UNIVERSITÀ DEGLI STUDI
DI TRENTO

CIMeC - Center for Mind/Brain Sciences

DOCTORAL SCHOOL IN COGNITIVE AND BRAIN SCIENCES XXIV CYCLE

Spatiotemporal aspects in audiovisual interaction

Thesis submitted for the degree of Doctor of Philosophy by

Stefano Targher

Supervisor: Prof. Massimiliano Zampini

ACADEMIC YEAR 2010/2011

Acknowledgements

First and foremost, I wish to thank my supervisors Massimiliano Zampini for his patience and support provided during my PhD and for the tenacious work on revising the present dissertation.

I would like to thanks also Professors David Melcher and Francesco Pavani for their collaboration and suggestion in two of the presented studies. I particularly wish to thank my colleague Alessio Fracasso for his help, advices and fruitful collaboration for my academic work and for the achievement of both the studies conducted with Professors Melcher and Pavani. I wish to thank Valeria Occelli for her collaboration and support with the studies on low vision.

Many thanks also to Angelika Lignau for her help and for the project we conduct together, to Silvia Savazzi and Leonardo Ricci for the data analysis suggestions and to Massimo Vescovi for his help with Matlab and all the hardware I used for my studies.

A particular thanks to all the low vision patients that took part to the studies and all the staff of the Italian Union of the Blind and Visually Impaired (UIC), the local section of the Institute for Research Training and Rehabilitation (IRiFoR), and the Presidium of Ophthalmology Hospital Santa Chiara in Trento for their collaboration with the Center for Maind/Brain Sciences.

Finally, thanks to all the staff in the Department of Cognitive Sciences and in particular to al my colleagues who have made my daily life a good one for these three years.

Last but not least, thanks to my family.

Stefano Targher has been supported by a grant provided by 'Fondazione Trentino Università' (FTU). All the studies have been realized also thanks to the support from the 'Provincia autonoma di Trento' and the 'Fondazione Cassa di Risparmio di Trento e Rovereto'.

Index

Outline of the Thesis

How humans perceive everyday reality is one of the most fascinating and enduring interest of different scientific disciplines. The aim of the present dissertation is to investigate - from both a theoretical and an empirical perspective - some aspects concerning the crossmodal interactions between hearing and vision. In the first part of the introduction special attention is given to the cortical and subcortical neural substrate involved in integrating different sensory modalities, and more specifically, audiovisual stimuli. Experimental studies in the literature designed to empirically investigate different aspects of audiovisual interaction and the potential existence of a sensory dominance between hearing and vision will be presented. In the last part, the introduction will be focalized to discuss some of the principal models aimed to predict the outcome of audiovisual integration and its relation with sensory dominance.

The following chapters present the experimental studies designed to empirically investigate different aspects of audiovisual interactions and the role of eye movements in auditory cognition.

In a first study, the effects of eye movements on auditory spatial representation will be explored. The aim is to disentangle controversial results in the literature emerged in two studies that used different type of sounds (i.e., free field sounds provided through loudspeakers vs. sounds provided intracranially through headphones) and different tasks.

In a second study, the disputed relation between perception and action will be investigated in presence of a crossmodal audiovisual illusion. The aim is to verify whether participants' visuo-motor behaviour might be biased by the visual illusion as emerged for perception, and eventually, whether perceptive and motor biases are correlated.

The last study presented in this dissertation will explore the effects of crossmodal audiovisual stimulation in low vision patients and the relation between the visual pathology. More precisely, the possible visual detection enhancement provided by a sound coupled with a visual stimulus will be investigated. To this purpose, in a first experiment, the effect of spatial disparity between audiovisual stimuli will be deepened while the last two experiments will be focalized on the effects of temporal audiovisual disparities of crossmodal stimuli.

The results of the studies described in the present dissertation provide evidence of an effect of eye movements in the auditory spatial cognition and a relation between the perceptive and visuo-motor systems in presence of an illusion induced by a sound. Moreover the presented findings report for the first time a significant crossmodal effect of audition on visual perception in low vision patients

The multisensory brain

Introduction

Human individuals continuously interact with an environment that provides a large amount of sensory information. Research has widely documented that inputs delivered by different sensory channels tend to be bound together by our brain. The process by which the human nervous system tends to merge together the available parts of different information in a unique event is commonly known as 'multisensory integration' (see Calvert, Spence & Stein, 2004, for a review).

A real event is often multisensory, thus, carried to the brain by different sensory channels (e.g., Alais, Newell & Mamassian, 2010). Neuroscientists trying to understand how the mind/brain works have also tended to focus on attempting to figure out how each of our senses operates in isolation first, before considering how the senses interact with one another to give rise to the rich multisensory precepts of everyday life. This reductionists approach was motivated primarily by the practical assumption that to understand such a complex structure as the brain one needs to take a modular approach (e.g., Fodor, 1983), breaking down perception, or any other cognitive process into its component parts. It may also partially reflect the fact that so much of our phenomenology is unisensory. For example, when people try to understand what another person is saying they typically think that all the information they get comes from what they hear, but as we will see shortly, research now shows that a lot of it actually comes from vision as well.

