes of crossmodal research is to establish the exact nature of the attentional system. Verifying analogies and differences between attentional mechanisms of several modalities (vision, auditory, touch) is maybe the best way to assess whether attention can be described as a unitary system able to shift attentional resources in different cognitive tasks or as an amount of several modality-specific systems largely independent from each other. As concerns the *supramodal hypothesis* (SMH), the attention would be based on anatomical circuits separated from the data processing systems. It was originally described as a unitary system (Posner & Petersen, 1990), although a more recent version of this model assumes two different attentional mechanisms: a posterior one, involved with spatial attention tasks, and an anterior one, involved in the control of wide brain structures in order to perform complex cognitive tasks (Posner & Dehaene, 1994). The *modality-specific hypothesis* (MSH) postulates the absence of any specific attentional system: attention would derive only from the activity of the same circuits that process sensory data (Rizzolatti & Camarda, 1987). According to this model, auditory attention would be a modality-specific system, which shows different features from the other modality-specific attentional systems. Obviously, there's a third possibility: attention could operate thought *separate-but-linked* systems (SBLH). As suggested by Spence and Driver (1996): "Visual and auditory attention neither operate in a purely supramodal manner, nor just within totally independent modality-specific systems. Instead, there are separable systems for vision and audition, but spatial synergies exist between them" (p. 1026).

Measuring the effect size

Although there is not definitive support for any of such hypotheses and the debate about this crucial issue is quite alive, it is possible to characterize two main lines of research existent in the literature. The first one is based on the cuing method and the rationale is strictly connected with SMH predictions. According to SMH, cuing effects should be observed when cues and targets are of different modalities, and they should be equal in size to those found in unimodal conditions. Farah et al. (1989) reported (in a study spoiled by some methodological bias, see section 'Early studies') that patients with right parietal lobe damage showed a similar pattern of cuing effects for both visual and auditory cues to visual targets and this similarity led the authors to conclude that parietal lobe mechanisms allocate attention to a supramodal representation of space. Similarly, Cielsielski, Knight, Prince, Harris and Handmaker (1995) reported a neurophysiological study wherein a clear reduction in the amplitude of the slow negative wave in a divided attention task with respect to a focused attention task. As such a reduction appears with both visual and auditory stimuli, authors postulated the existence of a *central distributor of attentional resource*, common to all modalities. Thus, the slow negative wave may be an index modality independent of the attentional resources available to the individual. However, it is worthwhile to underline that the similarity between pattern of cuing effect is not enough to assume the existence of a unitary attentional system (Spence & Driver, 1996). In other words, these findings actually do not support the SMH, as similarity of effects does not logically imply identity of mechanisms (Ferlazzo, Couyoumdjian, Padovani, & Olivetti Belardinelli, 2002).

Within the same line of research, however, other studies support the alternative hypothesis as concerns both exogenous and endogenous attentional mechanisms. Mondor and Amirault (1998) found that cuing effects are larger when cues and targets belong to same modality than when they belong to different modalities, even though they were both significant. According to the authors, these results are consistent with the existence of partially separated systems for auditory and visual exogenous orienting, given that if the effect of spatial cues was mediated by a single system, then the modality manipulation would be irrelevant. Similarly, if the visual spatial system dominated covert orienting, then an effect of visual cues upon auditory target identification would be expected, along with a null effect of auditory cues on visual target identification, but this did not happen. Analogously, Spence and Driver (1996) found the same pattern of results (a larger cuing effect within-modality with respect to crossmodality) studying endogenous attentional mechanisms. Besides, they found that participants were able to simultaneously allocate visual and auditory attention to two opposite spatial locations, if the likely target side for each modality was held constant for each block of trials (their 7th experiment is maybe the strongest evidence for the independence of visual and auditory attention systems). They also observed that when the same side was most likely for both modalities, cuing effects were larger than when visual and auditory attention had to be directed to different side. As the latter result is not consistent with completely separated visual and auditory attention systems, they formulate the SBLH.

Finding out a discrimination index

The second line of research is involved with the attempt to characterize an index able to discriminate, with respect to some relevant feature or cognitive process, visual and auditory attention systems. The rationale is that the more significant differences between modality-specific systems exist, the more is likely their independence and so the validity of the MSH. This could be represented by the spatial representation subserving visual and auditory attention systems. More specifically, a recent study of Ferlazzo et al. (2002) examined the so-called 'meridian effect' which is due (within the frame of the Premotor Theory of Attention) to the updating of the oculomotor program, necessary when individuals redirect attention across the visual horizontal or vertical meridians (Rizzolatti, Riggio, Dascola & Umiltà, 1987; Rizzolatti, Riggio, & Sheliga, 1994). The general version of Premotor Theory of attention holds spatial selective attention processes are embedded within cortical areas involved in programming motor actions related to specific sets of effectors. Thus, the attentional effects might be to the activity of the same areas, which subserve data processing or motor planning. In this context, the visual meridian effect provide a further instrument to investigate the nature of attention, as if one supramodal system exists, then a similar meridian effect would be found when persons were required to attend to both auditory and visual targets; in the opposite case, a different meridian effect would be found on auditory orienting with respect to the visual one, due to the programming of a different sets of effectors. Ferlazzo et al. (2002) found that when auditory targets were used, RTs were slower on trials in which cued and target locations were at the opposite side of the vertical head-centred meridian than on trials in which cued and target locations were at the opposite sides of the vertical visual meridian or were not separated by any meridian. The head-centred meridian effect for auditory stimuli was apparent when targets were cued by either visual (Experiments 2 and 3) or auditory cues (Experiment 5). Alternatively, participants showed a visual meridian effect when they were required to respond to visual targets (Experiment 4). These results strongly suggest that auditory and visual spatial attentional systems are separated, even though the authors seem to lean towards a less strong version of the MSH. Indeed, they see the above results as a better specification of the SBLH, in the sense that the ‘separable’ part might consist of the use of different spatial maps by the two attentional systems. In other words, it would be at least a difference as far as mechanisms involved in the meridian effect are concerned.

More recently, evidence from ERP studies seems to favour again the SMH (Eimer & Van Velzen, 2002; Eimer et al., 2002; McDonald, Teder-Salejarvi, Heraldez, & Hillyard, 2001; see also Eimer & Driver, 2001). For instance, Eimer and Van Velzen (2002) reported an experiment wherein participants had to detect peripheral tactile or visual targets on the attended side, while ignoring all stimuli on the unattended side and in the currently irrelevant modality. Both relevant locations and relevant modalities were specified on a trial-by-trial basis by auditory precues. Spatial orienting in the cue-target interval was reflected in anterior negativities and occipital positivities contralateral to the cued side, either when vision or touch was cued as relevant. Moreover, early effects of spatial attention on somatosensory and visual ERPs were of equivalent size for currently relevant and irrelevant modalities. These effects of ERP modulations during shifts of crossmodal attention imply, according to the authors, supramodal mechanisms in the control of spatial attention and demonstrate their independence of cue modality. A similar pattern of results in endogenous spatial attention between vision, audition, and touch was found by Eimer et al. (2002). Taken together, these results support the idea that crossmodal links in spatial attention are mediated by supramodal control mechanisms, likely involved with a frontoparietal neural circuit, and such a conclusion stands in apparent contrast with findings supporting SBL hypothesis (Ferlazzo at al., 2002; Spence & Driver, 1996). However, we argue that the existence of a cerebral frontoparietal structure, which shows a similar activity for crossmodal attention tasks, not necessarily rule out that attention (because of the existence of other unclear factors) could exhibit different behaviour in tasks involving different modalities, as shown by the meridian effect.

Implications for Traditional Theories and Models of Attention

Although any definitive word can be said about the nature and the structure of attention, above-mentioned data imply a radical changing in previous theories and models concerning attentional processes, and new theoretical efforts in formulating coherent and comprehensive framework for future research. Indeed, several theories and models of attention were proposed in the last decades, but none of them are able to take into account all data above reviewed. For instance, premotor theory suggested by Rizzolatti and colleagues (Rizzolatti & Camarda, 1987; Rizzolatti et al., 1987; Rizzolatti et al., 1994) postulated that attentional effects are due to the activity of the same areas, subserving data processing or motor planning. According to this theory, spatial selective attention processes are embedded within cortical areas involved in programming motor actions related to specific sets of effectors. However, this assumption stands in contrast with neurophysiological data that show the existence of a specific (frontoparietal) neural circuit involved in attentional processes and separated from acquisition of sensory data (Eimer & Van Velzen, 2002; Eimer et al., 2002; McDonald et al., 2001).

The feature integration theory (FIT) proposed by Treisman and colleagues (Treisman & Gelade, 1980; Treisman & Gormican, 1988) also appeared inconsistent with above-mention findings in two important respects. According to the FIT, focal attention is required to perform a discrimination or detection task only when a stimulus is processed for a conjunction of features (e.g., colour and form), not when performance is based entirely on a variation of single features. However, studies analysed in the section “Selection of auditory information” show attentional effects when listeners were required to attend only one relevant features (see for example Mondor et al., 1998. Experiment 1), as location and frequency information were conjoined. Moreover, the absence of any evidence in favour of a special or dominant role of location information in guiding auditory selective attention (Mondor et al, 1998; Mondor & Terrio, 1998) also appears to be inconsistent with the FIT. It is worth noting that FIT derived above all from visual attention findings. Then, it is not surprising to be inconsistent with auditory attention findings.

Duncan and Humphery’s (1992) attentional engagement theory (AET) provides an alternative conception of selective attention, which is proposed to emerge from three processing stages. In the first one, a structural description of the input stimulus is elaborated. In the second one, particular structural units from the description derived in the first stage are selected. Selection is directed by an attentional template that specifies the relevant stimulus attributes appropriate for the current goals of the observer. Critically, selection occurs on the basis of a similarity match between structural units and the attentional template. In the last stage, selected information is transferred to a limited-capacity visual short-term memory, where responses are determined. Although attentional template suggested by the AET seems to be consistent with several findings (Mondor & Bregman, 1994; Mondor et al, 1998; Mondor & Zatorre, 1995), it cannot be considered more than a starting point in the building of a comprehensive model of auditory and crossmodal (which is not taken into account by the AET) attention.

Computational models also provide a specific explanation of attentional mechanisms, modelling through neural networks the cellular activity of involved. Wang and Brown (1999) proposed a speech recognition system based on a neural oscillator network. In their model an external signal, which was constituted by a mixture of speech and noise, was analysed from a sort of peripheral auditory processing which consisted of a bank of cochlear filters. The gains of the filters are chosen to reflect the transfer function of the outer and middle ears. In turn, a model of hair cell transduction, giving a probabilistic representation of auditory nerve firing activity, processes the output of each filter channel. Through a *correlogram* (formed by computing a running autocorrelation of the auditory nerve activity in each filter channel) authors obtained the so-called *mid-level* auditory representation. The next stage comprised the core of their model, in which auditory organization took place within a two-layer oscillator network. The first layer produced a collection of segments that corresponded to the elementary structures of an auditory scene, and the second layer grouped segments into streams, in according to the Bregman's (1990) view. Conceptually, segments were the atomic elements of the represented auditory scene. The last stage of this model was a re-synthesis path, which allowed an acoustic waveform to be derived from the time frequency regions corresponding to a group of oscillators.

It is worth noting that this approach is consistent with neurophysiological findings, giving to the model a neurobiological foundation. Recently, Maldonado and Gerstein (1996) observed that neurons in the auditory cortex exhibit synchronous oscillatory firing patterns. Similarly, deCharms and Merzenich (1996) reported that neurons in separate regions of the primary auditory cortex synchronize the timing of their action potentials when stimulated by a pure tone. In fact, a neural oscillator network can be see as model of action potential generation or oscillatory burst envelope (Brown et al, 2001; Wang & Brown, 1999; Zahn, Izak, Trott, & Paschke, 1998), in which are simulated membrane potential and several level of activation of a number of ion channels. Single oscillator units can be considered as artificial processors able to synchronize their activity on a particular frequency region of input sounds: oscillators that belong to the same stream are synchronized (phase locked with zero phase lag), and are desynchronised from oscillators that belong to different streams.

However, this Wang and Brown's model lack of a directed simulation of an auditory attention system. Wrigley and Brown (2000, 2001), proposed a model strongly based on the above mentioned oscillator systems. The basic innovation is represented by the attentional leaky integrator (ALI), which is able to select one block of oscillators (if their oscillatory activity coincides temporally with a peak in the ALI activity) to become the attentional stream: each oscillators is, indeed, connected to the ALI by excitatory links and the strength of these connections is modulated by a sort of endogenous attention: ALI activity will coincide only with the activity of the channel within the attentional interest peak.

However, as a consequence of only recent development of an approach firmly based on principles of human auditory functions and neurobiological plausibility (Churchland & Sejnowski, 1995; Santangelo & Olivetti Belardinelli, 2002), today's computational models appear again too simplicistic to catch the complexity of attentional mechanisms so far outlined.

Overall, it is clear that above-mentioned theories and models are too limited to provide an explanation of recent findings here reviewed and to integrate them in a coherent and solid framework.

CHAPTER 2 - Crossmodal Divided Attention: Endogenous Comparison between Single and Double Cue Presentation

As illustrated in the previous chapter, the literature concerning the study of attentional processes suggests three different hypotheses about the unitary versus multimodal nature of the attention system. The *supramodal hypothesis* (SMH), according to which the attention system, based on anatomical circuits separated from the data processing systems, was originally described as a unitary system (Posner & Petersen, 1990). More recent versions of this theory assume two different attentional systems: a posterior system, which is independent from the sensorial data processing and involved in spatial attention tasks, and an anterior system involved in the control of wide brain structures in order to perform complex cognitive tasks (Posner & Dehaene, 1994). The *modality-specific hypothesis* (MSH) postulates the absence of any general attention system: attention would derive from the activity of the same circuits that process sensory data that is visual, auditory and tactile attention would be modality-specific systems, each of them showing different features from the others (Rizzolatti & Camarda, 1987; Rizzolatti, Riggio & Sheliga, 1994). Other evidence underlines the existence of spatial synergies between visual, auditory and tactile attention systems, in both exogenous and endogenous mechanisms (Spence and Driver, 1996, 1997). Such crossmodal links are the crucial point for the *separated-but-linked hypothesis* (SBLH).

As each of these hypotheses is supported by experimental evidence (see section “A supramodal or modality-specific system?”, Chapter 1), it seems quite difficult to achieve a conclusive point about the nature of attention system. This work is aimed at contributing to the debate by closer examining the real independence of the visual and the auditory attentional systems. We used a modified version of the *orthogonal cuing paradigm* (Spence & Driver, 1994) wherein participants were required to make an elevation discrimination task of visual targets preceded by visual cues, auditory cues, or both. Participants were instructed to pay attention to both visual and auditory (endogenous) cues, as the first or the second one, or both, indicated in the 80% of trials the correct side in which visual targets were delivered, facilitating their elevation discrimination task. The rationale was that if each attention modality-specific system is quite independent from the others (as postulated by the MSH), then no differences in participants' performance would exist between a condition wherein the attentional load engage only one modality (single visual or auditory cue) and another wherein the attentional load engaged two modalities (bimodal -visual + auditory- cue). On the other hand, if the attention system is supramodal (SMH), then to split attentional resources over two modalities would dramatically decrease performance in case of bimodal cue compared with single modality cues. Finally, if SBLH is true, then performance differences would exist but attenuated with respect to the previous case. In other words, if modality-specific attention systems are separated but strongly connected from a functional point of view, then it is reasonable to attend a significant decrease of performance in case of bimodal cues with respect to single cues, but also enough independence to avoid a sharp decrease of performance as though the system was supramodal.

Materials and method

Subjects

Fourteen participants (6 men) naive as to the purpose of the study were recruited for this experiment. They gave their informed consent for the experiment and reported normal (or corrected-to-normal) vision and hearing. Their mean age was 24.6 years [21-29]. All participants were above the criterion of 90% of correct responses and included in statistical analysis.

Apparatus and stimuli

Participants were comfortably seated at a table, facing straight ahead with a PC monitor of 15″ with a light-grey background. The fixation point consisted in a black plus (with a diameter of 0,5 cm) placed at the centre of the monitor. As concerns the stimuli, visual cues consisted in a black arrow (which replaced the fixation cross) pointing either to the left or to the right side. Auditory cues consisted in two syllables ('si' and 'de' which in Italian language are the first syllables of words 'left' and 'right') uttered by a male voice. Each syllable (synthesized at a sampling rate of 22000 Hz and linearly tapered for 5 ms in order to eliminate onset and offset clicks) had duration of 100 ms and was delivered stereophonically at 75 dB by the two loudspeakers cones (8 cm in diameter, situated at the middle line of screen, at ears level) at the same time, giving a subjective impression to be in correspondence to the centre of the screen. Participants responded by pressing one of two keys placed in a vertical arrangement on a response device situated in front of them. RTs were measured in tenth of milliseconds from target onset.

Procedure

Figure 2.1 shows the sequence of events in this experiment. Participants were seated in an armchair in a silent and darkened chamber. The fixation cross was presented at the beginning of each trial in the centre of the screen. Participants were instructed to maintain their fixation on this central cross. Involuntary head movements were prevented by using an adjustable chin-rest. After a random delay with range 600-1000 ms, a visual cue, an auditory cue, or both, equally likely to be on the left or on the right side, were delivered for 100 ms. In case of double cue presentation, the auditory cue (equally likely to be on the left or right loudspeaker) was delivered at a random interval between visual cue offset and visual target onset. Participants were instructed to pay attention to the cues, which indicated with a probability of 80% the correct side in which the target would occur. They were informed about this probability. After a further delay of 600 ms, the visual target was delivered at one of the corners of the screen for 100 ms. Participants were required to press the key farthest from them for a visual target from either of the upper target positions (regardless of the side) and the nearest key for a target from either of the lower positions, responding as rapidly and

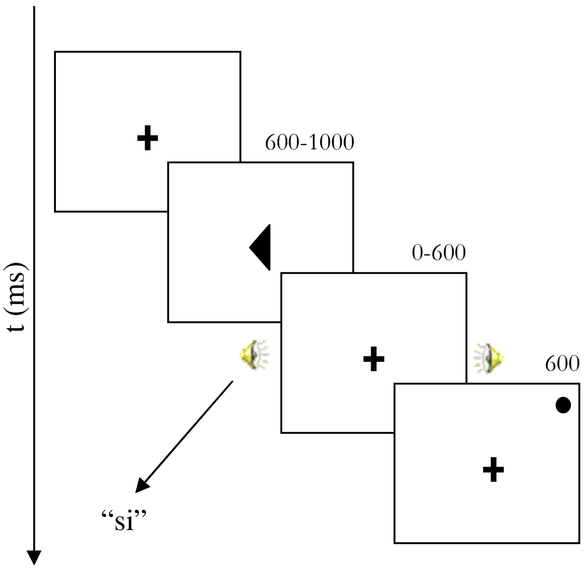


Figure 2.1: Sequence of events. The fixation cross was presented at the beginning of each trial in the centre of the screen. After a random delay with range 600-1000 ms, a visual or an auditory cue (or both), equally likely to be on the left or on the right side, were delivered for 100 ms. In case of bimodal cue presentation, the auditory cue (equally likely to be on the left or right loudspeaker) was delivered at a random interval between visual cue offset and visual target onset. After a further delay of 600 ms, the visual target was delivered at one of the corners of screen for 100 ms. Participants were required to press the key farthest from them for a visual target from either of the upper target positions (regardless of the side) and the nearest key for a target from either of the lower positions, responding as rapidly and accurately as possible with the index finger of either hand. Once a response was given (within a time limit of 2000 ms) a new trial began and so on until the of the block.

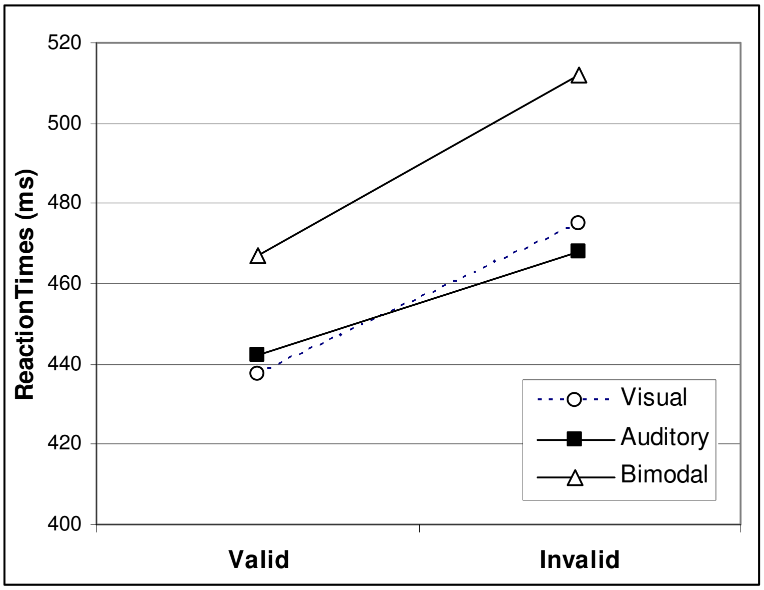


Figure 2.2: Mean RTs for valid and invalid trials in case of visual, auditory and bimodal cue.

accurately as possible with the index finger of either hand. Once a response was given (within a time limit of 2000 ms) a new trial began and so on until the end of the block. In order to stimulate participants, at the end of each block we gave them a feedback concerning their performance (number of correct responses, misses, and false alarms). Between blocks they were allowed to rest for a few minutes.