The idea that each sensory modality is processed independently through its own pathway and then is fused with the other sensory modalities in a later stage on the associations areas of the brain in a hierarchical fashion (e.g., Felleman & Van Essen, 1991) has progressively been abandoned. To date, the knowledge of physiological and cognitive mechanisms characterizing each single sensory modality has improved enormously. Nevertheless, it is not sufficient to know how information from each sensory modality is transduced and processed along its own pathway to explain how the final 'brain computation' will affect the behaviour. The growing amount in understanding unisensory processing and the evidences that multisensory interaction occurs much earlier than had been previously supposed, leaded neuroscientist to develop a multisensory approach of perception (e.g., Pavani, Murray & Schroeder, 2007). Indeed, a sensory modality is very often influenced by other sensory modalities and the final percept may vary significantly in relation to the weight that our brain will assign to each involved modality (e.g., Alais & Burr, 2004; Ernst & Banks, 2002). To perceive a real event, the brain has to sort through massive multiple streams of information and to find which information is related with another as parts of the same event. Furthermore, the brain has to discard information that is unrelated to the event of interest. Our different senses are designed to interact and enhance quantitatively and qualitatively the information characterizing physical phenomena in the surrounding environment. Redundancy of information through different sensory channels enhances the probability that an event will be detected (more rapidly and accurately; Calvert et al., 2004; Rowland, Quessy, Stanford & Stein, 2007) thus enhancing, from an evolutionary point of view, the probability to survive. A striking example of the potential benefits that efficient multisensory integration can hold comes from our ability to hear someone speaking in a noisy environment, such as a cocktail party situation. Under such noisy conditions, the ability to see the speaker's lips can enhance our ability to hear what the person is saying by an amount equivalent to a 15 dB increase in the sound intensity level (e.g., Sumby & Pollack, 1954).

Meredith and Stein (1990) established some key principles of sensory integration for a new approach in which ‘early’ interaction between different sensory modalities was the prominent view. Meredith and Stein (1990) demonstrated that visual, auditory and somatosensory information is integrated in the superior colliculus (SC) before to reach primary sensory cortices. In the SC there are neurons that respond to different sensory modalities (i.e., multisensory neurons). Furthermore, time and space seems to be two physical constrains for multisensory interaction taking place in SC’s multisensory cell (Meredith & Stein, 1990). Research suggests that spatial coincidence and temporal synchrony are the two key factors determining whether multisensory integration will take place to give us the rich multisensory perceptual objects of everyday life (cf. Driver & Spence, 2000; Slutsky & Recanzone, 2001; Stein & Meredith, 1993).

Another peculiarity in the SC’s multisensory neurons is the phenomenon called ‘superadditivity’. Superadditivity is characterized by higher neuron responses to visual and auditory stimulation presented together than the algebraic summation of the single responses when the stimuli are presented individually (see Stein & Meredith, 1993 for an extensive coverage of this topic). Furthermore, superadditivity has been found to be proportionally higher to weak unimodal sensory stimuli, the so called principle of inverse effectiveness (Stein & Meredith 1993; Stein, London, Wilkinson & Price, 1996). Moreover, superadditivity disappears in case of high unimodal stimulation (i.e., in this case the multisensory response often may vary between the best unisensory response and the algebraic summation of both the unisensory responses). Both, spatial and temporal constrains make sense in a world in which an event (very often multisensory) is spatially and temporally well defined, while superadditivity avoid weak stimuli being neglected.

At behavioural level, the final result of this multisensory interaction seems to be the 'product' of two main factors; first, an online cortical modulation built through direct experience and second, the stimuli reliability (Alais & Burr, 2004; Ernst & Banks, 2002; Ernst, & Bülthoff, 2004; Shams & Kim, 2010 for a review). Concerning the cortical modulation on low-level multisensory interaction it has been observed that when descending cortical afferences to the SC are blocked (e.g., by deactivation of visual anterior ectosylvian sulcus, AES; Jiang, Wallace, Jiang, Vaughan & Stein, 2001), cat's multisensory SC neurons will respond with the same intensity for both multisensory or unisensory stimulation. That is, SC multisensory neurons, in this case, lose their superadditive response property (Jiang et al., 2001). Furthermore, multisensory cells in the SC are either absent at birth or are not able to integrate multisensory input in both the cat and the monkey (Wallace, Carriere, Perrault, Vaughan, & Stein, 2006; Wallace & Stein, 2001). This means that the direct experience with real multisensory events is necessary to develop and calibrate temporal register and spatial maps in different sensory modalities on SC's multisensory neurons (Ernst, & Bülthoff, 2004; Wallace & Stein, 2001, 2007). Finally, only perceptually coherent events may elicit a multisensory response. Indeed, we build our knowledge of the world through experience. Causality is an example of an important 'rule' used by all the living being organisms to learn world physical effects. If to an action often corresponds the same effect or to a particular sound is often associated to a particular visual scene, this relation will be learned and 'stored' in the brain. Subsequently, this memory will be more likely recalled if all, or some of, the constituents are experienced (e.g. visual and acoustic constituents of an events already experienced like a barking dog will be more likely integrated in respect to those of a barking cat; Barraclough, Xiao, Baker, Oram & Perrett, 2005).

During the past decades visual perception has been considered as the dominant modality, highly modular, with separate brain areas and mechanisms involved in processing different information (e.g., motion, colour, form, location). The view of vision as the dominant sensory modality has been reinforced by classic studies of crossmodal interactions in which experimenters artificially imposed a conflict between visual information and other sensory modalities. For example, evidences of this visual sensory dominance view emerged in the ventriloquist effect in which the perceived sound's location is strongly biased towards the location of the visual stimulus (Thurlow, 1973; Warren, Welch & McCarthy, 1981). Not only, the dominance of vision has been reported also in relation to the tactile modality (Rock & Victor, 1964) and even in speech perception with the well known McGurk effect (McGurk & MacDonald, 1976).

Despite vision is the sensory modality that provides more reliable spatial information, it seems that visual perception can be strongly altered by sound or touch (see Shams & Kim, 2010 for a review). That is, stimulus reliability seems to be one of the important properties that lead multisensory perception. In fact, for instance, if in an audiovisual stimulation spatial visual cues are more reliable than auditory cues, then vision will lead multisensory perception. Thus, in case of integration of the two spatially disparate visual and auditory cues, the acoustic stimulus will be spatially shifted towards the visual stimulus and the crossmodal event will be perceived to come from the spatial position provided by the visual modality. Conversely, the dominance of vision over audition will be compromised in case of least reliable visual spatial cues (e.g., by blurring the visual target over a large region of space; see Alais & Burr, 2004). Not only, in some circumstances, spatial visual cues can be illusorily shifted towards the position cues provided by the sound (e.g., Hidaka et al., 2009).

A large amount of neurophysiological studies has shown that the integration between different sensory inputs occurs in various brain structures such as the superior temporal sulcus and prefrontal cortex (e.g., Alais et al., 2010; Andersen, Snyder, Bradley & Xing, 1997; Barth, Goldberg, Brett, & Di, 1995; Calvert & Thesen, 2004; Duhamel, Colby, & Goldberg, 1998; Falchier, Clavagnier, Barone & Kennedy, 2002; Macaluso & Driver, 2005; Wallace, Meredith, & Stein, 1992). The extreme complexity of the multisensory integration network and the mutual effects of the involved cortical and subcortical structures support the involvement of cognitive and attention-like (i.e., top-down) factors over early sensory integration outcomes (i.e., bottom-up). Wallace and colleagues (Wallace, Roberson, Hairston, Stein, Vaughan & Schirillo, 2004) found that the spatial bias of auditory stimuli towards the light was highly modulated by the participant's perception (judgment) of a unique visual-acoustic event (i.e., the 'unity assumption'; Welch & Warren, 1980; see p. 21 for more details). Unity assumption refers to the process of crossmodal binding in which observers will be more likely to assume that sensory inputs from different modalities have a common spatiotemporal origin (i.e., they originate from a common event) and hence will be more likely to bind them into a single unified percept. Conversely, in case of crossmodal segregation, the different sensory inputs will be perceived as different unimodal events.

If at the level of the SC structural constrains (i.e., spatial and temporal) seem to play a fundamental role, it is still a matter of debate whether cognitive components are also involved (i.e., behaviourally). For example, the experience with spatially or temporally disparate stimuli seems also to provide integration benefit (Wallace & Stein, 2007). Furthermore, the time necessary to learn new associations between different modalities or to disrupt old ones, seem to be very short (Lippert, Logothetis, & Kayser, 2007) and the principle of inverse effectiveness seems not to be so clear on cortical areas as on the SC (see Holmes, 2007; 2009).

Multisensory interaction in the Superior Colliculus

The superior colliculus (SC) is one of the most frequently investigated brain structures located in the vertebrate midbrain (Stein & Meredith, 1993). The SC is a structure that mediates orienting movements of the eye, the head and the ears towards the source position of visual, acoustic and somatosensory stimuli (e.g. Stein & Clamann, 1981). The SC is structured with different layers to which converge visual, auditory and somatosensory information. Neurons in the superficial layers are purely visual while those in the deep layers are often characterized by bimodal visual-auditory or visual-somatosensory inputs but may be even trimodal. Visual, acoustic and somatic spaces are topographically represented in the intermediate and deep layers (Middlebrooks & Knudsen, 1984). Multisensory neuron's receptive fields (RF) overlap, and for this reason multisensory cells respond to stimuli from the same external region of space despite the sensory modality input. Moreover, RFs are arranged to provide a functional map of the external space so that SC's multisensory neurons receptive fields are in 'spatial register' (Meredith & Stein, 1990). SC has also a pivotal role in saccade generation (e.g., Stein & Clamann, 1981). The deeper layers of the SC are presaccadic; they are also called visuo-motor as they exhibit small visual bursts before the actual motor bursts. They are also topographically organized: the neurons discharge before a saccade directed to a specific region of the visual field which is in register with the visual fields of the visual neurons situated in the superior layers. Different study highlights the effect of different sensory modalities on the same SC multisensory motor neuron involved in saccade generation (e.g., Stein & Clamann, 1981; Groh & Sparks, 1996a, b). This organization pattern characterizes the SC 'premotor map' allowing transducing different sensory cues into common motor responses. Indeed, SC's premotor map structure allows orienting eye movements towards the spatial position provided by one or more sensory information that take place in the same external position. Integrations of different sensory modalities allow, for example, orienting the visual system towards auditory or somatosensory stimulation that take place where the eyes cannot see (e.g., on the rear space or behind obstacles).