In sum, a peripheral (visual, auditory or bimodal) cue indicated one of two possible sides (2), the target occurred at one of four possible screen corners (4), and required either an up or a down button press (2). Each combination, which varied randomly, were delivered in 5 different blocks of 320 trials each, for an amount of 1600 trials, in which the 80% were valid, and the rest equally shared between visual-valid, auditory-valid and invalid trials.

Results

A 3x2 within-subjects analysis of variance (ANOVA) with the factors *cue type* (visual, auditory and bimodal) and *cue validity* (valid and invalid) was conducted on both reaction times (RTs) and percentages of error (PEs). Responses faster than 100 ms (premature) or slower than 1500 ms (miss) and errors were excluded from analysis. The data were collapsed across left and right responses and side of presentation. Visual- and auditory-valid trials were not analysed, as no specific theoretical hypotheses exist about that in relation to the aim of this study. Huyhn-Feldt procedures were applied to all F-tests to correct for possible violations of the sphericity assumption.

As concerns RTs, ANOVA found a main effect of the cue type [*F*(2, 26) = 3.388, *p* = 0.049], with a better performance for visual and auditory (means of 456 ms and 455 ms, respectively) than for bimodal cues (490 ms), as revealed by the post-hoc comparison (Duncan's test). An effect of cue validity was also found [*F*(1, 13) = 64.087, *p* = 0.001], with faster responses for valid (449 ms) then for invalid trials (485 ms). The interaction between these two factors was not significant [*F*(2, 26) = 2.296, *p* > 0.05] (Table 2.2 and Figure 2.2).

As concerns PEs (average of 3.2%), ANOVA failed to show significant effect of cue type [*F*(2, 26) = 0.609, *p* > 0.05], cue validity [*F*(1, 13) = 0.301, *p* > 0.05] and cue type X cue validity interaction [*F*(2, 26) = 3.169, *p* > 0.05] (Table 2.2 and Figure 2.3).

	RTs (ms)				**Errors (%)**			
	Valid		**Invalid**		**Valid**		**Invalid**	
Type of Trial	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Visual Cue	437.57	32.56	474.94	40.08	3.9	1.8	2.9	1.1
Auditory Cue	442.37	28.07	468.13	45.37	2.8	1.6	3.2	1.8
Bimodal Cue	467.18	41.33	512.01	50.00	2.5	1.0	3.7	2.0

Table 2.2: Means and standard deviations for both reaction times and percentages of error.

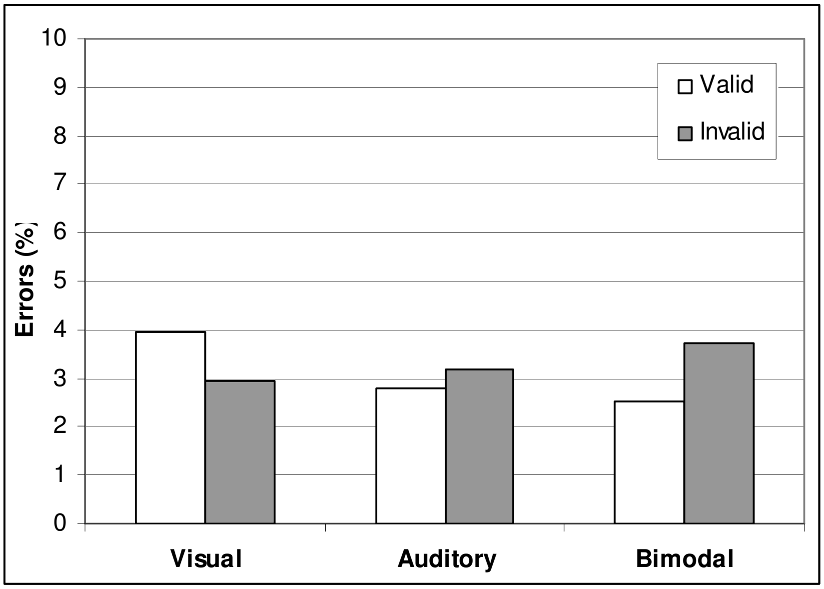


Figure 2.3: Mean PEs for valid and invalid trials in case of visual, auditory and bimodal double cue.

Discussion

The aim of this paper was to understand the real independence of visual and auditory attention systems. We used an experimental task that involved a crossmodal split of attentional resources (bimodal cue), and compared this behavioural performance with tasks that involved only one singular modality (either visual or auditory).

Results showed that reaction times were faster in case of singular unimodal (visual and auditory) cues than in case of bimodal cues. This effect seems to confirm that participants found more difficult to split their attentional load over two modalities than simply attend a unimodal cue. Although this data seems to support the SMH (if visual and auditory attention system had been separated then no differences between single and double cued conditions would have found), we must consider that such an effect was statistically reliable but not so large in absolute terms (about 30 ms between single and double cued conditions), and no differences in the error data analysis were found, while the SMH would postulate a larger impairment of performance in for bimodal cues with respect to visual and auditory cues. In other words, the decrease of performance in case of bimodal cues was not so dramatic as predicted by the SMH and it seems to rule out the existence of a singular unique attention system.

However, an alternative explanation is also possible. As we delivered (in case of bimodal cues) the second (auditory) cue at a random interval between the first (visual) cue offset and the target onset, it is possible that the decrease of performance in double cued conditions reflects the general working of a supramodal system which have to switch between visual and auditory spatial information. In other words, it might be possible that the difference in RTs performance between single and double cued conditions reflects the time needed by a single unique system to reallocate spatial orienting resources from the visual (cue) to the auditory (cue) modality. Future research should be addressed on this issue, simply using a synchronized bimodal cue presentation (see Chapter 4 and Chapter 5), evaluating with a simultaneous task-demand on visual and auditory attention whether a disruption of performance occurs or not.

Results showed also that responses were faster for valid than for invalid trials. Thus, participants were able to use advance spatial information in both single and double cued conditions in order to localize and then discriminate the following target. This crossmodal cuing effect strengthens the idea of a bi-directional connection or ‘spatial synergies’ (Spence and Driver, 1996, pp.1026) between visual and auditory attention systems.

Performance differences between visual, auditory and bimodal cues seem to rule out the existence of completely separated modality-specific systems, which should accomplish visual and auditory allocation of attentional resources independently from each other, as predicted by the MSH. Our results support the SBLH, as showed by the observed crossmodal links (faster RTs for valid than invalid trials when the visual target was preceded by auditory and double cues), and by the not dramatic decrease of performance in bimodal compared to single cued conditions. Moreover, this data provide at least two basic specifications of the SBLH: first, the time needed to allocate attentional resources on different spatial locations is function of the attentional load demanded by specific tasks, as showed by the increase of RTs in the double cued condition; second, the attention system is able to use spatial information derived from crossmodally different (audio-visual) cues in order to allocate attentional resources (as showed by the validity effect we found in double cued conditions in both experiments) in the elevation-discrimination task. This one seems to be a new relevant result with respect to the previous literature. Indeed, Spence and Driver (1996) reported a study in which they used a central visual arrow in order to indicate the likely (80%) side for target in one modality, but the unlikely (20%) side for target in the other, in a manner that the optimal strategy should be to shift endogenous auditory attention in one direction and endogenous visual attention in the opposite direction (Experiment 6). They found that it was difficult or impossible for participants to split their endogenous auditory or visual attention, but when the authors blocked the likely target side for each modality instead of cueing the likely target sides anew on each trial (Experiment 7), they found clear effects of spatial expectancy on RTs. The present study showed the existence of cuing effect in a crossmodal split condition also with a trial-by-trial procedure.

CHAPTER 3 - Crossmodal Divided Attention: Timing Integration between Audio-Visual Information

Although temporal dynamics seems to play a key role in the most of cognitive processes, just a few studies directly addressed on these central topic. The reasons for this are diverse, but mostly ascribable to an intrinsic difficulty to build experimental paradigms able to catch the whole complexity of this crucial domain.

Particularly interesting appears to be the temporal dynamics in case of multisensorial integration, because of time constraints likely determined the intrinsic nature of integration. For instance, multimodal brain structures as the superior temporal sulcus (STS) and the superior colliculus (SC) respond to both auditory and visual stimulation (Stein, Meredith & Wallace, 1993) but show a super-additive effect just in case of spatially and *temporally* coincident multimodal stimulation (Stein & Meredith, 1993). Thus, the timing relation of multisensorial events appears to be crucial for integration processes.

In this framework, it appears quite relevant to investigate whether it exits a temporal window in which audio-visual information can be bound together, and how attention allocates resources when the relevant information is delivered at shot time intervals. Although many studies have been recently addressed to multisensorial integration (e.g., Molholm, Ritter, Murray, Javitt, Schroeder, & Foxe, 2002; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002; see Chapter 1 for an extensive review, p. 18), evidence concerning the time course of audio-visual information is still relatively modest.

The aim of this experiment was to investigate the temporal features of audio-visual integration. We suggest a method strongly based on the bimodal condition depicted in the previous chapter (see p. 40), but using fixed intervals of milliseconds in which the auditory cue occurred. More specifically we delivered the second (auditory) cue at different Stimulus Onset Asynchronies (SOA) in order to establish the trend of cuing effect through time. If the information provided by the visual cue and the auditory cue can be bound together and used for the spatial discrimination task in a narrow time window, then we should find that the longer is the SOA (i.e., the longer is the temporal distance between the visual and the auditory cue) the longer are the RTs, with a consequent decrease of performance. Otherwise, if the process of audio-visual integration can be accomplished without substantial differences through time, then the cuing effect should be stable for each level of SOA.

Materials and method

Subjects

Fourteen participants (8 men) were recruited for this experiment. They were naive as concerns the purpose of the experiment and gave their informed consent. All participants (mean age of 23.9 years, range [19-27]) reported normal or corrected-to-normal vision and hearing. All of them were above the criterion of 90% of correct responses and included in statistical analysis.

Apparatus and stimuli

The apparatus and stimuli were the same of the experiment reported in the previous chapter (see p. 39).

Procedure

Participants were seated in an armchair in a silent and darkened chamber. The fixation cross was presented at the beginning of each trial in the centre of the screen. Participants were instructed to maintain their fixation on this central cross. Involuntary head movements were prevented by using an adjustable chin-rest. After a random delay with range 600-1000 ms, a visual cue, equiprobable on the left or right side, was delivered for 100 ms. An auditory cue (equally likely to be on the left or right loudspeaker) was delivered at the same time of the visual cue (0 ms interval) or at a SOA of 200, 400, 600 or 800 ms. Participants were instructed to pay attention to the cues, which indicated with a probability of 80% the correct side in which the target would occur. They were informed about this probability. The visual target was delivered after 800 ms of visual cue onset at one of the corners of the screen for 100 ms. Participants were required to press the key farthest from them for a visual target from either of the upper target positions (regardless of the side) and the nearest key for a target from either of the lower positions, responding as rapidly and accurately as possible with the index finger of either hand. Once a response was given (within a time limit of 2000 ms) a new trial began and so on until the end of the block. In order to stimulate participants, at the end of each block we gave them a feedback concerning their performance (number of correct

	RTs (ms)				**Errors (%)**			
	Valid		**Invalid**		**Valid**		**Invalid**	
SOA	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
0	313.00	55.62	332.79	52.43	3.2	2.1	3.4	3.5
200	311.14	52.77	333.29	51.32	3.3	2.4	3.7	2.3
400	311.21	46.11	338.86	47.38	3.5	2.4	3.4	2.5
600	376.29	40.83	382.14	39.20	2.6	1.5	3.6	2.3
800	406.86	35.18	405.57	45.55	3.0	1.6	3.2	2.6

Table 3.1: Mean and standard deviations of RTs and PEs in valid and invalid trials at the different SOAs.

responses, misses, and false alarms). Between blocks they were allowed to rest for a few minutes.

In sum, a visual cue that indicated one of two possible sides (2) as well as an auditory cue (2) preceded at various SOAs (5) the target, which occurred at one of four possible screen corners (4), and required either an up or a down button press (2). Each combination, which varied randomly, were delivered in 5 different blocks of 320 trials each, for an amount of 1600 trials, in which the 80% were valid, and the rest equally shared between visual-valid, auditory-valid and invalid trials.

Results

A 5x2 within-subjects analysis of variance (ANOVA) with the factors *SOA* (0, 200, 400, 600, 800 ms) and *cue validity* (valid and invalid) was conducted on both reaction times (RTs) and percentages of error (PEs). Responses faster than 100 ms (premature) or slower than 1500 ms (miss) and errors were excluded from analysis. The data were collapsed across left and right responses and side of presentation. Huyhn-Feldt procedures were applied to all F-tests to correct for possible violations of the sphericity assumption.

As concerns the RTs, ANOVA found a main effect of the SOA [$F(4, 52) = 53.823$, $p = 0.001$], with a better performance in case of 0, 200 and 400 ms of SOA (means of 323, 322 and 325 ms, respectively) than in case of 600 ms of SOA (379 ms), which in turn produced faster responses than 800 ms of SOA (406 ms). An effect of cue validity was also found [$F(1, 13) = 25.729$, $p = 0.001$], with faster responses for valid (344 ms) then for invalid trials (359 ms). These two factors significantly interacted [$F(4, 52) = 6.612$, $p > 0.002$], indicating absence of validity effect when the SOA grew larger, i.e. 600 and 800 ms of SOA (see Table 3.1 and Figure 3.1).

As concerns the PEs, ANOVA failed to find significant effects (see Table 3.1 and Figure 3.2). Overall, participants made very few mistakes (average of 3.3%).

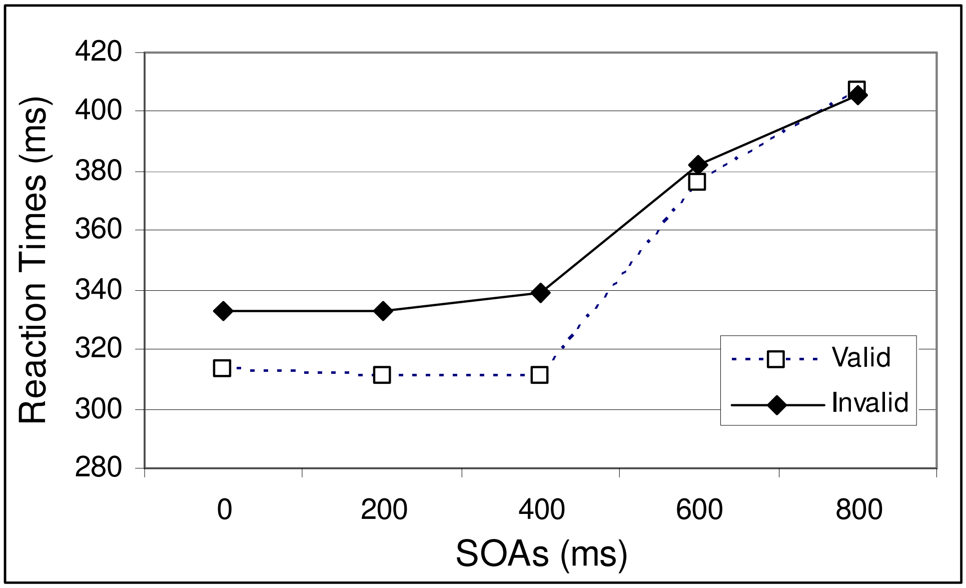


Figure 3.1: Mean PEs for valid and invalid trials at the different SOAs.

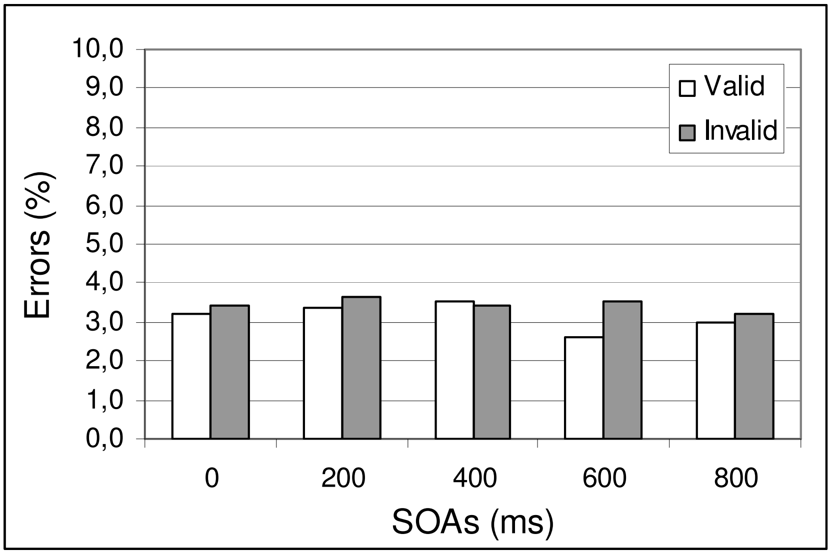


Figure 3.2: Mean PEs for valid and invalid trials at the different SOAs.

Discussion

The primary purpose of this study was to investigate the time intercourse of audio-visual integration by means of an experimental procedure wherein a visual target was preceded by two informative cues delivered at different SOAs. Results clearly showed that performance remains good till 400 ms of SOA. No substantial differences were found, indeed, between 0, 200 and 400 ms of SOA. When SOA grew larger, however, RTs sharply decreased, indicating that the second (auditory) cues might be too close to the target to be used in the spatial discrimination task.

This assumption seems to be supported by the observed attentional orienting effects. Until 400 ms of SOA, robust cuing effects were found (magnitude of 20, 22 and 28 ms, for 0, 200 and 400 ms of SOA, respectively), while the cuing effect vanished for the longer SOAs (magnitude of 6 and -1 ms for 600 and 800 ms of SOA, respectively) (see Figure 3.3). Thus, data seems to indicate the existence of a *temporal window* of 400 ms in which the audio-visual information can be bound together and fruitful used in the visual spatial discrimination task.

However, it exists at least an alternative explanation for the observed results. Obviously, when subjects were required to perform a task, they formed a mental representation of instructions.

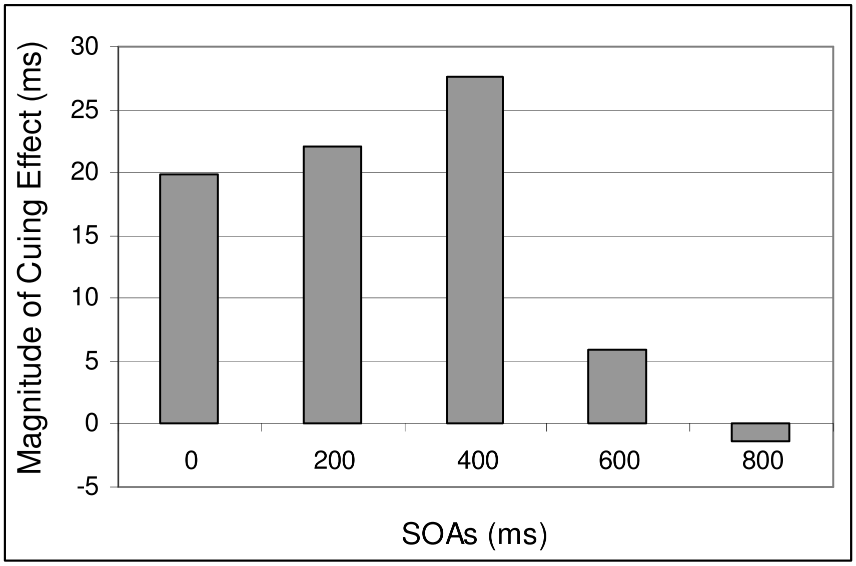


Figure 3.3: Magnitude of cuing effects at the different Stimulus Onset Asynchronies (SOA).