Neural recordings at the single unit level of SC have highlighted several peculiarities governing the multisensory interaction (see Stein & Meredith, 1993 for an extensive coverage of this topic). A first characteristic is that the neurons of the SC have spatially corresponding receptive fields that map different modalities (i.e., vision, touch and audition; see Stein, 1998; Stein & Meredith, 1993; Stein, Stanford, & Rowland, 2009). The so-called 'spatial principle of multisensory integration' postulates that the neural response enhancement produced by multisensory stimuli is dependent on the spatial alignment and/or overlap of the excitatory receptive fields of their individual sensory components (e.g., Stein & Meredith, 1993; Wallace et al., 1992). Therefore, for instance, visual and auditory stimuli originating from the same spatial position are more likely to be bound together, thus enhancing the neuronal responses and facilitating their detection at a behavioural level (e.g., Andersen & Mamassian, 2008; Doyle & Snowden, 2001; Frassinetti, Bolognini, Bottari, Bonora, & Làdavas, 2005; Frassinetti, Bolognini, & Làdavas, 2002; Frassinetti, Pavani, & Làdavas, 2002; Hairston, Laurienti, Mishra, Burdette, & Wallace, 2003; Marks, Szczesiul, & Ohlott, 1986). However, facilitatory multisensory interaction (i.e., neuronal response enhancement) can be observed even when the stimuli are spatially misaligned in their external positions, provided that the relevant neurons contain sufficiently large RFs, such that each stimulated position falls within their excitatory zones (Wallace & Stein, 2007). If this is not the case, no facilitation or even a response depression can be observed (Kadunce, Vaughan, Wallace, Benedek, & Stein, 1997; Meredith, Nemitz, & Stein, 1987). In case of response depression the intensity of the multisensory response may be below the weaker unimodal response.

A second principle driving multisensory interaction concerns the relative timing of the two sensory events. For example, at the level of the SC, the multisensory 'spatial' enhancement is maximized when the peak activities of the auditory and visual inputs overlap. It typically happens when the stimuli are presented simultaneously, although sensory integration can take place also between stimuli that are not temporally coincident but fall within the 'temporal window' of integration (e.g., Spence & Squire, 2003). Indeed, at the level of SC, multisensory enhancement may emerge also in case of temporally disparate audiovisual stimuli. Despite the asynchronous onset of the two modalities, the evoked responses in multisensory neurons can have significantly more impulses than those evoked by the strongest of the individual stimuli (Meredith et al., 1987). Differences in physical propagation of acoustic and visual information (i.e., visual speed is 900000 times faster) and transduction times together with neural latencies, may also explain this flexibility of the brain to integrate acoustic and visual stimuli. In fact, it has been shown that the temporal window of integration operates up to a distance of around 10 meters, a distance in which our brain my compensate physical with transduction speed differences between visual and acoustic stimuli (Sugita & Suzuki, 2003). Furthermore, other mechanisms like the temporal ventriloquism effect (e.g., Bertelson & Aschersleben, 2003; Morein-Zamir, Soto-Faraco, & Kingstone, 2003) in which the onset of a visual stimulus is shifted toward the onset of an acoustic stimulus may play a fundamental role to account of multisensory temporal perception. Temporal ventriloquism reflects the higher reliability of the auditory modality on the temporal domain.

Another feature of multisensory interaction is that the related neuronal response is proportionally greater when unimodal stimuli are less effective (i.e., the principle of inverse effectiveness; Rowland & Stein, 2008; Stein et al., 1996). In particular, the principle of inverse effectiveness, originally proposed by Stein and colleagues, states that the strength of multisensory integration increases as the neural response to stimuli presented in isolation decreases (see Stein & Meredith 1993 for an overview; though see Holmes 2007, 2009). Whereas the co-occurrence of weak stimuli contributes to the increase of the multisensory neural activity, a little increase can be observed when stimuli coming from different modalities are highly effective. Multisensory neuron response is often categorized as subadditive, additive or superadditive. Additive, when response intensity is around the algebraic summation of responses to unimodal stimuli, subadditive when response intensity is below to the algebraic summation of responses to unimodal stimuli, superadditive in case of response intensity above the algebraic summation of responses to unimodal stimuli (often associated to weak unimodal stimuli). Stanford and colleagues (Stanford, Quessy & Stein, 2005) provided a clear quantification of enhancement effects of visual acoustic stimulation in multisensory SC's neurons. Their results show that the majority of bimodal combinations yielded additive responses while superadditivity was observed in only 24% of neurons and subadditivity in 7% of neurons.

A very interesting issue is that SC's multisensory integration/segregation properties seem to develop through experience. In fact, multisensory cells are either absent at birth or are not able to integrate multisensory input (Wallace & Stein, 2000, 2001). Different inputs from the cortex to the SC seem to be fundamental for the multisensory development. For example, in absence of inputs from the association cortex, multimodal cells in the cat's SC are unable to integrate different sensory inputs (Stein et al., 2009). There are two main areas in the cat's association cortex that are involved in the multisensory development: the anterior ectosylvian sulcus (AES) and the lateral suprasylvian sulcus (rLS). The SC is a primary site of converging inputs from different senses (Meredith & Stein, 1986; Stein & Meredith, 1993) and it receives sensory inputs from many unisensory structures (Wallace, Meredith & Stein, 1993). For example an audiovisual multisensory neuron receives descending visual afferences from the anterior ectosylvian visual area (AEV) and auditory inputs from the auditory field of the anterior ectosylvian region (FAES) while the output of this multisensory visual acoustic neuron project mostly to motor related areas of the brainstem and spinal cord involved in the control of orientation behaviours (Moschovakis & Karabelas, 1985; Peck, 1987; Stein & Meredith, 1993). Consequently, the principles that govern the multisensory interaction of unimodal visual and acoustic stimuli at the level of the single SC's neuron (i.e., inputs) seem to be reflected in orientation behaviours (i.e., output; Jiang, Jiang & Stein, 2002; Jiang, Jiang, Rowland & Stein, 2007; Stein, Meredith, Huneycutt & Mc Dade, 1989; Wilkinson, Meredith & Stein, 1996).

Another interesting question is whether in the same multisensory neuron the computation for the integration of stimuli of the same modality is the same as for different modalities. It seems that at least for some SC's multisensory neurons of the cat, the computation is very different (Alvarado, Vaughan, Stanford & Stein, 2007; Wallace & Stein, 1994). In fact, while the integration of different modalities can provide a multisensory cell response that can be additive, subadditive, superadditive or depressed as a function, for instance, of the spatiotemporal characteristic of the involved different modalities, integrating stimuli from the same sensory modality seems to provide a response which is very close to the lowest of the two (i.e., subadditive; Stain & Stanford, 2008 for a review; Wallace & Stein, 1994). The computation of inputs from different receptive fields providing information from different modalities seems to be one key factor to obtain enhancement of the multisensory neuron response. In fact, in case of unisensory stimulation, the excited multisensory neuron's receptive field is the same for both unisensory stimuli. These representative samples are exemplary of the characteristic differences between the neural computations that underlie multisensory and unisensory integration in multisensory neurons.

Multisensory integration in the cortex

Multisensory interaction seems to characterize most of the cerebral cortex despite of the hierarchical process of information view characterizing cognitive science of few decades ago (e.g., Felleman & Van Essen, 1991). Recent imaging studies show that multisensory interaction may be present in the association areas but also in primary sensory and motor areas (see Klemen & Chambers, 2011 for a recent review). Indeed, in primary sensory areas inputs from the sensory modality have the major weight on the information processing, nevertheless, other sensory modalities can exert a relevant influence. Non-human primates constitute a good model to study the neuronal mechanisms of multisensory interaction. Anatomical and electrophysiological approaches have been developed to assess whether crossmodal influences can occur at early stages of sensory processing. A recent animal anatomical study revealed the presence of direct links between the primary visual area (V1) and auditory areas (Cappe, Rouiller & Barone, 2009). Recording of single unit activity in the primary visual cortex of the monkey does not reveal pure auditory responses but, in line with the principle of inverse effectiveness, a modulation (i.e., reduction) in neuronal response latencies when visual and auditory stimuli are at middle intensity but not when are at high intensity (Wang, Celebrini, Trotter & Barone, 2008). The primary striate visual cortex V1 of the monkey is highly connected with the extra-striate visual cortex V2 and V3. Furthermore, few inputs converge to V1 from multisensory areas like the superior temporal polisensory region (STP) and areas of the posterior parietal cortex (PPC) more precisely from lateral and ventral intraparietal areas (Kaas & Collins, 2004).

In humans and non-human primates studies have focused on the posterior parietal cortex (PPC) where visual, vestibular, tactile and auditory sensory information converges to constitute a common multisensory spatial map (Graziano, 2001). Indeed, in PPC sensory signals are coded in common coordinate frames, such as auditory-visual or visual-somatosensory maps in eye-centred coordinates (Cohen & Andersen, 2000). Lateral intraparietal (LIP), anterior intraparietal (AIP), medial intraparietal (MIP) and ventral intraparietal (VIP) areas are different PPC subregions involved in various aspects of spatial attention and action. For example, the LIP neurons encode visual and auditory stimuli with respect to current eye position, requiring auditory receptive fields dynamically realign with changes in gaze direction (Avillac, Deneve, Olivier, Pouget & Duhamel, 2005; Cohen & Andersen, 2000; Schlack, Sterbing-D'Angelo, Hartung, Hoffmann & Bremmer, 2005). Also in MIP neurons, a subregion that seems to be involved in limbs reaching movements, visual and auditory targets are encoded in a common eye-centred coordinate map (Batista, Buneo, Snyder, Andersen, 1999; Cohen & Andersen, 2000). By using single cells recording, Avilac and colleagues (Avillac et al., 2005; Avillac, Ben Hamed, & Duhamel, 2007) found that in the macaque VIP area the majority of neurons perform visual tactile multisensory integration following the general multisensory interaction principles (e.g., spatial and temporal coincidence) that are closely similar to those described in other cortical and subcortical regions. Surprisingly, the response direction is not as that observed in the SC. In fact, visual and tactile stimuli presented in the same position of space mainly produce response depression (i.e., the magnitude of the multisensory response is lower than that of the response to the best unisensory stimulus). Furthermore, multisensory integration occurs most often when the visual and tactile stimuli are spatially aligned but some neurons combine multisensory events for incongruent stimuli and few show multisensory integration for stimuli that are either spatially congruent or incongruent. Moreover, integration produced by spatially incongruent stimuli could provide either response enhancement or depression with the same probability. Avillac and colleagues (2007) also found that most of the unisensory neurons response in macaque VIP area is influenced by another modality even if cells do not respond directly to that modality as found more recently by Wang and colleagues (2008).