In this experiment the mental representation dealt with the three consecutive stimuli: the visual cue, the auditory cue and the target. Participants knew that the target was always preceded by the two cues. Thus, when they were delivered in a close temporal proximity (let's say within a window of 400 ms), participants knew the target was quite imminent, producing an increase in their *alertness state*. Instead, When the interval between the two cues corresponded to 600 and 800 ms, the auditory was probably too close to the target to allow participants to increase their alertness state, with a consequent decrease of performance. Given that it was no possible to control for the role of the alertness effect in this study, results should be considered cautiously. Much more research appears to be needed to take into account possible biases and to really understand the effective contribution of multisensorial integration in the allocation of attentional resources in space. For this reason, this central topic is going to be deeply investigated also in the following two chapters and, particularly, in Chapter 5, wherein the influence of multisensorial integration on attention orienting mechanisms will be examined by means of event-related brain potentials.

CHAPTER 4 - Spatial Attention Triggered by Unimodal, Crossmodal and Bimodal Exogenous Cues: A Comparison on Reflexive Orienting Mechanisms

The literature concerning reflexive (exogenous) mechanisms of spatial attention has shown that reaction time (RT) decreases when a spatial cue (e.g., a flash of light or a burst of sound), being unpredictive with regard to the location of a forthcoming target, is delivered on the target side (i.e. a validly cued trial) as compared to when it occurs on the non-target side (i.e. an invalidly cued trial; Posner, 1980). This 'cuing effect' has been found both in unimodal (Posner, 1980; Posner & Cohen, 1984; Henderson & Macquistan, 1993; Hopfinger & Mangun, 1998) and in crossmodal settings (Spence & Driver, 1997; Schmitt, Postma & De Haan, 2000; McDonald, Teder-Sälejärvi, Di Russo & Hillyard, 2003; Teder-Sälejärvi, McDonald, Di Russo & Hillyard, 2002; Van der Lubbe & Postma, in press). The classical interpretation of these cuing effects is that a cue automatically attracts attention which enhances the processing of a subsequent target near the cued location, whereas attention has to be reoriented when the target occurs at another location.

Several electrophysiological studies have attempted to map the underlying neurophysiological correlates of attention orienting. Hopfinger and Mangun (1998) investigated the influence of reflexive attention in unimodal settings with visual stimuli (i.e. visual cues and visual targets), and found support for an influence of exogenous cues on processing within extrastriate visual brain areas, which indicates that early sensory processes are affected by reflexive attention. McDonald et al. (2003) investigated exogenous crossmodal orienting by employing a task where visual targets were preceded by abrupt sounds. They found that activation of unimodal visual areas (the fusiform gyrus of the ventral occipito-temporal cortex, and the peri-sylvian cortex of the inferior parietal lobe) was preceded by modulations in multimodal areas (superior temporal sulcus/gyrus, STS/STG). Therefore, the authors suggested that the orienting of attention towards sudden sounds initially facilitates processing of spatially coincident visual events in areas in which visual and auditory input converge, and that the subsequent modulation in unimodal extrastriate areas is a consequence of re-entrance feedback from multisensory integration regions (STS/STG) into the ventral processing stream. Thus, unimodal studies point to a locus of the effect of covert exogenous orienting in extrastriate areas (Hopfinger & Mangun, 1998), while crossmodal studies to an initial locus in STS/STG, followed by an effect in extrastriate areas (McDonald et al., 2003).

While these spatial attention effects have been typically obtained for single cue situations, it has rarely been examined what happens when two cues from different modalities are presented at the same time. In daily life, however, bimodal stimuli occur rather frequently. There is considerable evidence that simultaneous inputs from different modalities support each other in an additive or super-additive fashion. In non-human mammals, neurons in the STS and superior colliculus (SC) respond to both auditory and visual stimulation (Stein, Meredith & Wallace, 1993) and show a super-additive effect in case of spatially and temporally coincident multimodal stimulation (Stein & Meredith, 1993). In other words, multimodal stimuli significantly enhanced the responses of STS and SC neurons above those evoked by the sum of either unimodal stimuli (Wallace, Meredith & Stein, 1998). In addition, Teder-Sälejärvi, McDonald, Di Russo and Hillyard (2002), compared event-related potentials (ERP) to auditory (noise bursts) and visual (flashes) stimuli alone to combined audiovisual stimulation (simultaneous noise bursts and flashes) and found enhanced neural activity (ERPs) in the later case, localized in the anterior temporal peri-sylvian cortex. The proximity between this structure and STS/STG regions (McDonald et al., 2003) leaves open the possibility that these effects might reflect the activity of the same neural circuit operating in case of multimodal (audio-visual) tasks.

Thus, a relevant question is whether spatial attention triggered by bimodal events acts differently as compared to singular unimodal or crossmodal events, which may provide important information regarding the architecture of exogenous orienting mechanisms. Indeed, bimodal cuing paradigms can provide important insights in the level at which multiple cues from different locations are integrated as well as in the exact time course of the integration process. In order to investigate these issues in our first experiment, we compared a condition wherein a visual target was preceded by both visual and auditory exogenous cues delivered together at the same side (bimodal cue), with conditions wherein the visual target was preceded by either the visual (unimodal cue) or the auditory cue (crossmodal cue) in a spatial attention discrimination task. Furthermore, to examine possible differences in activation patterns over time we varied the time interval between cue and target (stimulus onset asynchrony: SOA). Namely, SOAs were either 200, 400, or 600 ms.

With regard to the underlying mechanisms of exogenous orienting, three different theoretical viewpoints and predictions may be distinguished. In case of the existence of a supramodal spatial attention module (e.g., see Farah, Wong, Monheit & Morrow, 1989; Eimer & van Velzen, 2002) comparable cuing effects on RT in case of unimodal, crossmodal and bimodal cues may be predicted to occur, as the supramodal module can be triggered by these different types of cues, which from that level on affects visual processing. Summing up of cuing effects may be predicted to occur when unimodal and crossmodal cues affect different processes. In other words, the allocation of attentional resources may be accomplished by independent visual and auditory attentional systems, that are triggered by completely separate pathways (Rizzolatti, Riggio & Sheliga, 1994). Finally, in case of *interactive effects* of which AV (i.e., audio-visual interaction) is not equal to the sum of A + V (either underadditive or overadditive), separate-but-linked auditory and visual attention systems (Spence & Driver, 1997) with overlapping (or at least partially overlapped) pathways might underlie mechanisms of exogenous attentional orienting towards unimodal, crossmodal and bimodal cues.

In other words, if attentional orienting has an advantage of the integration of different modalities then the cuing effect on RT should be larger in case of bimodal cue presentation than in case of unimodal or crossmodal cue presentation.

EXPERIMENT 1

Method

Participants

Seventeen participants, all students of Utrecht University, participated. They had normal or corrected-to-normal vision and normal hearing, and were paid € 10 for their participation. Before the start of the experiment, they had to indicate the side of auditory stimuli at the cued locations, and a score of at least 95% was required to participate. One participant was excluded from the analysis, because of too many eye movements (> 40%), leaving sixteen participants (mean age 23.4 years, range 18-34, 6 male, 15 right-handed, 1 ambidextrous). The study was approved by a local ethics board of the University of Utrecht, and informed consent was obtained from all participants.

Stimuli and Materials

Stimuli were presented on three crossmodal units (21 * 12 cm), consisting of a sound passing 8 * 8 green LED display (10 * 10 cm) in front of a loudspeaker. The units were hung at 160 cm from participant, either 29.4° to the left or the right, or in front of the participant. The fixation point was a dot (0.7 * 0.7°) and a line (0.2 * 3.1°) presented for 100 ms served as visual cue. The auditory cue was a burst of white noise presented for 100 ms. As visual targets, triangles (2.6 * 1.4°) were used pointing up– or downwards, which were presented for 100 ms.

Tasks and Procedure

Three conditions were included in which visual and/or auditory cues preceded visual targets. The modality of cues was varied between conditions, and their order was counterbalanced. In the *unimodal* condition, visual targets were preceded by a visual cue; in the *crossmodal* condition, visual targets were preceded by an auditory cue; in the *bimodal*

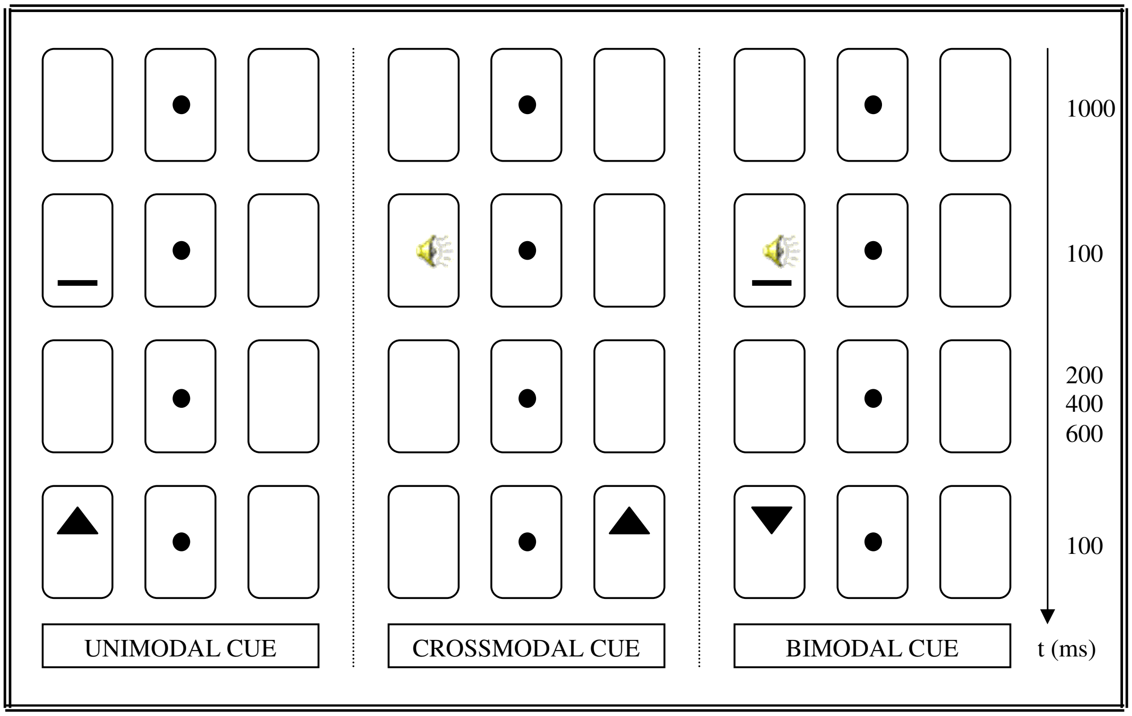


Figure 3.1: Sequence of events and their duration (starting from the top) in unimodal, crossmodal, and bimodal conditions. Each trial began with a fixation point, which remained until the end of trial. After 1000 ms, a visual cue (unimodal), an auditory cue (crossmodal), or both of them (bimodal) appeared for 100 ms equiprobable on the left or right side, depending on cued condition. After a variable interval depending on SOAs of 200, 400, or 600 ms, the target (triangle) appeared for 100 ms equiprobable on the left or right side, requiring a left or right button press when triangle pointed up- or downwards, respectively.

condition, visual targets were preceded by both visual and auditory cues simultaneously delivered at the same unit.

Participants were seated in an armchair in a silent and darkened chamber. Left and right response buttons were placed on a hand rest in front of the participant. Each trial started with a fixation dot on the middle unit. After a delay of 1000 ms, the cue (either unimodal, crossmodal or bimodal, depending on the condition) was presented (for 100 ms) with equal probability on the left or right unit. After a further delay with SOAs varying randomly on a trial-by-trial basis from 200, 400, to 600 ms, the visual target was presented, equiprobable on the left or the right unit (see Fig. 1). Unimodal, crossmodal and bimodal cues were unpredictive with regard to the forthcoming target position, and were either valid or invalid. Participants were required to press a left or right button as fast and accurately as possible for either a triangle pointing up– or downwards. The next trial started after a response or at 1500 ms after target onset.

Each type of trial (72 combinations - type of cue (3) x side of cue (2) x side of target (2) x required response (2) x SOA (3)), was repeated 15 times.

EOG Recording

Presentation of stimuli and the emission of triggers signalling the moment and the type of stimulus was controlled by a CMO-module (version 3.7f, IGF, Physics department). The triggers were received by Vision Recorder (version 1.0b, BrainProducts, GmbH), which measured the participants' eye movements (the horizontal and vertical electro-oculogram (EOG)) and button presses. EOG was recorded at a rate of 1000 Hz (TC = 5.0 s, low-pass 100 Hz) from Ag/AgCl ring electrodes placed above and below the left eye and at the outer canthi of both eyes. The baseline was determined from –100 to 0 ms before presenting the cue. A criterion of +/- 60 µV from cue onset until target offset was used, which implies that trials with eye movements larger than approximately 3° (determined by Vision Analyzer 1.03) were excluded.

Data Analysis

Responses faster than 100 ms (premature) or slower than 1500 ms (miss) and errors were excluded from RT analysis. The data were collapsed across left and right responses and side of presentation. Reaction times (RTs) and proportion of correct responses (PCs) were evaluated by repeated-measures ANOVAs with the factors Cue type (unimodal, crossmodal or bimodal), SOA between onset and target (200, 400, 600 ms), and Cue validity (valid or invalid). Huyhn-Feldt procedures were applied to all F-tests to correct for possible violations of the sphericity assumption.

Results

Mean RTs are displayed in Table 4.1. Trials with eye movements (8.3% of the trials) were excluded from further analyses. Both premature responses and misses occurred seldomly (0.1% and 1.85% of the trials, respectively). ANOVA showed faster responses for validly cued (548 ms) than for invalidly cued targets (559 ms) [$F(1, 15) = 7.23$, $p = 0.017$]. Responses were slower in case of an SOA of 200 ms (567 ms) than in case of an SOA of 400 (546 ms) and 600 ms (544 ms), [$F(2, 30) = 17.50$, $p = 0.001$]. Responses were faster when the cue was crossmodal (525 ms) than when it was bimodal (558 ms), which, in turn, was faster than when it was unimodal (574 ms), [$F(2, 30) = 8.33$, $p = 0.001$].

The SOA * Cue validity interaction was significant [$F(2, 30) = 7.953$, $p = 0.002$], showing a validity effect at 200 [$F(1, 15) = 13.775$, $p = 0.002$] and 400 ms of SOA [$F(1, 15) = 9.604$, $p = 0.007$], but no effect at 600 ms of SOA [$F(1, 15) = 0.300$, $p = 0.592$]. The significant SOA * Cue type interaction [$F(4, 60) = 3.863$, $p = 0.023$] indicated that responses became faster for the medium and the long SOAs with respect to the short SOA in case of bimodal cues (sizes of reduction of 29 and 39 ms, respectively), but less so in case of unimodal cues (23 and 23 ms) and crossmodal cues (13 and 8 ms). ANOVA showed a nearly significant SOA * Cue type * Cue validity interaction [$F(4, 60) = 2.49$, $p = 0.057$], and a significant Cue type * Cue validity interaction [$F(2, 30) = 3.53$, $p = 0.043$], pointing out different magnitudes of cuing effects, which seem to depend on SOA. Contrast analyses showed that this was due to a reverse effect (longer RTs for validly than for invalidly cued targets) for bimodal cues at an SOA of 600 ms, as revealed by the Cue type * Cue validity

	Visual				**Auditory**				**Bimodal**			
	Valid		**Invalid**		**Valid**		**Invalid**		**Valid**		**Invalid**	
SOA	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
200	581.2	107.8	597.4	115.7	526.3	106.4	538.8	98.3	570.3	117.6	590.8	132.7
400	556.5	116.8	575.7	132.8	512.5	93.0	526.2	95.5	549.3	115.2	553.8	117.5
600	564.4	112.5	568.7	109.9	520.2	94.1	528.7	96.8	551.7	107.6	531.3	92.6

Table 4.1: Mean RTs and standard deviations for unimodal, crossmodal and bimodal cues in Experiment 1.

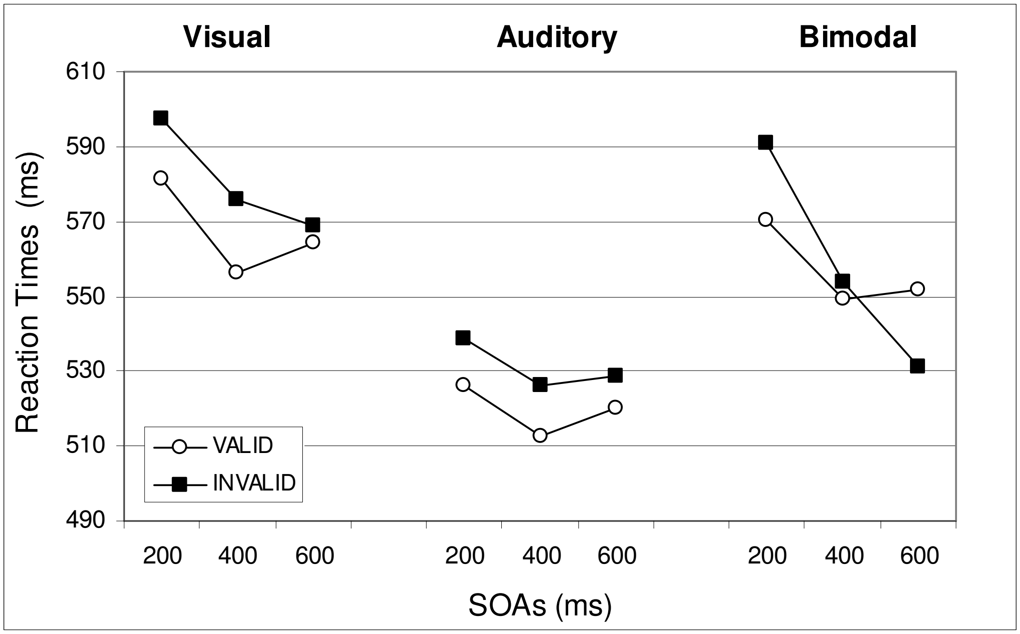


Figure 4.2: Mean RTs (valid and invalid trials) for unimodal, crossmodal and bimodal cues at 200, 400 and 600 ms of SOA in Experiment 1.

	Visual				**Auditory**				**Bimodal**			
	Valid		**Invalid**		**Valid**		**Invalid**		**Valid**		**Invalid**	
SOA	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
200	93.2	4.7	93.0	6.4	96.4	2.5	94.9	3.8	94.3	5.0	91.3	9.5
400	94.2	3.8	93.5	4.2	96.0	4.0	96.0	4.3	95.5	4.8	92.0	7.0
600	94.7	3.1	95.6	2.9	96.6	3.5	95.5	3.4	94.2	5.8	94.3	6.8

Table 4.2: Mean PCs and standard deviations for unimodal, crossmodal and bimodal cues in Experiment 1.

interaction [$F(2, 30) = 6.20$, $p = 0.006$], while this interaction was far from significant with SOAs of 200 ms [$F(2, 30) = 0.36$, $p = 0.695$] and 400 ms [$F(2, 30) = 1.88$, $p = 0.170$] (see Figure 4.2).

With regard to PCs (see Table 4.2), ANOVA with the factors cue type (unimodal, crossmodal or bimodal), SOA (200, 400, 600), and cue validity (valid or invalid) showed a main effect of cue validity [$F(1, 15) = 6.55$, $p = 0.022$], with less errors in case of validly (5.0%) than in case of invalidly cued (6.0%) targets, and a main effect of SOA [$F(2, 30) = 4.29$, $p = 0.023$], with less errors when SOA increases from 200 (6.7%), 400 (5.5%), until 600 ms (4.8%). The number of errors was less when the cue occurred on the target side with SOAs of 200 and 400 ms, but this effect vanished at an SOA of 600 ms (SOA * Cue validity) [*F*(2, 30) = 3.21, *p* = 0.057] (see Figure 4.3).

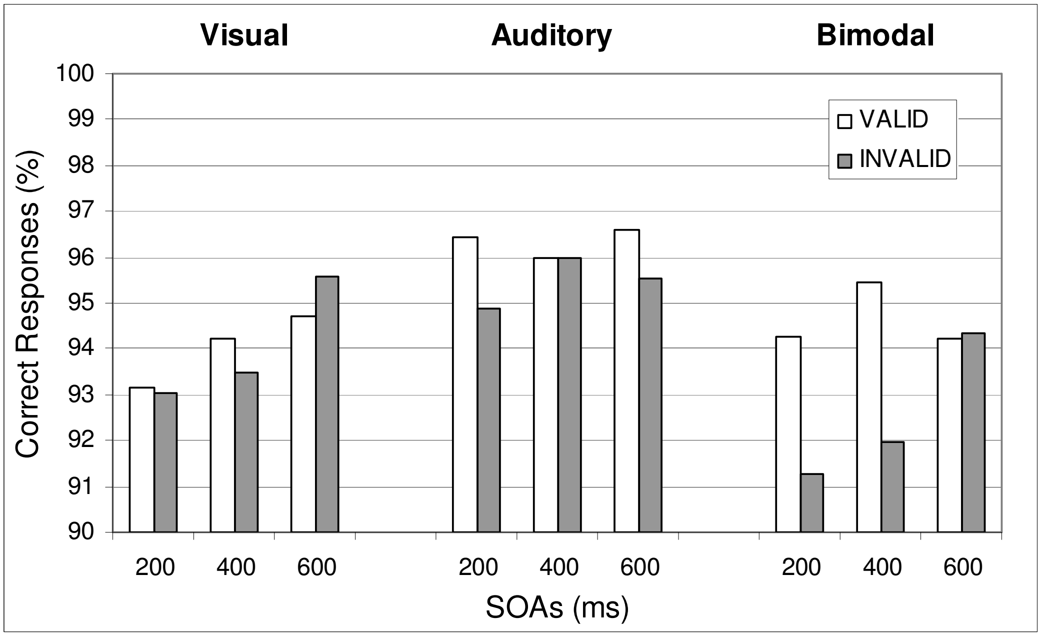


Figure 4.3: Mean PCs (valid and invalid trials) for unimodal, crossmodal and bimodal cues at 200, 400 and 600 ms of SOA in Experiment 1.