In humans, electric functional imaging studies (event-related potential, ERP and the magnetoencephalography, MEG) providing high temporal resolution, highlighted changes in primary sensory areas of early-evoked potentials related to crossmodal stimulation (i.e., between 50 and 200 ms after stimulus onset). Indeed, it has been found enhancement of auditory evoked responses when an additional somatosensory stimulus was applied to the hand (Murray, Molholm, Michel, Heslenfeld, Ritter, Javitt, Schroeder & Foxe, 2005), while another crossmodal visual acoustic crossmodal study reported early multisensory effects in visual, auditory and fronto-temporal areas (Giard & Peronnet, 1999; Fort, Delpuech, Pernier & Giard, 2002). In another audiovisual crossmodal study Molholm and colleagues (Molholm, Ritter, Murray, Javitt, Schroeder & Foxe, 2002) found that RTs to crossmodal stimuli, when presented simultaneously, were significantly faster than when stimuli were presented alone. Furthermore, EEG studies for the McGurk illusion (McGurk & MacDonald, 1976) show visual modulation over classical auditory areas (Colin, Radeau, Soquet, Demolin, Colin & Deltenre, 2002; Mottonen, Krause, Tiippana & Sams, 2002).

Single cells recording in animals and non-human primates clearly indicate that interaction between sensory modalities is spread in the brain also at early stages of sensory processing (e.g., Kaas & Collins, 2004; Cappe et al, 2009). These results are absolutely not in line with a hierarchical view of sensory information processing.

Indeed, recent evidence highlights that multisensory interactions occur in a richly interconnected system (Klemen & Chambers, 2011). To further underline these results, electrophysiological studies in humans provide evidences of much similar early interaction in different multimodal tasks (e.g., Fort et al., 2002; Molhom et al. 2002; Murray at al., 2005).

Humans' neuroimaging studies (e.g., functional magnetic resonance imaging, fMRI) report multisensory interaction in various cortical areas (see Klemen & Chambers, 2011 for a recent review) however, most of such studies did not find superadditivity maybe due to low temporal resolution constrains of this method and a small population of multisensory neurons involved in the process of sensory information. Furthermore, interpretation of the crossmodal sensory signal obtained during human brain imaging studies might be attributed to the presynaptic inputs from different sensory systems (i.e., both excitatory and inhibitory presynaptic activity) rather than the resulted sensory output (see Logothetis, Pauls, Augath, Trinath & Oeltermann, 2001). For this reasons activations observed in imaging studies after multisensory stimulation might be not considered as a real multisensory response.

Despite fMRI methodology difficulties and data interpretation issues, this methodology seems to be widely used and prominent to provide very good insights. First, because for the recent developments in multivoxel pattern analysis (MVPA) and second, because multimodal neuroimaging technique provides results of more than one measurement technique acquired on the same task, from the same participants at the same time (see Klemen & Chambers, 2011). Kayser and coworkers highlighted crossmodal modulation effects of somatosensory and visual stimuli in the auditory cortex in two different studies (Kayser, Petkov, Augath, Logothetis, 2005; Kayser, Petkov, Augath & Logothetis, 2007). Ghanzanfar and colleagues also found a modulation of most cells in primary auditory cortex (A1) provided by visual stimuli

(Ghazanfar, Maier, Hoffman & Logothetis, 2005). Furthermore, Lakatos and colleagues also found auditory and somatosensory interactions at nearly every recording site in A1 (Lakatos, Chen, O'Connell, Mills & Schroeder, 2007). Subadditive multisensory response in STS has been found with acoustic and tactile stimuli (Beauchamp, Lee, Argall & Martin, 2004) and with both audiotactile and visuotactile stimuli (Beauchamp, Yasar, Frye & Ro, 2008).

In a recent study, Stevenson and colleagues (Stevenson, Geoghegan & James, 2007) by varying signal strength of audiovisual stimuli found clearly superadditive responses for weak stimuli showing the effects of principle of inverse effectiveness in the superior temporal sulcus (STS). Foxe and colleagues (Foxe, Wylie, Martinez, Schroeder, Javitt, Guilfoyle, Ritter & Murray, 2002) by presenting to the participants tactile and auditory stimuli found that unimodal tactile stimulation activated areas of auditory association cortex and left superior temporal gyrus (STG). Moreover, when the auditory and haptic stimuli were presented simultaneously, superadditivity was observed in the left STG. Calvert and colleagues (Calvert, Campbell & Brammer, 2000) presented participants with audiovisual speech. Compared to unimodal stimulation responses, simultaneous audiovisual speech produced superadditive response while asynchronous presentation provided subadditive response in the left STS.