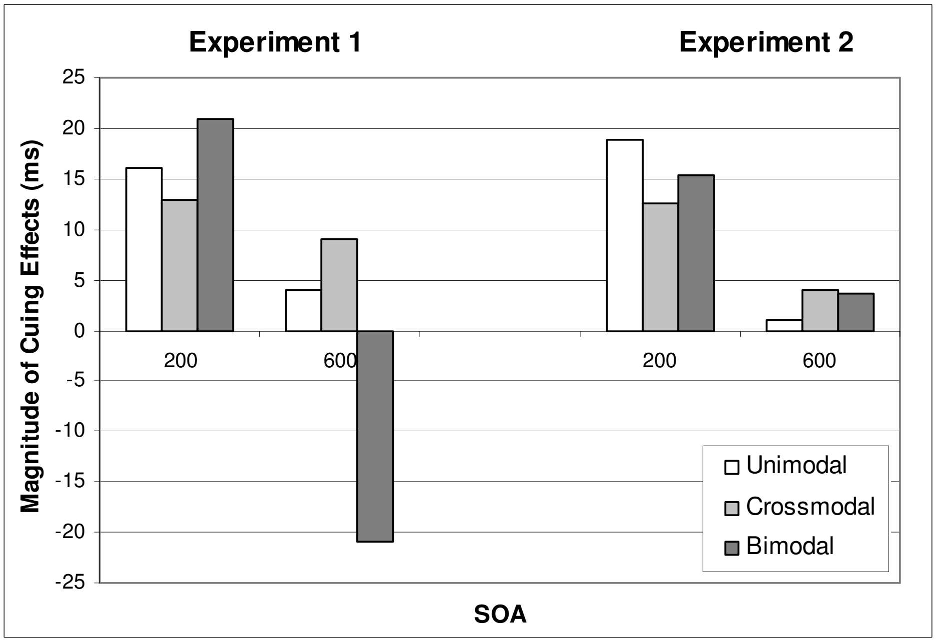


Figure 4.4: Magnitude of cuing effects (ms) for unimodal, crossmodal, and bimodal conditions at 200 and 600 ms of SOA in Experiment 1 and 2.

Discussion

The results showed comparable magnitudes of cuing effects on RTs in unimodal, crossmodal and bimodal cuing conditions for the SOAs of 200 (16, 13 and 21 ms, respectively) and 400 ms (19, 14 and 5 ms, respectively), but a reversal in case of bimodal cues (an inhibitory effect) for the longest SOA (4, 9 and -21 ms, respectively; see Figure 4.4). The comparable magnitudes (or absence of an interactive effect) for the SOAs of 200 and 400 ms do not support the separate pathways view or the separate-but linked systems view, but seem to favour the existence of a supramodal spatial attention module. However, the results for the longest SOA appear more difficult to understand. It may be argued that the current data once more show that the integration of multiple multimodal events (see Stein et al., 1993; Teder-Sälejärvi et al., 2002) is a process unrelated to exogenous orienting mechanisms as no improvement occurred when attention was reflexively attracted by an abrupt bimodal event as compared to a unimodal or crossmodal (either visual or auditory) event. Clearly, as argued by McDonald, Teder-Sälejärvi and Ward (2001), “Attributing modulations in behaviour or neural activity to a spatial attention mechanism is a nontrivial task” (p. 1791). Indeed, it cannot be taken for granted that multisensory audio-visual integration processes plays a role in generating crossmodal spatial attention effects. Thus, on the basis of present results we may state that if audio-visual integration took place in case of bimodal cues then this integration was ineffective at the level of attentional orienting.

One important problem for the supramodal view concerns the inhibitory effect for bimodal cues at an SOA of 600 ms. Namely, in case of the existence of a supramodal spatial attention module one may expect to observe comparable time courses of cuing effects. However, this inhibitory effect may have a non-attentional cause, as cues might additionally invoke response tendencies (e.g. see Van der Lubbe, Vogel, & Postma, apr), and these tendencies may become stronger in case of bimodal cues. In our second experiment, we controlled for this possibility by including a large amount of catch trials, which should reduce response tendencies, but is unlikely to affect attentional orienting. Additionally, we decided to employ more repetitions per type of trial to improve reliability. Most relevant results in Experiment 1 were obtained for the shortest (200 ms) and the longest (600 ms) SOA, therefore, we no longer included the intermediate SOA (400 ms).

EXPERIMENT 2

Method

The method was the same as Experiment 1, except for three differences. First, in this experiment we used only two SOAs (either 200 or 600 ms). Second, on 20% of the trials no target occurred (catch trials). Third, the number of repetitions for each trial (48 combinations - cue type (3) *x* cue side (2) *x* target side (2) *x* required response (2) *x* SOA (2)), was increased from 15 to 50 times (2400 trials) plus 600 catch trials (200 for cue type, i.e. 100 repetition for side of cue presentation) for an amount of 3000 trials, which were delivered in six different blocks, two for each cue type. EEG was additionally measured and reported in Chapter 5.

Each type of trial (48 combinations - cue type (3) *x* cue side (2) *x* target side (2) *x* response buttons (2) *x* SOA (2)), was repeated 50 times

Participants

Informed consent was obtained from twelve students, who were paid 25 € for their participation (mean age 22.1 years, range 19-32, 4 male, 11 right-handed, 1 left-handed) .

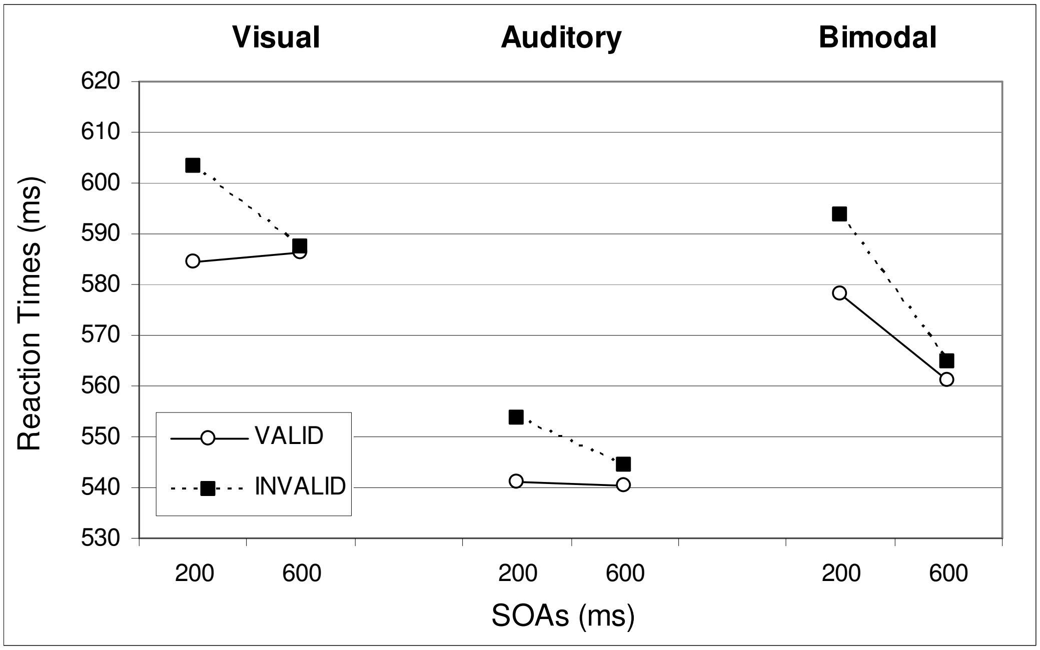


Figure 4.5: Mean RTs (valid and invalid trials) in unimodal, crossmodal and bimodal cues at 200 and 600 ms of SOA in Experiment 2.

	Unimodal				**Crossmodal**				**Bimodal**			
	Valid		**Invalid**		**Valid**		**Invalid**		**Valid**		**Invalid**	
SOA	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
200	584.5	82.7	603.3	92.2	541.0	96.2	553.6	92.6	578.3	81.8	593.7	89.7
600	586.4	78.6	587.4	88.6	540.5	89.1	544.5	90.3	561.2	83.6	564.8	88.7

Table 4.3: Mean RTs and standard deviations for unimodal, crossmodal, and bimodal conditions in Experiment 2.

Results

Mean RTs are displayed in Table 4.3. Trials with eye movements (8.7% of the trials) were excluded. Both premature and misses responses occurred seldomly (0.01% and 0.1% of the trials, respectively). ANOVA showed faster responses when the target was delivered at the cued side (565 ms) than in the opposite case (575 ms) [$F(1, 11) = 5.95$, $p = 0.033$]. Responses were slower in case of an SOA of 200 ms (576 ms) than in case of an SOA of 600 ms (564 ms) [$F(1,11) = 13.54$, $p = 0.004$]. Besides, responses were faster when the cue was crossmodal (545 ms) than when it was bimodal (575 ms), which, in turn, was faster than when it was unimodal (590 ms) [$F(2, 22) = 5.70$, $p = 0.015$]. A nearly significant interaction SOA * Cue type [$F(2, 22) = 3.61$, $p = 0.057$], seems to indicate that responses became faster for the long SOA in case of bimodal cues (size of reduction of 23 ms), but less so in case of unimodal cues (7 ms) and crossmodal cues (5 ms). The interaction SOA * Cue validity was nearly significant as well [$F(1, 11) = 4.752$, $p = 0.052$], showing a validity effect for the shortest SOA [$F(1, 11) = 6.852$, $p = 0.024$] but no effect for the longest SOA [$F(1, 11) = 0.805$, $p = 0.389$].

Crucially, neither the SOA * Cue type * Cue validity interaction [$F(2, 22) = 0.32$, $p = 0.727$] nor the Cue type * Cue validity interaction were significant [$F(2, 22) = 0.30$, $p = 0.922$], which supports the presence of equivalent magnitudes of cuing effects with unimodal, crossmodal and bimodal cues at both SOAs (see Figure 4.5).

As concerns the PCs (Table 4.4), ANOVA with the factors cue type (unimodal, crossmodal or bimodal), SOA (200 or 600 ms), and cue validity (valid or invalid) failed to reveal significant effects (see Figure 4.6). Overall, participants made few mistakes (average of 8.8%).

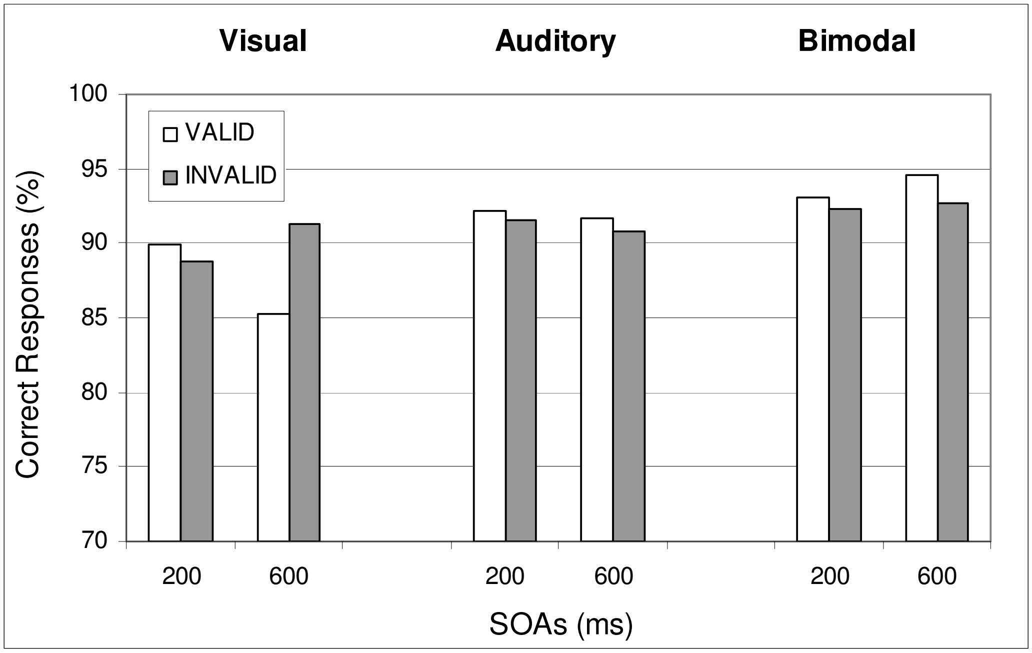


Figure 4.6: Mean PCs (valid and invalid trials) in unimodal, crossmodal and bimodal cues at 200 and 600 ms of SOA in Experiment 2.

	Unimodal				**Crossmodal**				**Bimodal**			
	Valid		**Invalid**		**Valid**		**Invalid**		**Valid**		**Invalid**	
SOA	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
200	89.9	9.0	88.8	9.0	92.2	7.0	91.5	6.6	93.0	4.3	92.3	5.1
600	85.3	24.4	91.3	6.6	91.7	7.8	90.8	6.7	94.6	3.4	92.7	4.3

Table 4.4: Mean PCs and standard deviations for unimodal, crossmodal, and bimodal conditions in Experiment 2.

Discussion

This second experiment by and large replicated the results of Experiment 1. No reversal for the bimodal condition with an SOA of 600 ms was found (as depicted in Figure 2) instead no cuing effect was present (1, 4 and 4 ms for unimodal, crossmodal and bimodal cues, respectively), whereas comparable magnitudes of cuing effects were found in case of the shortest SOA (19, 13 and 15 ms, respectively; see Fig. 2). The elimination of the reversal seems due to the use of catch trials, which reducing or even eliminated a stronger response tendency for bimodal cues.

GENERAL DISCUSSION

The primary purpose of this study was to investigate whether spatial attention triggered by multimodal events (bimodal cue) acts differently as compared to unimodal (visual cue) and crossmodal (auditory cue) events. If a supramodal module underlies mechanisms of spatial attention orienting, then no difference should be found in the magnitude of cuing effects in unimodal, crossmodal, and bimodal conditions (Farah, Wong, Monheit & Morrow, 1989; Eimer & van Velzen, 2002); while interactive effects (either $A + V = AV$ or $A + V \neq AV$) should be revealed by results if the allocation of attentional resources is accomplished by independent visual and auditory attentional systems that are triggered by separate pathways (Rizzolatti, Riggio & Sheliga, 1994) or in case of separate-but-linked auditory and visual attention systems as a consequence of the spatial synergies operating between them (Spence & Driver, 1997).

The results showed comparable magnitudes of cuing effects on RTs between unimodal, crossmodal and bimodal conditions. In other words, the bimodal cue did not produce a larger cuing effect (both RTs and PCs) with respect to both the unimodal and the crossmodal cue.

Although in Experiment 1 the average magnitude of the cuing effect at an SOA of 200 ms was higher for bimodal cue in terms of absolute values (21 ms for bimodal cue against 16 ms for unimodal and 13 ms for crossmodal cue), this difference was far from significant. Moreover, in Experiment 2 the higher magnitude in terms of absolute values was for unimodal cue (19 ms against 13 ms for crossmodal and 15 ms for bimodal cue). At an SOA of 600 ms the average cuing effect was significantly different (Experiment 1), but this result may be due to a stronger response tendency in case of bimodal than in case of unimodal and crossmodal cues. To examine this, a second experiment was conducted with catch trials, which should reduce response tendencies. Importantly, now a comparable magnitude of the cuing effect was found for bimodal, unimodal and crossmodal cues.

Taken together, the foregoing findings support the existence of a supramodal spatial attention module, which appears to be flexible enough to account for involuntary allocation of attentional resources toward stimuli coming from different modalities, and multimodal objects (bimodal cue) that combines different sensorial (audio-visual) attributes.

One might argue that the comparable sizes of cuing effects are due to a ceiling effect, i.e. no difference was found because no further improvement was possible for valid with respect to invalid trials, independently from the experimental manipulation of the cue type. However, this possibility may be ruled out taking into account the different type of alertness effect occurred for crossmodal (auditory) and bimodal cues. The observation that responses were faster after auditory cues than after bimodal cues in both experiments (main effect of cue type) indicate that responses could have been faster, thereby suggesting that more room for improvement was present in the bimodal condition (i.e., performance in case of bimodal cue did not reach the ceiling!). Indeed, as Stein and Meredith (1993) noted, multimodal stimuli that are presented simultaneously will not always reach multimodal integration areas (as the superior colliculus) simultaneously. The reason is that there are substantial differences in the processing and conduction times from the receptor surface to the SC in different modalities. For instance, it takes approximately 13 ms for a sound occurring near the ear to reach the SC, whereas visual stimuli require around 80 ms (incidentally, the effect of these differences in conduction and transduction latencies could explain the general pattern of alertness occurred in unimodal, crossmodal and bimodal conditions). If performance in case of bimodal cue got the ceiling, then it is reasonable to assume that attentional orienting mechanisms used all possible benefits derived from bimodal cues (both visual and auditory components). It follows that the auditory component in the bimodal stimulus (which arrives around 65-70 ms before the visual component) should have provided an higher alertness (just in case of the single auditory cues), determining faster orienting and responses. Otherwise it appears clear that performance occurring for bimodal cue reflects a real constrain of a supramodal spatial orienting module rather than a ceiling effect. Of course, this constrain deals with the elaboration of spatial information derived from the visual component of bimodal cue. It is likely, indeed, that when the elaboration of the auditory component is finished and the allocation of attentive focus can start, the system is busy with the elaboration of the visual component, which arrives several tens of milliseconds after the auditory component to the specific cortical areas of elaboration.

To conclude, this study revealed no difference in reflexive attentional orienting towards unimodal, crossmodal and bimodal events. A similar mechanism appears to be involved when spatial attention is triggered by both single (either visual or auditory) and double (audio-visual) stimuli, which supports the existence of a supramodal spatial attention module.

CHAPTER 5 – On the Influence of Crossmodal Integration on Exogenous Orienting: An Event-Related Potential Study

The study reported in the previous chapter has essentially shown the absence of any kind of integration operated by spatial orienting mechanisms when attentional resources are triggered by multimodal objects. The comparable effect obtained comparing single and double cue conditions seems to be a clear evidence of a single and supramodal spatial attention module that orients the attentional focus towards unimodal (either auditory or visual) and multimodal (audio-visual) events in a complete similar manner. In other words, the synchronous presentation of audio-visual information did not fasten orientation processes and the integration of these information appears to be accomplished during a following stage which does not involve any spatial orienting mechanism. This result could be due to a failure in audio-visual integration. This failure could indicate either that there was no crossmodal integration, or that crossmodal integration took place at a later processing level not affecting exogenous orienting.

However, an investigation conducted only with behavioural measures (reaction times and percentages of correct responses) seems to be not enough specific to shed light on underlying (neural) mechanisms of spatial orienting of attention, and consequently, to provide definitive evidence on this issue. Evidence can be found in the literature showing no proof of audio-visual integration in case of behavioural measures. On the contrary, clear audio-visual integration processes were found when the approach is electrophysiological. For instance, Fort et al. (2002) investigated the existence of early audio-visual interactions during bimodal (non-redundant) target identification. Although no specific facilitation in RTs was found in case of bimodal targets compared with unimodal targets, the event-related potential analysis revealed early (< 200 ms latency) crossmodal activities in sensory-specific and non-specific cortical areas. Thus, a more sensible approach appears to be useful in order to provide a stronger evidence also in case of spatial attention processes.

The aim of this study was to deeply investigate the neural basis and electrophysiological mechanisms of reflexive orienting of attention towards unimodal and multimodal events. To examine whether crossmodal integration was effective we measured ERPs elicited by *task-irrelevant* bimodal, crossmodal, and unimodal cues, and determined their exogenous orienting effects on behaviour in a visual discrimination task. The rationale for using task-irrelevant cues derives from the attempt to examine only exogenous orienting mechanisms and not the neural activity related to the specific discrimination task. Indeed, as Teder-Sälejärvi et al. (2002) argued: "Studies of cross-modal interactions in sensory processing need to take into account task-related neural activity that may be elicited following each stimulus but may not necessarily have anything to do with integration of information across different modalities. [...] On the other hand, designs that do not assign any task-relevance to the unimodal and bimodal stimuli […] would not be likely to encounter this problem." (p. 114).

Obviously, if orienting attentional mechanisms take advantage when the trigger is a multimodal event, this advantage should result in ERP components; otherwise, no difference will be found in ERP components elicited by single and double cue conditions. More specifically and following the same rationale used by Teder-Sälejärvi et al. (2002), the neural basis of crossmodal interactions was investigated by subtracting the ERPs to the single auditory (A) and the single visual (V) cue conditions from the ERP to the double audio-visual (D) cue condition. A crossmodal interaction is found if D - (A+V) > 0 and if this effect is lateralized over left and right hemisphere, as a particular activation of sensory–specific cortical areas.

Method

The method (including *participants, stimuli and materials, task and procedure*) was the same as Experiment 2 reported in the previous chapter (see p. 60).

Electrophysiological Recording

Presentation of stimuli and the emission of triggers signaling the moment and the type of stimulus was controlled by a CMO-module (version 3.7f). The triggers were received by Vision Recorder (version 1.0b, BrainProducts, GmbH), which measured participants' electroencephalographic (EEG) signals, eye movements (the horizontal and vertical electro-oculogram (EOG)) and button presses.

EEG signals were recorded from 64 Ag/AgCl ring electrodes, including 60 Standard sites from the 10-10 system (Fpz, Fp1, Fp2, AFz, AF3, AF4, AF7, AF8, Fz, F1, F2, F3, F4, F5, F6, F7, F8, FCz, FC1, FC2, FC3, FC4, FC5, FC6, FT7, FT8 C1, C2, C3, C4, C5, C6, T7, T8, CPz, CP1, CP2, CP3, CP4, CP5, CP6, TP7, TP8, Pz, P1, P2, P3, P4, P5, P6, P7, P8, POz, PO3, PO4, PO7, PO8, Oz, O1, O2). Vertical and horizontal EOG was recorded using two Ag/AgCl ring electrodes placed above and below the left eye and at the outer canthi of both eyes, respectively. Trials with amplitudes on the EOG channels exceeding 60 μV from cue onset until target offset (determined by Brain Vision Analyzer 1.03) were excluded, which implies that no trials remained with eye movements larger than approximately 3º. Electric activity was initially referenced to Cz Electrode. Impedances were kept below 5 kΩ.

Data Analysis

The EEG form each electrode site was digitized at 250 Hz (time constant of 4 s) with an amplifier bandpass of 0.04 to 100 Hz (half amplitude low- and high-frequency cut-offs, respectively) together with a 50 Hz notch filter and was stored for off-line analyses. Computerized artefact rejection was performed before signal averaging to discard epochs related to cue onset in which deviations in eye position, blinks, or muscular activity occurred. On average, about 8 % of the trials were rejected for violating of artefact criteria.

Responses faster than 100 ms (premature) or slower than 1500 ms (miss) and errors were excluded from analysis. The data were collapsed across left and right responses and side of presentation. Repeated-measures analysis of variance (ANOVA) was used to evaluate attention effects on the different ERP components. The ANOVA factors were cued condition (unimodal, crossmodal and bimodal), SOA between onset and target (200, 600), and cue validity (valid or invalid). Separate ANOVAs were conducted on amplitudes and latencies for each ERP component of interest (i.e. ipsilateral and contralateral P1, ipsilateral and contralateral N2) with factors hemisphere (left and right), cued side (left and right), and electrode (11 posterior sites). Crossmodal interaction was investigated by subtracting the ERPs to unimodal (U) and crossmodal (C) cues from the ERP to the bimodal (CU) cues. Component amplitudes were measured as peak voltage deflections within specified time intervals with respect to a 100 ms pre-stimulus baseline, at which all ERPs were aligned. These analyses were carried out at the electrode sites where the components were maximal in amplitude. Huyhn-Feldt procedures were applied to all F-tests to correct for possible violations of the sphericity assumption.

Dipole Source Analysis

Estimation of the dipolar sources of ERP components was carried out using Brain Electrical Source Analysis (BESA version 2.2). The BESA algorithm estimates the location and the orientation of multiple equivalent dipolar sources by calculating the scalp distribution that would be obtained for a given dipole model (forward solution) and comparing it to the original ERP distribution. Interactive changes in the location and in the orientation in the dipole sources lead to minimization of the residual variance (RV) between the model and the observed spatio-temporal ERP distribution (Scherg, 1990). The energy criterion of the BESA was set at 15% to reduce the interaction among dipoles. Dipoles pairs were fit sequentially over specific latency (see section below “Source Localization”) to correspond with the distinctive components in the waveform. The approach for spatio-temporal dipole modeling was, indeed, to compare the localization of source in case of bimodal, unimodal and crossmodal at the points of interest outlined by the ERPs analysis.

Results

Behavioral data

Behavioural results were previously reported in Experiment 2, Chapter 4 (see p. 61).

Electrophysiological data

Figure 5.1 and Figure 5.2 shows the ERPs elicited by unimodal, crossmodal and bimodal cues delivered at the right side and left side, respectively. The waveforms showed a typical morphology that included auditory and visual P1 and N2 components. P1 tended to appear between 130-170 ms cue onset, while N2 between 200-250 ms after cue onset.

Figure 5.3 shows the waveforms (CU - (C + U)) obtained by subtracting the ERPs to the crossmodal (C) and unimodal (U) cue from the ERP to the bimodal (CU) cue presentation for most relevant channels. The response to the bimodal cue is not simply the sum of the separate unimodal and crossmodal cue, as an enlargement was found, being most pronounced above posterior sites. Particularly, this effect tended to be specific for posterior sites (CPz, CP1, CP2, CP3, CP4, CP5, CP6, TP7, TP8, Pz, P1, P2, P3, P4, P5, P6, P7, P8, POz, PO3, PO4, PO7, PO8, Oz, O1, O2), while anterior sites showed no significant ERPs difference (i.e., CU = (C + U)), as confirmed by a T test performed on each scalp electrode (see Table 5.1).

Repeated-measures ANOVA with the factors Site (11 posterior electrodes), Hemisphere (left and right), and Cue side (left and right) demonstrated that this effect was not lateralized as a consequence of the not significant Hemisphere * Cue side interaction for both P1, [$F(1, 11) = 2.86$, $p = 0.12$] and N2, [$F(1, 11) = 1.75$, $p = 0.21$]. Thus, this effect appears not to be

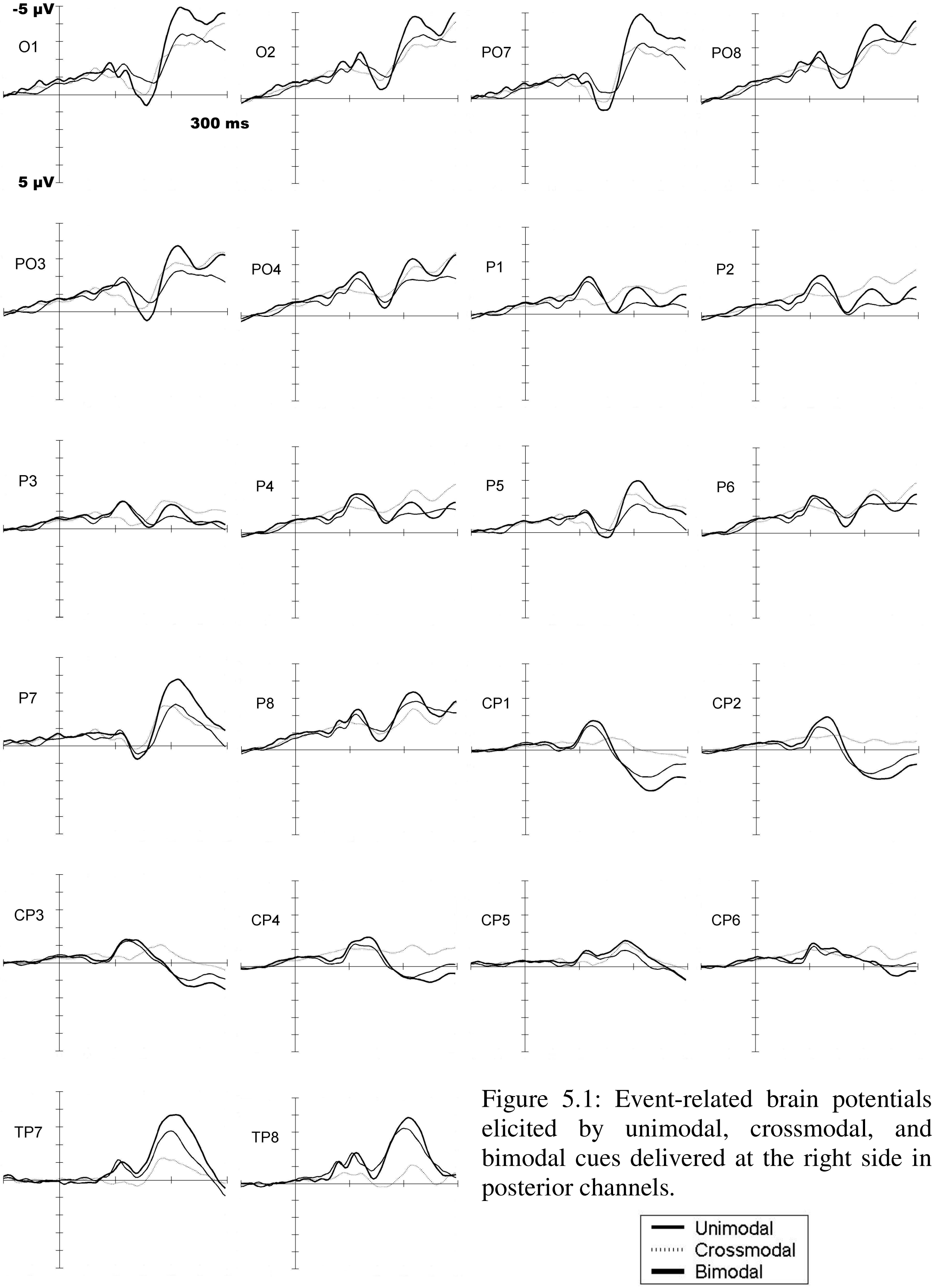


Figure 5.1: Event-related brain potentials elicited by unimodal, crossmodal, and bimodal cues delivered at the right side in posterior channels.

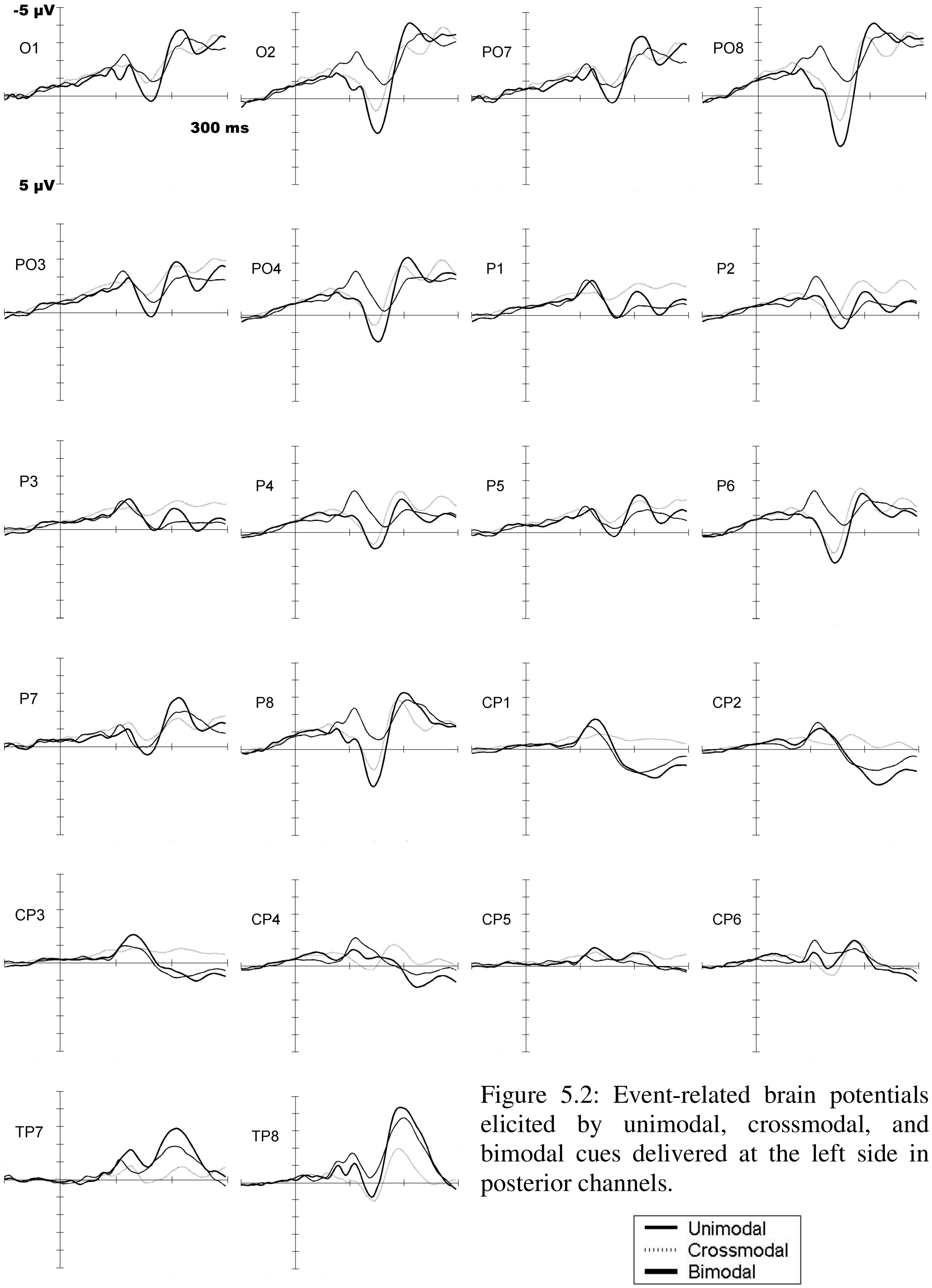


Figure 5.2: Event-related brain potentials elicited by unimodal, crossmodal, and bimodal cues delivered at the left side in posterior channels.

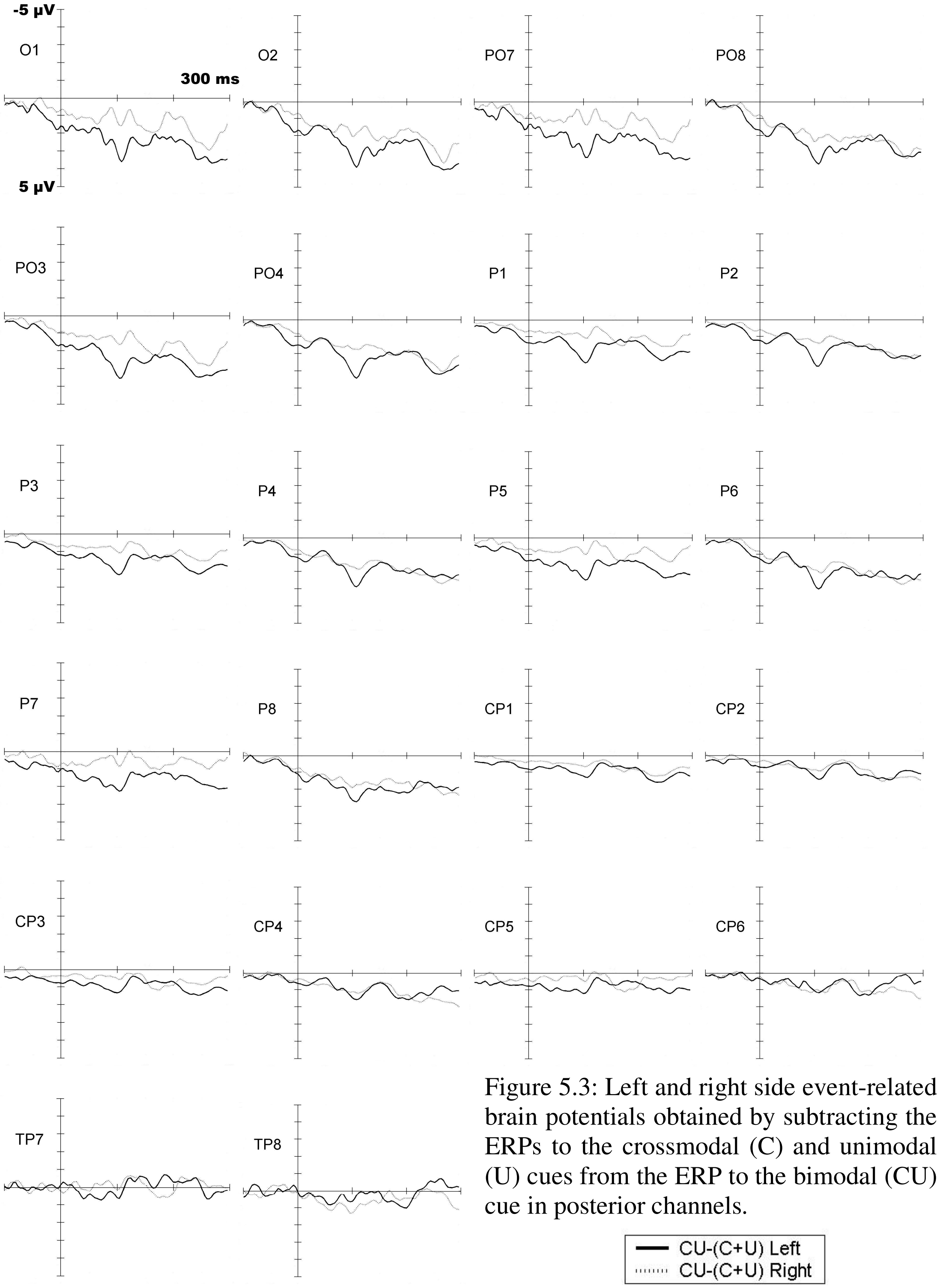


Figure 5.3: Left and right side event-related brain potentials obtained by subtracting the ERPs to the crossmodal (C) and unimodal (U) cues from the ERP to the bimodal (CU) cue in posterior channels.