Reaction times facilitation for audiovisual multisensory stimuli compared to unisensory stimuli reported also in macaque monkeys (Cappe, Murray, Barone & Rouiller, 2010) has been recently confirmed in humans (Senkowski, Saint-Amour, Höfle & Fox, 2011). This facilitation has been observed only with subthreshold but not with suprathreshold auditory stimuli. Similarly, Noesselt and colleagues (Noesselt, Boehler, Budinger, Heinze & Driver, 2010) have shown that the simultaneous presentations of auditory stimuli enhance the behavioural visual detection for low intensity visual stimuli but not for high intensity visual stimuli. Furthermore, Noesselt and colleagues' results indicate that multisensory enhancement of detection sensitivity for low-contrast visual stimuli by co-occurring sounds reflects a brain network involving not only established multisensory STS and sensory specific cortex but also visual and auditory thalamus. These very recent psychophysical and behavioural results provide further evidences consistent with the principle of inverse effectiveness. Furthermore, these results highlighting the same behavioural facilitation mechanisms in humans and non-human primates, allow to better interpreting human imaging studies due to the evidence of similarities between human's and non-human's primate brain structure.

Taken together, single cells recording, electrophysiological and fMRI imaging studies highlight the extreme complexity of interaction between different modalities indicating the presence of both feedback and feedforward connections between early processing of sensory information (i.e., subcortical structures and primary sensory cortices) and high-level association areas (e.g., FEF, STS, STG, LIP, VIP). However, the outcome of this complex multisensory interaction network is not easily predictable because of the interaction between environmental sensory information and their online interpretation that highly depend to the past recipient experience. Thus, the outcome behaviour may differ significantly among different recipients explaining why it is possible to obtain highly different outcomes from the same physical event. While at the level of the SC multisensory interaction clearly seem to obey to structural constrains (i.e., spatiotemporal and stimulus intensity), this is not so clear if cognitive components are also involved (i.e., behaviourally). Is undeniable that many cortical regions exhibit crossmodal interactions, however, some principles very clear at the level of the SC are not so pronounced at the level of cortical areas (Holmes, 2007; 2009). One of the possible explanations may be hypothesized by observing results of experiments in which 'particular' spatiotemporal relation characterizes external crossmodal events. For example, the experience (i.e., adaptation) with spatially or temporally disparate stimuli seems also to provide integration benefit (Wallace & Stein, 2007). Wallace and colleagues reared cats in an altered sensory environment in which visual and auditory stimuli were temporally coupled but originated from different locations. Neurons in the superior colliculus of those cats developed a form of multisensory integration in which spatially disparate visual-auditory stimuli were integrated. The animals learned to associate disparate stimuli because those were the environmental constrains. The data suggest that the principles governing multisensory integration are highly plastic and that there is no a priori spatial relationship between different modalities for their integration. Not only, the time necessary to learn new associations between different modalities or to disrupt old ones, seem to be very short (Lippert et al.,2007). Lippert and colleagues were able to eliminate in just one experimental section, the enhancement effect provide by the sound in a synchronous crossmodal audiovisual stimulation in a visual detection task. In a first experiment in which the sound was always synchronous with the visual stimulus, the authors found visual enhancement in the crossmodal stimulation in respect the unimodal visual condition. With another group of participants, by varying the stimulus onset asynchrony between sound and target in the crossmodal conditions (i.e., −400, −150, 0, +150 and +400 ms), the authors were able to eliminate the visual detection enhancement effect provided by the sound even in the synchronous condition. Taken together, Wallace and Stein (2007) and Lippert et al. (2007) findings indicate that our cognitive system seems to learn and adapt very fast to different associations between modalities in a useful way to optimize task demands (i.e., environmental constrains). The outcome of such adaptation may exploit crossmodal integration enhancement only if there is a stable relation between spatial and temporal cues of different modalities. In this way, integrations principles governing early stages of sensory processing (e.g., at the level of the SC) are likely to be established by the modulation (e.g., by means of feedforward connection) of high-level association cortices that have to 'find' the useful (i.e., constant) relation to learn and exploit in the future. How fast may be this plasticity, how match well established relation might be modified and what kind of limits there are in establishing new relations is still to be deepened. It is likely that differences between single cells recording results at the level of subcortical and associations areas may be due just because in the latter are mostly involved aware flexible decision factors and such factors are the basis to establish new relations at early integration levels.

An important difference in multisensory integration between subcortical and cortical areas seems to be that in the latter, multisensory events have to be coherent or already experienced in a way that the event can be recognized by means of all its constituents (i.e., experienced association between modalities). In this multisensory context, coherence means that a sound stimulus, for instance, should be an ecologically valid match to a given visual object (e.g., a 'barking' sound would be coherent with an image of a dog while a 'meowing' sound would not). This issue, concerning which particular inputs from one sense should be jointly weighted together with which particular selection of inputs from other senses, refer to one of the most discussed actual topic of 'unity assumption' (Welch & Warren, 1980).

Stimulus reliability, optimal integration and the unity assumption

How the cognitive system is able to bind information from different modalities and which are the rules governing such phenomenon is still debated. Different models have been proposed to predict the outcome of crossmodal stimulations in relation to physical information provided by the constituent unimodal stimuli (i.e., reliability of each modality cues).