	P1 (100-120 ms)				N2 (200-250 ms)			
	LEFT CUE		RIGHT CUE		LEFT CUE		RIGHT CUE	
SITES	T Value	Sig.	T Value	Sig.	T Value	Sig.	T Value	Sig.
FPZ	-4.621	.001	-2.331	.040	-3.827	.003	-2.665	.022
FZ	-4.332	.001	-3.415	.006	-2.996	.012	-2.780	.018
CPZ	4.189	.002	2.262	.045	2.192	.051	4.455	.001
POZ	6.444	.000	2.111	.059	3.939	.002	3.491	.005
F1	-4.691	.001	-3.124	.010	-2.838	.016	-2.440	.033
FC1	-3.584	.004	-1.699	.117	-0.949	.363	0.588	.568
C1	0.463	.652	0.183	.858	2.890	.015	1.727	.112
CP1	4.122	.002	3.079	.010	3.161	.009	3.408	.006
P1	5.421	.000	2.616	.024	3.405	.006	3.321	.007
FP1	-3.381	.006	-3.000	.012	-2.388	.036	-4.029	.002
AF3	-4.902	.000	-2.810	.017	-3.627	.004	-3.063	.011
F3	-5.314	.000	-2.717	.020	-2.992	.012	-2.155	.054
FC3	-4.080	.002	-2.023	.068	-1.134	.281	-0.004	.997
C3	0.202	.844	-0.416	.685	2.632	.023	1.746	.109
CP3	3.710	.003	3.275	.007	3.755	.003	3.237	.008
P3	4.627	.001	2.957	.013	3.552	.005	3.534	.005
PO3	5.800	.000	2.458	.032	4.281	.001	3.258	.008
O1	5.859	.000	1.959	.076	4.317	.001	2.236	.047
AF7	-4.807	.001	-2.824	.017	-3.783	.003	-3.351	.006
F5	-5.112	.000	-2.565	.026	-3.484	.005	-3.030	.011
FC5	-3.402	.006	-2.510	.029	-1.717	.114	-1.923	.081
C5	0.078	.939	-0.773	.456	0.659	.523	-0.100	.922
CP5	2.889	.015	1.283	.226	2.969	.013	1.384	.194
P5	4.722	.001	1.992	.072	3.607	.004	1.745	.109
PO7	4.953	.000	1.815	.097	3.532	.005	1.742	.109
F7	-4.293	.001	-1.925	.080	-3.597	.004	-3.959	.002
FT7	-1.756	.107	-2.496	.030	-2.222	.048	-3.989	.002
T7	0.024	.982	-1.616	.134	-0.354	.730	-1.398	.190
TP7	0.695	.501	-0.520	.613	-0.880	.398	-1.182	.262
P7	3.609	.004	1.218	.249	2.467	.031	0.800	.441
AFZ	-4.371	.001	-2.713	.020	-3.617	.004	-3.292	.007
FCZ	-2.483	.030	-1.317	.215	-0.691	.504	-0.270	.792
PZ	5.620	.000	2.350	.038	2.924	.014	3.647	.004
OZ	6.533	.000	1.949	.077	5.075	.000	2.198	.050
F2	-4.882	.000	-3.100	.010	-4.044	.002	-2.886	.015
FC2	-2.736	.019	-1.102	.294	-0.382	.710	-.046	.964
C2	1.480	.167	2.293	.043	1.825	.095	3.293	.007
CP2	3.779	.003	3.270	.007	2.285	.043	5.132	.000
P2	5.427	.000	2.746	.019	2.512	.029	3.724	.003
FP2	-4.876	.000	-2.022	.068	-3.794	.003	-2.624	.024
AF4	-4.849	.001	-3.529	.005	-4.248	.001	-3.534	.005
F4	-5.046	.000	-3.076	.011	-3.686	.004	-2.699	.021
FC4	-2.539	.028	-0.858	.409	-0.955	.360	-0.826	.426
C4	1.241	.240	2.363	.038	1.101	.294	4.603	.001
CP4	4.156	.002	3.440	.006	2.559	.027	5.238	.000
P4	5.332	.000	3.339	.007	2.928	.014	4.172	.002
PO4	6.359	.000	2.381	.036	3.222	.008	3.423	.006
O2	6.546	.000	2.294	.042	4.304	.001	2.458	.032
AF8	-4.298	.001	-1.118	.287	-4.459	.001	-3.032	.011
F6	-4.482	.001	-3.174	.009	-4.970	.000	-3.934	.002
FC6	-2.808	.017	-0.374	.716	-1.904	.083	-1.894	.085
C6	0.286	.781	1.443	.177	-0.024	.981	0.724	.484
CP6	2.909	.014	2.476	.031	1.970	.075	2.498	.030
P6	5.714	.000	3.357	.006	3.383	.006	3.766	.003
PO8	6.792	.000	2.803	.017	3.263	.008	2.630	.023
F8	-4.829	.001	-0.764	.461	-5.384	.000	-3.855	.003
FT8	-3.271	.007	-0.587	.569	-2.568	.026	-2.061	.064
T8	-0.715	.490	1.404	.188	-0.348	.734	-0.523	.612
TP8	0.675	.514	3.714	.003	0.152	.882	0.363	.723
P8	4.456	.001	2.481	.031	2.563	.026	2.166	.053
CZ	0.810	.435	1.568	.145	1.820	.096	2.479	.031

Table 5.1: T-test results for all sites. Test value = 0, sig. 2-tailed.

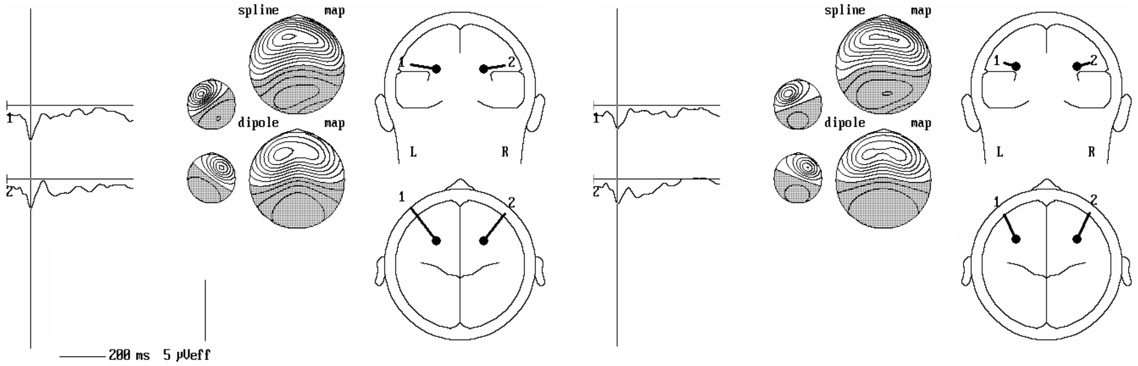
related to an attentional orienting mechanism, but to an increasing in the arousal state of participants in case of bimodal cue. Indeed, the same analysis repeated taking exclusively into account unimodal cue presentation revealed a lateralized effect for both P1 [F(1, 11) = 5.68, p = 0.036] (Hemisphere * Cue side interaction) and N2 [F(10, 110) = 2.53, p = 0.042] (Site * Hemisphere * Cue side interaction) components; similarly, in case of bimodal cue we found Hemisphere * Cue side interaction for P1 [F(1, 11) = 9.80, p = 0.010] component and Site * Hemisphere * Cue side interaction for N2 [F(10, 110) = 4.26, p = 0.012] component.

Source Localization

Figure 5.3 illustrates the results of dipole source analysis at the specific time points (112 and 192 ms) outlined from ERPs analysis. At the left of each picture (from *a* to *l*) it is possible to see the activation of the sources over time. In the middle different maps at the time points of interest are reported: upper part is the activation at that moment as indicated by spline (or isopotential) map, which gives an indication of activity across the whole head, while the figure below is the activation as explained by the dipole model. The Residual Variance (RV), which indicates the percentage of measured activity at that moment which is not explained by the model, is reported for all the pictures. Particularly, the dipole solutions were reported to the right, whereby the dot is the locus of the dipole and the tail is the length of the dipole tail, which gives an indication of its strength at that moment.

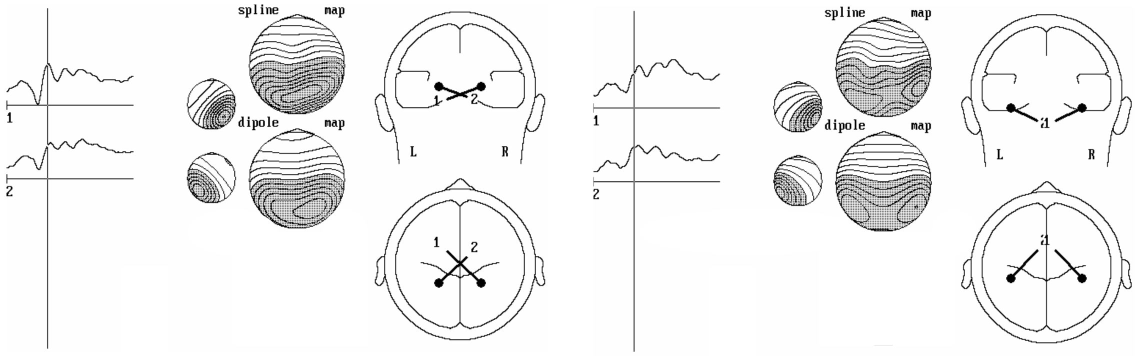
Dipole source analysis failed to individuate the crossmodal (auditory) peaks at 192 ms and the unimodal (visual) peaks at 112 ms. Thus, only the auditory peaks at about 112 ms, and the visual at 192 ms were reported. For bimodal cue and subtraction CU - (C + U) both peaks (112 and 192 ms) were found and reported. Dipole pairs were chosen because the scalp voltage and current density distributions of the components were essentially bilaterally symmetrical, indicative of mirror-image sources in the two hemispheres (Scherg, 1990). The bilateral dipoles in case of auditory cues (with mirrored location and orientation parameters) had a slightly anterior focus (Fig. 5.3, *a* and *b*). The situation was quite different in case of visual cue, wherein the bilateral dipoles had a posterior focus (Fig. 5.3, *c* and *d*).

Of most interest is that comparable locations were obtained both in the bimodal and in the single (unimodal and crossmodal) conditions. Indeed, the bilateral dipoles corresponding to peaks at 112 ms for bimodal cues had a slightly anterior focus as well as for auditory cues (Fig. 5.3, *e* and *f*); while the bilateral dipoles corresponding to peaks at 192 ms had a posterior focus that mirrored dipoles for visual cues (Fig. 5.3, *g* and *h*). Thus, neural correlates of reflexive orienting of attention towards multimodal events appear to be quite similar to those involved when spatial attention is attracted by single unimodal and crossmodal stimuli.



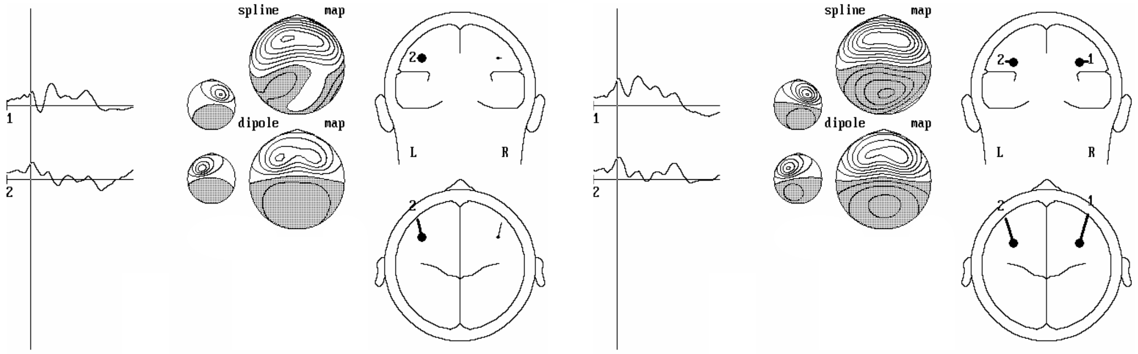
a: Auditory Cue Left; Lat = 112 ms; RV = 3.82 %

b: Auditory Cue Right; Lat = 112 ms; RV = 4.07 %



c: Visual Cue Left; Lat = 192 ms; RV = 7.3 %

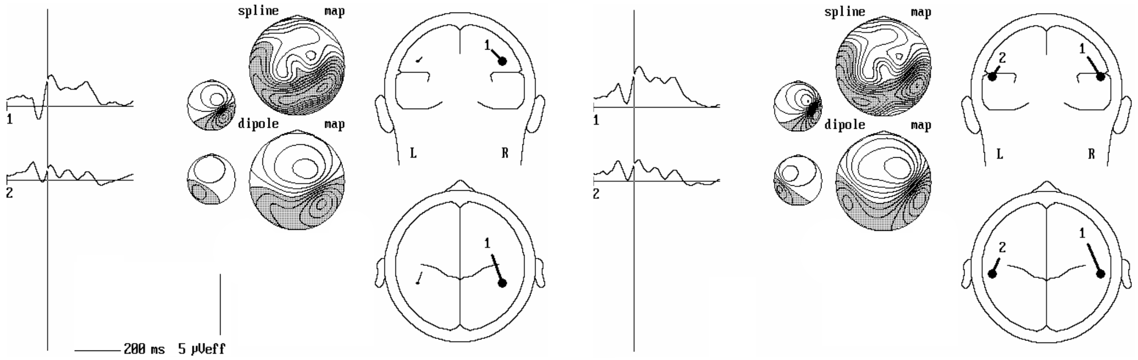
d: Visual Cue Right; Lat = 192 ms; RV = 9.68 %



e: Double Cue Left; Lat = 112 ms; RV = 9.52 %

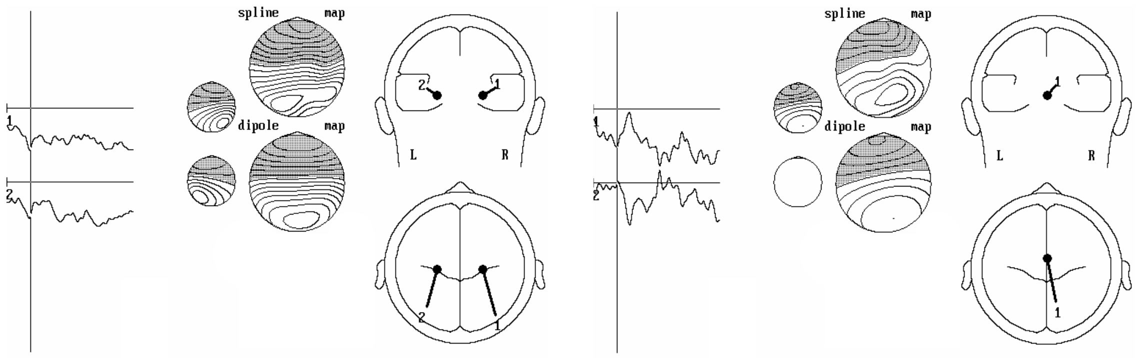
f: Double Cue Right; Lat = 112 ms; RV = 3.86 %

Figure 5.3 (*a-l*): Results of dipole source analysis at specific time points (112 and 192 ms). At the left of each picture (from *a* to *l*) it is possible to see the activation of the sources over time. In the middle different maps at the time points of interest are reported: upper part is the activation at that moment as indicated by spline (or isopotential) map, which gives an indication of activity across the whole head. The spacing, which indicates the difference in



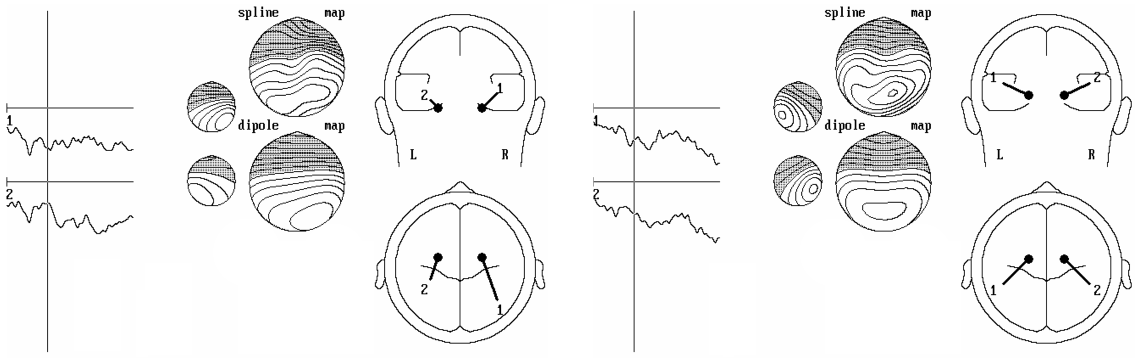
g: Double Cue Left; Lat = 192 ms; RV = 37.5 %

h: Double Cue Right; Lat = 192 ms; RV = 37.4 %



i: D – (A+V) Left; Lat = 112 ms; RV = 5 %

j: D – (A+V) Right; Lat = 112 ms; RV = 5.9 %



k: D – (A+V) Left; Lat = 192 ms; RV = 7.02 %

l: D – (A+V) Right; Lat = 192 ms; RV = 5.62 %

activity between two isopotential lines, is of 0.5 μV. Hatching is negative, non-hatching positive; the lower figure is the activation as explained by the dipole model. The Residual Variance (RV), which indicates the percentage of measured activity at that moment which is not explained by the model, is reported for pictures all pictures. To the right you see the dipole solutions, whereby the dot is locus of the dipole and the tail is the length of the dipole, which gives an indication of its strength at that moment.

The signal left after subtraction (CU - (C + U) is clearly no noise, as it was accurately described by ERP analysis (see previous section). The source analysis seems to confirm this result. Indeed, the bilateral dipoles in case of both left and right stimuli were originated from a central locus (i.e., in the middle between dipoles obtained for auditory and visual cues) and their strength appears to be high, as showed by the length of tails (Fig. 5.3, *i, j, k* and *l*).

Discussion

The aim of this study was to examine the neural basis and electrophysiological mechanisms of reflexive orienting of attention towards unimodal, crossmodal and bimodal events, and to investigate the influence of crossmodal integration on exogenous orienting. Starting from the behavioural results reported in the previous chapter, a possible scenario consisted in comparable effect between singular (both unimodal and crossmodal) and bimodal cues, indicating that either there was no crossmodal integration in exogenous orienting processes, or crossmodal integration takes place at a later processing stage which does not affect exogenous orienting. A more sensible approach with respect to behavioural measures, based on analysis of event-related brain potential and localization of dipole sources, was used in order to deeply investigate this issue.

The existence of crossmodal interactions or integration processes of audio-visual information accomplished by reflexive orienting mechanisms of attention, was investigated by subtracting ERP components of singular unimodal (U) and singular crossmodal (C) cues from the ERP components elicited by bimodal (CU) cues. Although ERP components elicited by bimodal cues were larger than the sum of ERP components elicited by single auditory and visual cues (i.e., CU - (C + U) > 0), this effect appear generated only by an increase of the arousal state inducted by the synchronous presentation of audio-visual information, rather than a particular activation of sensory–specific cortical areas. This assumption was demonstrated by the absence of significant interaction hemisphere * cue side, which clearly imply that ERP components left after subtraction (CU – (C + U)) are not lateralized over left and right hemisphere. Thus, as already pointed out in the general discussion of previous chapter (chapter 4, see p. 63), the integration of multimodal features seems to be a process completely unrelated to exogenous orienting mechanisms, as no further improvement appears to be possible when the attentional focus is reflexively attracted by an abrupt event, event though when it combines multisensorial (audio-visual) information (i.e., the orienting towards auditory or visual cues is equal to the orienting towards multimodal audio-visual cues). In other words, crossmodal integration seems to fulfil no special role for exogenous orienting as no enlarged orienting effect was found, even though components elicited by bimodal cues were larger than the sum of a single auditory (crossmodal) and visual (unimodal) stimulus.

The present study supplies also further evidence in favour of a supramodal spatial attention module, that seems to be triggered by multimodal (audio-visual) as well as unimodal (either visual or auditory) events, without showing remarkable difference in the cuing effect size. This module, underlying reflexive orienting mechanisms of attention, appears to be flexible enough to account for involuntary allocation of attentional resources toward stimuli coming from different sensorial modalities, and multimodal objects (double cue) that combines different sensorial (audio-visual) attributes.

Finally, as indicated by the analysis of dipole source localization, this supramodal spatial attention module seems to involve a circuit of fronto-parietal structures (Posner & Dehaene, 1994; Eimer & van Velzen, 2002). Indeed, a slightly frontal activation was found in case of reflexive orienting towards auditory cues, while posterior activities were pointed out in case of visual cues. The double cue presentation appeared to involve both structures activated by the single cue presentation.

CHAPTER 6 - A Proposal for a Synthesis

The theoretical framework illustrated in Chapter 1 and the empirical evidence reported in chapters from 2 to 5 are bound together into an innovative model of multimodal attention mechanisms. An experimental paradigm to test the predictions of model is proposed.

A Debate Still Alive

The debate concerning supramodal vs. modality-specific systems appears to be still a crucial point in attention research. Although any definitive word cannot be said about this central issue, it appears difficult to support again a completely independence between specific attentional systems, and in other words, a strong version of the modality-specific hypothesis (Rizzolatti, Riggio, & Sheliga, 1994). Indeed, too many evidence of spatial synergies operating between different sensorial modalities were found in last years. The separated-but-linked hypothesis received so many supports since the time it was firstly formulated (Spence & Driver, 1996) that it seems to be one of the most stable landmark as concerns the theoretical framework of crossmodal attention research. However, many other investigations occur to better understand the real degrees of separation and modalities of interaction between visual and auditory (and tactile) attention systems, or, broadly speaking, what is separated and what is linked in the mechanisms and structures of crossmodal attention.

In fact, the separated-but-linked model seems to be too general to explain in depth mechanisms involved with multimodal attention effects. We know that modality-specific attention systems are separated as far as spatial representation maps subserving such systems are concerned (Ferlazzo et al., 2002). We know that spatial synergies exist between auditory, visual and tactile systems, and, in other words, that to pay attention to a spatial cue delivered in one modality entails an enhancement of performance in detection or discrimination of a spatial congruent target in the same or different sensorial modality (Spence and Driver, 1998a, 1998b). Thus, we (partially) know which are the separated and the linked components of the process which involve selective spatial attention. However, we are faraway from understanding the real nature of these processes. Specifically, we do not know 1) how the attention system accomplish these crossmodal interactions; 2) how the allocation of attentional resources spreads from a modality-specific systems to another; 3) whether it exists a specific stage of elaboration for the allocation of crossmodal attention resources (maybe independent from the structural point of view) or whether this is an intrinsic capability of each sensorial specific attention system.

Thus, the separated-but-linked model results intriguing and fascinating from a descriptive point of view, but it is lacking in an equally strength in accounting for the attentional processes it has showed. Moreover, the separated-but-linked model appears to be largely incompatible with the amount of findings supporting the supramodal hypothesis of attention system (see for instance Eimer and van Velzen, 2002; Eimer et al., 2001).

Binding Together the Empirical Evidence

The empirical investigations conducted and reported in chapter 2 to 5 seem to perfectly reflect the state of the art depicted in the literature (Chapter 1), providing evidence which stands in apparent contrast between them. The investigation on reflexive orienting mechanisms showed a substantial similarity when spatial attention is triggered by unimodal (single cues) or multimodal events (bimodal cues), as concerns both behavioural (Chapter 4) and electrophysiological (Chapter 5) measures. This comparable effect seems to be in agreement with the existence of a supramodal spatial attention module, that operates in the allocation of attentional resources in case of both multimodal and unimodal events. This module, underlying reflexive orienting mechanisms of attention, appears to be flexible enough to account for involuntary allocation of attentional resources toward stimuli coming from different sensorial modalities, and multimodal objects (bimodal cues) that combines different sensorial (audio-visual) attributes.

A different conclusion derives from the study on endogenous attention (Chapter 2) that showed a clear distinction between single and bimodal cues. Paying attention to a bimodal cue, which engages two different sensorial modalities, produced a decrease of performance with respect to pay attention just to one modality. Given that the decrease of performance in case of bimodal cues was not so dramatic as predicted by the supramodal hypothesis, the existence of a single unique attention system seems to be ruled out. On the other hand, performance differences between single and bimodal cues seem to rule out the existence of completely separated modality-specific systems, which should accomplish visual and auditory allocation of attentional resources independently from each other, as predicted by the modality-specific hypothesis. These results lean towards the direction outlined by the separated-but-linked hypothesis, as shown by crossmodal links (faster RTs for valid than invalid trials when the visual target was preceded by auditory and bimodal cues), and by a not dramatic decrease of performance in bimodal compared with single cues.

Thus, some evidence points to a unimodal structure that operates between different modalities, while other evidence points to independent but synergic systems that are specific for each modality. Given that this uncertain situation reflects the more general state of the art in literature, the need for a new model of crossmodal attention appears unequivocally clear.

A Shared Attention Resources (ShAR) Model

To solve this issue it is possible to imagine a model in which multimodal and modality-specific processes coexist and work together to accomplish tasks based on selective spatial attention processes. The architecture of the *Shared Attention Resources* (ShAR) model is illustrated in Figure 6.1. This model essentially moves on two different planes. The first one concerns a modality-specific level in which both the visual and the auditory attention system operate. The preferential input to this level might be constituted by the top-down processes, that specify a voluntary engagement of either visual or auditory attention. Specifically, this level of processing could be involved in sustained attention tasks, wherein attention is endogenously directed and maintained over a source of interest for a relatively large amount of time.

The second plane concerns a modality-unspecific level in which several crossmodal shared resources support the activity of modality-specific attention systems when it is required by the task. The preferential input to this level might be constituted by the bottom-up processes, that specify a reflexive involvement of spatial attention towards abrupt changes in the surrounding visual or auditory scene.

More specifically, this level of processing might be constituted by different modules that could manage different kind of selective spatial attention activities, and could be used in turn by visual or auditory modality-specific systems. Particularly, one of these shared resources, the Temporal Allocator Sub-System (TASS), might be responsible for the rapid shift of attentional resources toward abrupt events, as those provided by visual or auditory exogenous cues. It is worth noting that TASS also provide a connection between visual and auditory

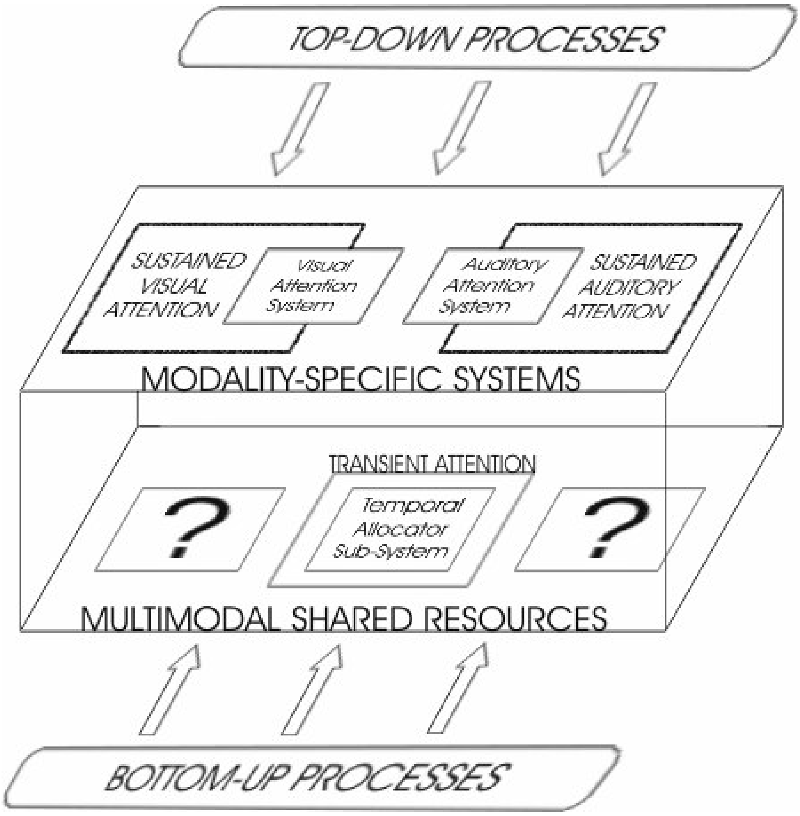


Figure 6.1: Architecture of the Shared Attention Resources (ShAR) model. It is constituted by two different planes: the first one concerns a modality-specific level in which both visual and auditory attention system operate. Its preferential input come from the top-down processes, that specify a voluntary engagement of either visual or auditory attention. This level of processing is involved in sustained attention task, wherein attention is endogenously directed and maintained over a source of interest for a relatively large amount of time; the second one concerns a modality-unspecific level in which several crossmodal shared resources support the activity of modality-specific attention systems when it is required by the task. Its preferential input come from the bottom-up processes, that specify a reflexive involvement of spatial attention towards abrupt changes in the surrounding visual or auditory scene. This level of processing is constituted by different modules (as the Temporal Allocator Sub-System, which is responsible for the shift of attentional resources toward abrupt events) that could manage different kind of selective spatial attention activities, and could be used in turn by visual or auditory modality-specific system.

attention. Indeed, streams of information may move from one to another plane by means of this unimodal shared resource. The existence of TASS is suggested by the analysis of many studies on selective spatial attention, which showed similar pattern of spatial advantages (i.e., crossmodal links, see page 22) and spatial disadvantages (i.e., inhibition of return effect, see page 26) in tasks involving different sensorial modalities. These comparable patterns of activity imply, indeed, the necessity to consider that the same mechanisms (operating according specific needs of attention system) might be responsible for facilitatory and inhibitory effects.

The architecture suggested by the ShAR model appears able to explain the empirical findings reported from chapters 2 to 5. Indeed, within the framework outlined by the ShAR model, the result reported in Chapter 2 (i.e., the decrease of performance in case of the double cue condition compared with the single cue one) might be due to the fact that the request posed by the task primary involved the modality-specific level. Participants were instructed to pay attention to the cues, which indicated with a probability of 80% the correct side in which the target would occur. Given that they were informed about this probability, the best strategy to accomplish the task was voluntarily decide to trust into the cues, which clearly involved an activation of top-down processes. The increase of RTs in case of bimodal cues could be due to the nearly-synchronous request by the visual and auditory attention systems to the TASS, in order to allocate resources on the spatial location cued by the information provided by the different modalities.

Conversely, the result reported in Chapter 4 and 5 (i.e., the comparable effect in case of reflexive orienting towards both unimodal or multimodal events) is a consequence of a primary involvement of the multimodal shared resources level. Given that both single and bimodal cues were completely uninformative, the stream of information deriving from the bottom-up processes directly involve the activation of TASS, and from that level engages visual and/or auditory attention systems in the modality-specific level.

A Simple Way to Test the ShAR Model

If crossmodal attention resources are really shared between the different modality-specific attention systems, then each resources would be available just for one system at a time. In other words, it would be impossible to use the same resources at the same time, with clear implications for the working of attention system. The TASS has been postulated to be one of the shared resources specifically involved with the spatial allocation necessary in cuing effect. If such a shared resource is busy with an auditory task, then it would not be possible to find a visual cuing effect and vice versa. Figure 6.2 shows a possible sequence of events to test the model's prediction. Participants should be required to stare a central fixation point, avoiding to move their eyes. After a short interval (with range 800-1200 ms), a square equally probable on the left or right side, and a sound of 500 Hz from a central location should be presented at the same time. They should be followed by either a sound of 1000 Hz (on 75% of

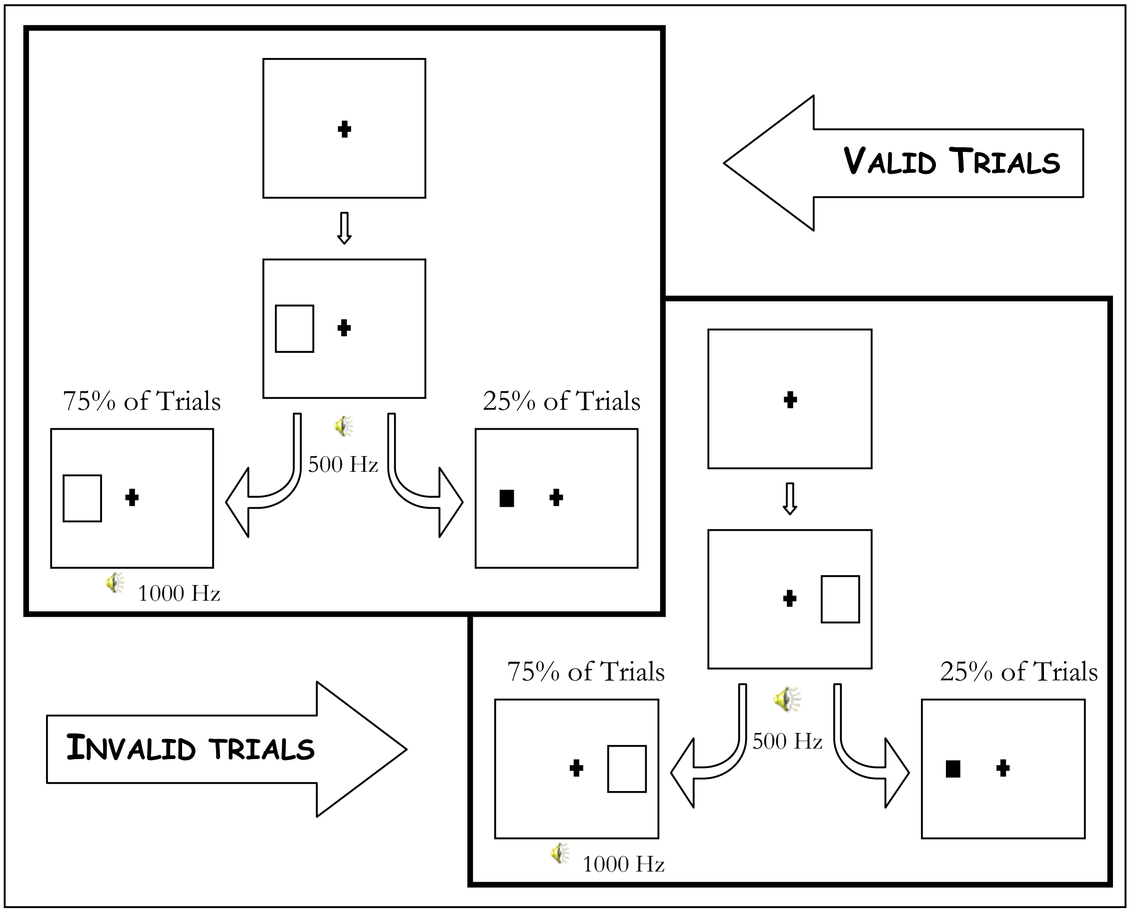


Figure 6.2: Sequence of events to test the ShAR model. Participants are required to stare a central fixation point (+), avoiding to move their eyes. After a short interval (with range 800-1200 ms), a square equally probable on the left or right side, and a sound of 500 Hz from a central location are presented at the same time. They are followed by either a sound of 1000 Hz (on 75% of all the trials) or a smaller (black) square (on the remaining 25% of trials) that can match the previous square location (valid trial) or not (invalid trial). Participants are informed about the probabilities of above-mentioned conditions and they are required to press as quickly and accurately as possible a response button whenever they hear the 1000 Hz sound or whenever they see the black square.

all the trials) or a smaller (black) square (on the remaining 25% of trials) that can match the previous square location (valid trial) or not (invalid trial). Participants should be informed about the probabilities of above-mentioned conditions and required to press as quickly and accurately as possible a response button whenever they hear the 1000 Hz sound or whenever they see the black square. Within this arrangement, the TASS should be busy with the auditory modality, as most of the trials (75%) participants are required to detect a switch of sound from 500 to 1000 Hz. Thus, no attentional resources should be available for the request submitted to the TASS by the visual detection task (i.e., to press for the black square).

At the moment I am using this paradigm to verify the above-mentioned hypothesis in a research in progress. Thus, at short time it should be possible to have relevant data concerning the empirical plausibility of the ShAR model.

Conclusion

The ShAR model provides a new theoretical framework for selective spatial attention processes, substantially contributing to the development of this crucial domain of human cognition. Review of literature (Chapter 1) essentially showed the lack of a comprehensive model of crossmodal attention that unifies empirical evidence in a structured and solid theoretical framework. Indeed, most of the hypotheses formulated in the literature appear to be in contrast between them (e.g., supramodal vs. modality-specific vs. separated-but-linked hypothesis of attention), and the conflicts generated by the alternative positions seems to be irresolvable by a means of the previously proposed models. Thus, the need of a model able to coherently integrate as much evidence as possible appears to strongly emerge from literature inspection. The ShAR model with its distinctive features and capabilities to integrate previous and current findings (i.e. the data reported in the present work) might represent an interesting solution in the debate on crossmodal mechanisms of attention.

In order to shed light on multimodal mechanisms of attention and to develop the ShAR model both behavioural and electrophysiological investigations were conducted. Chapter 2 examined the real independence of visual and auditory attention systems using at the same time both visual and auditory cues. The rationale was that if the attention systems are quite independent from each other (as postulated by the modality-specific hypothesis, Rizzolatti, Riggio, & Sheliga, 1994), then no performance differences should exist between a condition in wherein the attentional load engage one modality (control condition) and another wherein the attentional load engaged two modalities (experimental condition). On the other hand, if the attention system is supramodal (Farah et al., 1989; Eimer & van Velzen, 2002), then to split attentional resources over two modalities should be a disaster as concerns the performance to experimental condition with respect to the control condition. Finally, if SBLH is true, then performance differences should exist even though less remarkable than in the previous case (i.e., without a sharp decrease in the experimental condition). In other words, if modality-specific attention systems are separated but strongly linked from a functional point of view, then it should be expected a significant decrease in experimental condition performance with respect to control condition performance, but also enough independence so that performance does not collapse as if the system was supramodal. Results showed little costs in terms of RTs and no costs in terms of errors when participants pay attention to both visual and auditory cues in both exogenous and endogenous conditions, leaning toward the direction outlined by the separated-but-linked hypothesis (Spence and Driver, 1996). In other words, participants did not find so difficult to attend and use information derived from crossmodally separated cues, which should be impossible for a single unique attention system. At the same time, this data provide at least two basic specifications of the separated-but-linked hypothesis: first, the time needed to allocate attentional resources on different spatial locations is function of the attentional load demanded by specific tasks, as showed by the increase of RTs in the double cued condition; second, the attention system is able to use spatial information derived from crossmodally different (audio-visual) cues in order to allocate attentional resources (as showed by the validity effect we found in double cued conditions in both experiments) in the elevation-discrimination task.

The study reported in Chapter 3 tried to establish the time course of audio-visual integration by means of bimodal cues delivered at different intervals of milliseconds between them. The rationale was that if the information provided by the visual cue and the auditory cue could be bound together and used for the spatial discrimination task in a narrow time window, then the longer was the SOA (i.e., the longer was the temporal distance between the visual and the auditory cue) the longer were the RTs, with a consequent decrease of performance. Otherwise, if the process of audio-visual integration could be accomplished without substantial differences through time, then the cuing effect should be stable for each level of SOA. Results showed that performance remains good till 400 ms of SOA. No substantial differences were found, indeed, between 0, 200 and 400 ms of SOA. When SOA grew larger, however, RTs sharply decreased, indicating that the second (auditory) cues might be too close to the target to be used in the spatial discrimination task (as showed by the disruption of cuing effects at 600 and 800 ms of SOA).

Evidence for a supramodal attention system was achieved by the studies reported in Chapter 4 and Chapter 5. They were specifically addressed at examining whether spatial attention triggered by multimodal events acts differently as compared to unimodal events. A comparison between a condition wherein a visual target was preceded by both visual and auditory exogenous cues delivered together at the same side (bimodal cue), with conditions wherein the visual target was preceded by either the visual (unimodal) or the auditory (crossmodal) cue was accomplished. The rationale was to understand whether orienting attentional mechanisms take advantage when the trigger is a multimodal event or not. Behavioural measures (RTs and PCs) showed a substantial overlap between performance in single and double cue condition (Chapter 4). Similarly, single (both unimodal and crossmodal) and bimodal cue conditions elicited comparable ERP components: the only difference appeared to be an increase in the arousal state in double compared with single cue condition, an increase not related to the activity of sensorial-specific areas. No enlarged orienting effect was found, indeed, showing as well that crossmodal integration fulfils no special role for exogenous orienting (Chapter 5). Thus, the comparable effects found in these two studies are a clear evidence in favour of a modality-unspecific spatial attention module, that operates in the allocation of attentional resources in case of both multimodal and unimodal events, without showing remarkable differences.

Chapter 6 was directly addressed at operating a synthesis and an integration of the empirical evidence reported in previous chapters. This was achieved by a means of the ShAR model (see picture 5.1, pp. 81), which essentially moves through different planes. The first one concerns a modality-specific level in which both visual and auditory attention systems operate. The preferential input to this level is constituted by the top-down processes, that specify a voluntary engagement of either visual or auditory attention. Specifically, this level of processing could be involved in sustained attention task, wherein attention is endogenously directed and maintained over a source of interest for a relatively large amount of time. The second plane concerns a modality-unspecific level in which several crossmodal shared resources support the activity of modality-specific attention systems when it is required by the task. The preferential input to this level is constituted by the bottom-up processes, that specify a reflexive involvement of spatial attention towards abrupt changes in the surrounding visual or auditory scene. More specifically, this level of processing is constituted by different modules that could manage different kind of selective spatial attention activities, and could be used in turn by visual or auditory modality-specific system. Particularly, one of these shared resources, the Temporal Allocator Sub-System, might be responsible for the rapid shift of attentional resources toward abrupt events, as those provided by visual or auditory exogenous cues. As showed in the previous chapter, the architecture suggested by the ShAR model appears able on one side to explain the empirical findings reported from chapter 2 to chapter 4, and on the other side to integrate most of the evidence present into the literature, providing a new framework for the research on crossmodal mechanisms of attention.

References

Aitkin, L. (1986). *The auditory midbrain.* Clifton, NJ: Humana.

Alain, C., & Woods, D. L. (1993). Distractor clustering enhances detection speed and accuracy during selective listening. *Perception & Psychophysics, 54*(4), 509-514.

Andersen, R. A., Snyder, L.H., Bradley, D.C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience, 20*, 303-330.

Baltz, G. W., & Hock, H. S. (1997). The effect of attentional spread on spatial resolution. *Vision Research, 37*, 1499-1510.

Bedard, M. A., Massioui, F. E., Pillon, B., & Nandrino, J. L. (1993). Time for reorienting of attention: A premotor hypothesis of the underlying mechanism. *Neuropsychologia, 31*(3), 241-249.

Bertelson, P. (1999). Ventriloquism: a case of crossmodal perceptual grouping. In G. Ashersleben, T. Bachmann, & J. Müsseler, ed., *Cognitive Contribution to the Perception of Spatial and Temporal Events* (pp. 347-362). Amsterdam: Elsevier Science, B.V.

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review, 64(3)*, 205-215.

Bloch, M. I., & Hellige, J. B. (1989). Stimulus intensity, attentional instructions, and the ear advantage during dichotic listening. *Brain and Cognition, 9*, 136-148.

Bregman, A. S. (1990). *Auditory scene analisys*. Cambridge, MA: MIT Press.

Bryden, M. P. (1978). Strategy effects in the assessment of hemispheric asymmetry. *Strategies of information processing.* G. Underwood ed., (pp. 117-149). London: Academic Press.

Buchtel, H. A., & Butter Charles M. (1988). Spatial attentional shifts: Implications for the role of polysensory mechanisms. *Neuropsychologia, 26*(4), 499-509.

Butter, C. M., Buchtel, H. A., & Santucci, R. (1989). Spatial attention shifts: further evidence for the role of polysensory mechanisms using visual and tactile stimuli. *Neuropsychologia, 27*(10), 1231-1240.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodroff, P. W., Iversen, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*: 593-596.

Cameron, E. L., Tai, J. C., & Carrasco, M. (2002). Covert attention affects the psychometric function of contrast sensitivity. *Vision Research, 42*, 949–967.

Carrasco, M., Giordano, A. M., & McElree, B. (2004). Temporal performance fields: visual and attentional factors. *Vision Research, 44*, 1351-1365.

Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Science, 98*, 5363–5367.

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision Research, 40*, 1203-1215.

Carrasco, M., Talgar, C. P., & Cameron, E. L. (2001). Characterizing visual performance fields: effects of transient covert attention, spatial frequency, eccentricity, task and set size. Spatial Vision, 15, 61–75.

Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: support for signal enhancement. *Journal of Vision, 2*, 467–479.

Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 673-692.

Christ, S. E., McRae, C. S., & Abrams, R. A. (2002). Inhibition of return in static and dynamic displays. *Psychonomic Bulletin & Review, 9*, 80-85.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, Mass: MIT Press.

Ciesielski, K. T., Knight, J. E., Prince, R. J., Harris, R. J., & Handmaker, S. D. (1995). Event-related potentials in crossmodal divided attention in autism. *Neuropsychologia, 33*(2), 225-246.

deCharms, R. C., & Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature, 381*, 610-613.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222.

Deutsch, D. (1974). An auditory illusion. *Journal of the Acoustical Society of America, 55* , 518-519.

Deutsch, J. A., & Deutsch, D. (1963). Attention: some theoretical considerations. *Psychological Review, 70*, 80-90.

Driver, J., & Spence, C. (1998a). Attention and crossmodal construction of space. *Trends in Cognitive Science, 2*(7), 254-262.

Driver, J., & Spence, C. (1998b). Crossmodal links in spatial attention. *Proc. R. Soc. London Ser., B 353*, 1-13.

Driver, J., & Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current Biology, 10*, R731-R735.

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

Eimer, M. (1999). Can attention be directed to opposite locations in different modalities? An ERP study. *Clinical Neurophysiology, 110*, 1252-1259.

Eimer, M., Cockburn, D., Smedley, B., & Driver, J. (2001). Cross-modal links in endogenous spatial attention are mediated by common external locations: evidence from event-related brain potentials. *Experimental Brain Research, 139*, 398-411.

Eimer, M., & Driver, J. (2001). Crossmodal links in endogenous and exogenous spatial attention: evidence from event-related brain potential studies. *Neuroscience and Biobehavioral Reviews, 25*(6), 497-511.

Eimer, M., & Van Velzen, J. (2002). Crossmodal links in spatial attention are mediated by supramodal control processes: evidence from event-related potentials. *Psychophysiology, 39*(4), 437-449.

Eimer, M., Van Velzen, J., & Driver, J. (2002). Cross-modal interactions between audition, touch, and vision in endogenous spatial attention: ERP evidence on preparatory states and sensory modulations. *Journal of Cognitive Neuroscience, 14*(2), 254-271.

Farah, M. J., Wong, A. B., Monheit, M. A., & Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal? *Neuropsychologia, 27*(4), 461-470.

Ferlazzo, F., Couyoumdjian, A., Padovani, T., & Olivetti Belardinelli, M. (2002). Head-centred meridian effect on auditory spatial attention orienting. *The Quarterly Journal of Experimental Psychology A, 55*(3), 937-963.

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

Graziano, M. S. A., & Gross, C. G. (1996). Multiple pathways for processing visual space. *Attention and performance: information integration in perception and communication* (Vol. 26, pp. 181-207). MIT Press.

Graziano, M. S. A., & Gross, C. G. (1998). Spatial maps for the control of movement. *Current Opinion in Neurobiology, 8*, 195-201.

Heffner, R. S., & Heffner, H. E. (1992). Visual factors in sound localization in mammals. *Journal of Comparative Neurobiology, 317*, 219-232.

Henderson, J.M., & Macquistan, A.D. (1993). The spatial distribution of attention following an exogenous cue. *Perception & Psychophysics, 53,* 221-230.

Henry, R. G. J. (1979). Monaural studies eliciting an hemispheric asymmetry: a bibliography. *Perceptual and Motor Skills, 48*, 335-338.

Henry, R. G. J. (1983). Monaural studies eliciting an hemispheric asymmetry: a bibliography II. *Perceptual and Motor Skills, 56*, 915-918.

Hiscock, M., Inch, R., & Kinsbourne, M. (1999a). Allocation of attention in dichotic listening: differential effects on the detection and localization of signals. *Neuropsychology, 13*(3), 404-414.

Hiscock, M., Inch, R., & Kinsbourne, M. (1999b). Allocation of attention in dichotic listening: Effects on the detection and localization of targets within lists. *Journal of Clinical and Experimental Neuropsychology, 21*(2), 265-278.

Hiscock, M., & Stewart, C. (1984). The effect of asimmetrically focused attention upon subsequent ear differences in dichotic listening. *Neuropsychologia, 22*(3), 337-351.

Hopfinger, J. B., & Mangun, G. R. (1998). Reflexive attention modulates processing of visual stimuli in human extrastriate cortex. *Psychological Science, 9(6)*, 441-447.

Hugdahl, K., & Andersson, L. (1986). The "forced-attention paradigm" in dichotic listening to CV-syllables: A comparison between adults and children. *Cortex, 22*, 417-432.

Johnson, P. R. (1977). Dichotically-stimulated ear differences in musicians and nonmusicians. *Cortex, 13*, 385-389.

Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., & Klapp, S. T. (1995). Tests of attentional flexibility in listening to polyrhythmic patterns. *Journal of Experimental Psychology: Human Perception and Performance, 21*(2), 293-307.

Kennett, S., Eimer, M., Spence, C., & Driver, J. (2000). Tactile-visual links in exogenous spatial attention under different postures: convergent evidence from psychophysics and ERPs. *Journal of Cognitive Neuroscience, 13*(4): 462-478.

Kimura, D. (1967). Functional asymmetry of the brain in the dichotic listening. *Cortex, 3*, 163-178.

Kinchla, R. A. (1992). Attention. *Annual Review of Psychology, 43*, 771-42.

Klein, R., Brennan, M., & Gilani, A. (1987). Covert cross-modality orienting of attention in space. *Paper Presented at the Annual Meeting of the Psychonomic Society, Seattle, WA.*

Kuyk, T., & Niculescu, D. (2001). The psychophysical periphery effect crosses the vertical meridian. *Visual Neuroscience, 18(4)*, 657-661.

Lloyd, D. M., Merat, N., McGlone, F., & Spence, C. (2003). Crossmodal links between audition and touch in covert endogenous spatial attention. *Perception & Psychophysics, 65(6)*, 901-924.

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research, 38*, 1183-1198.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in area V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology, 77*, 24-42.

Macaluso, E., Frith, C., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science, 289*: 1206-1208.

Maldonado, P. E., & Gerstein, G. L. (1996). Neuronal assembly dynamics in the rat auditory cortex during reorganization induced by intracortical microstimulation. *Experimental Brain Research, 112*, 431-441.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 1057-1074.

McDonald, J. J. , Teder-Sälejärvi, 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-Sälejärvi, W.A., & Ward, L.M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science, 292,* 1791.

McGurk, H., & McDonald, J. J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schoroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research, 14*, 115-128.

Mondor, T. A. (1999). Predictability of the cue-target relation and the time-course of auditory inhibition of return. *Perception & Psychophysics, 61*(8), 1501-1509.

Mondor, T. A., & Amirault, K. J. (1998). Effects 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.

Mondor, T. A., & Breau, L. M. (1999). Facilitative and inhibitory effects of location and frequency cues: Evidence of a modulation in perceptual sensitivity. *Perception & Psychophysics, 61*(3), 438-444.

Mondor, T. A., Breau, L. M., & Milliken, B. (1998). Inhibitory processes in auditory selective attention: Evidence of location-based and frequency-based inhibition of return. *Perception & Psychophysics, 60*(2), 296-302.

Mondor, T. A., & Bregman, A. S. (1994). Allocating attention to frequency regions. *Perception & Psychophysics, 56*(3), 268-276.

Mondor, T. A., & Bryden, M. P. (1991). The influence of attention on the dichotic REA. *Neuropsychologia, 29*(12), 1179-1190.

Mondor, T. A., & Bryden, M. P. (1992a). On the relation between auditory spatial attention and auditory perceptual asymmetries. *Perception & Psychophysics, 52*(4), 393-402.

Mondor, T. A., & Bryden, M. P. (1992b). Orienting of auditory spatial attention: effects of a lateralized tone cue. *Neuropsychologia, 30*(8), 743-752.

Mondor, T. A., & Lacey, T. E. (2001). Facilitative and inhibitory effects of cuing sound duration, intensity, and timbre. *Perception & Psychophysics, 63*(4), 726-736.

Mondor, T. A., & Terrio, N. A. (1998). Mechanisms of perceptual organization and auditory selective attention: The role of pattern structure. *Journal of Experimental Psychology: Human Perception and Performance, 24*(6), 1628-1641.

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and occusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*(2), 387-409.

Mondor, T. A., Zatorre, R. J., & Terrio, N. A. (1998). Constrains on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 66-79.

Morgan, M. J., Ward, R. M., & Castet, E. (1998). Visual search for a tilted target: tests of spatial uncertainty models. *The Quarterly Journal of Experimental Psychology A, 51*, 343-370.

Murray, J., Allard, F., & Bryden, M. P. (1988). Expectancy effects: cost-benefit analysis of monoaurally and dichotically presented speech. *Brain and Language, 35*, 105-118.

Niemi, P. (1981). Constant vs variable stimulus intensity and visual simple reaction time. *Perceptual Motor Skill, 53(2)*, 615-619.

Norman, D. A. (1968). Towards a theory of memory and attention. *Psychological Review, 75*, 522-536.

Olivetti Belardinelli, M., & Santangelo, V. (in press). The head-centered meridian effect: auditory attention orienting in conditions of impaired visual spatial information. *Disability and Rehabilitation.*

Pillow, J., & Rubin, N. (2002). Perceptual completion across the vertical meridian and the role of early visual cortex. *Neuron, 33(5)*, 805-813.

Pizzamiglio, L., De Pascalis, C., & Vignati, A. (1974). Stability of dichotic listening test. *Cortex, 10*, 203-205.

Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

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

Posner, M. I., & Cohen, Y. (1980). Consequences of visual orienting. *Paper presented at the annual meeting of the Psychonomic Society* St. Louis.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma and D. Bowhuis ed., *Attention and performance Vol. 10* (pp. 531-556). Hillsdale, NJ: Erlbaum.

Posner, M. I., Davidson, B. J., & Nissen, M. J. (1976). The process of stimulus detection. *Paper presented at the annual meeting of the Psychonomic Society* St. Louis.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in Neuroscience, 17*, 75-79.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25-442.

Posner, M. I., Rafal, R. D., Choate, L., & Vaughan, J. (1985). Inhibition of return: neural basis and function. *Cognitive Neuropsychology, 2*, 211-228.

Quinlan, P. T., & Bailey, P. J. (1995). An examination of attentional control in auditory modality: futher evidenve for auditory orienting. *Perception & Psychophysics, 57*(5), 614-628.

Rafal, R. D., Calabresi, P. A., Brennan, C. W., & Sciolto, T. K. (1989). Saccade preparation inhibits reorienting to recently attended locations. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 673-685.

Rafal, R. D., Egly, M., & Rhodes, G. (1994). Effects of inhibition of return on voluntary and visually guided saccades. *Canadian Journal of Experimental Psychology, 48*(2), 284-300.

Reuter-Lorenz, P. A., & Rosenquist, J. N. (1996). Auditory cues and inhibition of return: The importance of oculomotor activation. *Experimental Brain Research, 112*, 119-126.

Rhodes, G. (1987). Auditory attention and the representation of spatial information. *Perception & Psychophysics, 42*(1), 1-14.

Rizzolatti, G., & Camarda, R. (1987). Neural circuits for spatial attention and unilateral neglect. In M. Jeannerod ed., *Neurophysiological and neuropsychological aspects of spatial neglect* (pp. 289-313). Amsterdam: North-Holland.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia, 25*, 31-40.

Rizzolatti, G., Riggio, L., & Sheliga, B. M. (1994). Space and selective attention. In C. Umiltà, M. Moscovitch ed., *Attention and performance, XV. Conscious and nonconscious information processing* (pp. 231-265). Cambridge, Mass: Mit Press.

Röder, B., Spence, C., & Rösler, F. (2000). Inhibition of return and oculomotor control in the blind. *NeuroReport, 11(13)*, 3043-3045.

Rorden, C., & Driver, J. (1999). Does auditory attention shift in the direction of an upcoming saccade? *Neuropsychologia, 37*(3), 357-377.

Rorden, C., & Driver, J. (2001). Spatial deployment of attention within and cross hemifields in an auditory task. *Experimental Brain Research, 137*, 487-496.

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophisics, 54*, 406-416.

Santangelo, V., & Olivetti Belardinelli, M. (2002). Spatial modeling: Simulating place cells activity by an action potential timing network. *Cognitive processing, 3(1-2)*, 123-129.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Perception & Psychophysics, 42*(3), 215-223.

Scherg, M. (1990). Fundamentals of dipole source potential analysis. In F. Grandori, M. Hoke, G. L. Romani ed., *Advances in Audiology, Vol. 5*, (pp. 40-69). Karger, Basel.

Schmidt, W. C. (1996). 'Inhibition of return' without visual input. *Neuropsychologia, 34*(10), 943-952.

Schmitt, M., Postma, A., & De Haan, E. (2000). Interactions between exogenous auditory and visual spatial attention. *The Quarterly Journal of Experimental Psychology A, 53*(1), 105-130.

Simon, J. R. (1967). Ear preference in a simple reaction time task. *Journal of Experimental Psychology, 75*, 49-55.

Sparks, R., Goodglass, H., & Nickel, L. B. (1970). Ipsilateral vs. controlateral extinction in dichotic listening resulting from hemispheric lesions. *Cortex, 6*, 249-260.

Spence, C., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanism. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 355-374.

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

Spence, C., & Driver, J. (1997b). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics, 59*(3), 389-403.

Spence, C., & Driver, J. (1998). Auditory and audiovisual inhibition of return. *Perception & Psychophysics, 60*(1), 125-139.

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *NeuroReport, 11*, 2057-2061.

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

Stein, B.E., Meredith, M.A., & Wallace, M.T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Progress in Brain Research, 95,* 79-90.

Stenberg, S. (1969). The discovety of processing stages: extensions of Donders' method. In W. J. Koster (Eds.), *Attention and Performance II. Acta Psychologica, 30,* 276-315.

Stenberg, S. (1984). Stage models of mental processing and the additive-factor method. *Behavioral and Brain Sciences, 7,* 82-84.

Stenberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica, 106,* 147-246.

Talgar, C. P., & Carrasco, M. (2002). Vertical meridian asymmetry in spatial resolution: visual and attentional factors. *Psychonomic Bulletin and Review, 9(4)*, 714–722.

Tassinari, G., Aglioti, S., Chelazzi, L., Marzi, C. A., & Berlucchi, G. (1987). Distribution in the visual field of the costs of voluntarily allocated attention and of the inhibitory after-effects of covert orienting. *Neuropsychologia, 25*, 55-71.

Tassinari, G., & Berlucchi, G. (1995). Covert orienting to non-informative cue: reaction time studies. *Behavioural Brain Research, 71*, 101-112.

Tassinari, G., Biscaldi, M., Marzi, C. A., & Berlucchi, G. (1989). Ipsilateral inhibition and controlateral facilitation of simple reaction time to non-foveal visual targets from non-informative visual cues. *Acta Psychologica, 70*, 267-291.

Tassinari, G., & Campara, D. (1996). Consequences of covert orienting to non-informative stimuli of different modalities: A unitary mechanism? *Neuropsychologia, 34*(3), 235-245.

Teder-Sälejärvi, W. A., McDonald, J. J. Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research, 14*, 106-114.

Terry, K. M., Valdes, L. A., & Neill, W. T. (1994). Does "inhibition of return" occur in discrimination tasks? *Perception & Psychophysics, 55*(3), 279-286.

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*, 242-248.

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

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence for search asymmetries. *Psychological Review, 95*, 15-48.

Van der Lubbe, R.H.J., Vogel, R.O., & Postma, A. (apr). Different effects of exogenous cues in a visual detection and discrimination task: Delayed attention withdrawal and/or speeded motor inhibition?. *Journal of Cognitive Neuroscience.*

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

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on perception. *Journal of Experimental Psychology: Human Perception and Performance, 26*(5): 1583-1590.

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.

Wang, D. L., & Brown, G. J. (1999). Separation of speech from interfering sounds based on oscillatory correlation. *IEEE Transactions on Neural Networks, 10*(3), 684-697.

Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus -driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology, 48*(2), 242-259.

Wood, S., Hiscock, M., & Widrig, M. (2000). Selective attention fails to alter the dichotic listening lag effect: Evidence that the lag effect is preattentional. *Brain and Language, 71*, 373-390.

Woods, D. L., Alain, C., Diaz, R., Rhodes, D., & Ogawa, K. H. (2001). Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 65-74.

Woods, D. L., Alho, K., & Algazi, A. (1994). Stages of auditory feature conjunction: An event-related brain potential study. *Journal of Experimental Psychology: Human Perception and Performance, 20*(1), 81-94.

Wrigley, S. N., & Brown, G. J. (2001). A neural oscillator model of auditory attention. In G. Dorffner, H. Bischop, & K. Hornik, ed., *ICANN 2001, LNCS 2130* (pp. 1163-1170). Berlin, Heidelberg: Springer-Verlag.

Wrigley, S. N., & Brown, G. J. (2000). A neural oscillator model of auditory selective attention. In T. G. Dietterich, S. Becker, & Z. Ghahramani, ed., *Advances in neural information processing systems 14:* MIT Press.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature, 396*, 72-75.

Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research, 39*, 293-306.

Yeshurun, Y., & Carrasco, M. (2000). The locus of attentional effects in texture segmentation. *Nature Neuroscience, 3*(6), 622-627.

Zahn, T. P., Izak, R., Trott, K., & Paschke, P. (1998). A paced analog silicon model of auditory attention. In L. S. Smith, & A. Hamilton, ed., *Neuromorphic systems: Engineering silicon from neurobiology: 1st European workshop on neuromorphic systems* (pp. 99-112). Singapore: World scientific publishing.

Zatorre, R. J., Mondor, T. A., & Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *Neuroimage, 10*, 544-554.

Acknowledgements

During my Ph.D. I was fortunate to have the advice of several colleagues and friends who helped me and offered suggestions. They were Demis Basso, Anna Bazzucchini, Cristiana Brunetti, Riccardo Brunetti, Maddalena Chiorazzo, Antonio Cortese, Anna Damiani, Alessandro D'ausilio, Franco Delogu, Loredana Dinapoli, Sabrina Fagioli, Debora Jeric, Alessandro Londei, Simone Macrì, Alessandra Padova, Massimiliano Palmiero, Emanuele Pasqualotto, Augusto Petacchi, Laura Pieroni, Emiliano Sabadello, Stefano Sdoia, Carlo Sestieri, Gemma Todrani. Although no words are enough, I thank them all for the gift of time, knowledge, and wisdom.

I wish to express a truthful thank to Albert Postma and the *Neurocognition Space Group* (Helmholtz Research Institute, Psychonomic Department, University of Utrecht) to have given me the possibility to work with them on mutual research interests, and to setup and conduct several experiments on crossmodal spatial attention and integration, in the six months I stayed in their Psychological Laboratory. Particularly, I wish to thank Rob van der Lubbe. I learned immensely from the friendly debates his comments often prompted. I thank him for the patience with which he read different version of chapters 4 and 5 and helped me to improve them.

I am grateful to Dunjia Fluijt, Erika van Hell, Marieke Lansbergen, Matthijs Noordzij, Styrmir Sævarsson, and Alain Wouterlood for their helpful collaboration in data collection. I am grateful as well to Michela Santangelo for her precious and careful collaboration in graphical composition of pictures and tables. I am not less grateful to Marta Olivetti Belardinelli for what she have taught me through the years of my Ph.D. and for the fruitful help she gave me in the preparation of the manuscript.

My family helped me to sustain and overcome every difficulty occurred during these last years. I would not even try to thank them for their support.