The 'modality precision hypothesis' proposed by different authors (Choe, Welch, Guilford & Juola, 1975; Fisher, 1968; Kaufman, 1974; Howard & Templeton, 1966) states that the more precise information cues provides one modality, the more it will bias the integration outcome towards those cues. For instance, in normal conditions, vision dominates audition for spatial tasks, and audition dominates vision for temporal tasks. In the well know audiovisual 'ventriloquist effect' in which the movements of the dummy's mouth alter the perceived location of the ventriloquist's voice (Howard & Templeton 1966) it seems that such crossmodal localisation bias depends on the spatial and temporal relationships between the stimuli, with bias declining as a function of increasing temporal and spatial disparity (e.g., Hairston, Wallace, Vaughan, Stein, Norris & Schirillo 2003; Lewald & Guski 2003; Lewald, Ehrenstein & Guski, 2001; Radeau & Bertelson 1987; Slutsky & Recanzone 2001; Welch & Warren 1980). Despite the general agreement on this issue, the degree of tolerated disparity and the amount of elicited bias, range widely, probably reflecting the substantial differences between the used experimental paradigms.

The 'modality appropriateness hypothesis' proposed by Freides (1974) is not very different to some extent to the 'modality precision hypothesis' as it states that each modality provide different information cues but is the more appropriate to provide a particular type of cue (e.g., spatial or temporal). For example, vision is the more appropriate sensory modality to provide spatial information. However, while the precision hypothesis seems to be more flexible in a way that as less is the precision of one modality in a particular situation, as less will bias multisensory integration, the appropriateness hypothesis seems do not admit exceptions. For the modality appropriateness hypothesis, the visual modality is always the best to provide spatial cues and for this reason, in all spatially incongruent crossmodal tasks, the resulting percept will be biased toward the visual stimulus location. However, both the reported hypotheses cannot explain why in some circumstances the behavioural bias is toward the least precise/appropriate modality. Visual dominance can be reversed in spatial tasks when visual signals are degraded (inverse ventriloquism; Alais & Burr, 2004) and other 'reversing' dominance effects can be obtained by manipulating unimodal stimuli reliabilities in other sensory modalities (e.g., visual-haptic estimation of height; Ernst & Banks, 2002).

A model that formalizes the effect of modalities reliability is the maximum likelihood estimation (MLE; Ernst & Banks, 2002) or optimal combination model (Alais & Burr, 2004) two models based on the 'stimulus reliability hypothesis'. Conceptually this two models works with the same mechanisms; they are a weighted linear sum that combines two or more signals that are weighted by their reliability. Thus, the modalities with reliable cues receive a high weight, while those with less reliable cues receive a low weight. The combination rule is considered statistically optimal because it provides a result that is the most reliable (i.e., most probable or least variable). Results provided by visuotactile (Ernst & Banks, 2002) and audiovisual integration (Alais & Burr, 2004) studies provide evidences to the MLE model. Furthermore, other studies have shown that human crossmodal perception closely matches predictions from the MLE model even in trimodal contexts (Wozny, Seitz, & Shams, 2008), as well as between independent cues within a single modality (Hillis, Ernst, Banks & Landy, 2002). Thus, the MLE seems to be a good predictor for the behavioural outcome or, in case of ambiguity, for the direction of the bias.

Other studies have been focalized on the role of audition in crossmodal integration. Audition is more accurate than vision to provide temporal information. A sound presented in close temporal proximity to a visual stimulus can alter the perceived temporal dimensions of the visual stimulus (temporal ventriloquism; Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Burr, Banks & Morrone, 2009; Morein-Zamiret al., 2003). In a first study, Bertelson & Aschersleben (2003) asked to participants to judge the order of occurrence of sound bursts and light flashes separated in time with different stimulus onset asynchronies (SOAs) and delivered either in the same or in different locations. Results pointed out significant longer SOAs needed by participants to correctly respond which of the two stimuli (i.e., flash or sound) were delivered first when the two stimuli where delivered in the same spatial location (i.e., at fixation) in respect to different positions (i.e., one at fixation and the other in the periphery). This result highlights the strong effect of the sound in a temporal judgment task while emphasize the role of congruent spatial cues that seem to have a particular role in the crossmodal binding. The temporal dominance of the sound was effective only when acoustic and visual stimuli were presented in the same spatial position.

In a second sensory motor study Aschersleben & Bertelson (2003) asked to participants to produce tapping movements in synchrony with a sequence of repeated reference signals. In a first experiment, the reference signals were light flashes, each one preceded or followed with different stimulus onset asynchronies (SOAs) by an auditory stimulus that the participant was instructed to ignore. Results show that the timing of the tap was strongly biased toward the occurrence time of the sound (i.e., distractor that had to be ignored). In a second experiment, the reference signals were auditory while light flashes have to be ignored (i.e., distractors). The timing of the taps was biased significantly toward the flash distracters, but the effect was far weaker than that provided by the sounds distractors in the first experiment indicating that the sound seems to lead vision in the temporal domain by strongly biasing the time of occurrence of visual perception. However, the differences between the instructions in the two experiments seem to provide different results biased towards the modality to be ignored. This is not explainable by just the MLE model.

In temporal order judgment task (TOJ) were participants have to report which light between two is presented first, Morein-Zamir and colleagues (2003) for instance, show that a sound trailing the second visual stimulus provide an improvement in the visual temporal accuracy as compared to the baseline condition in which two temporal disparate visual stimuli are presented with two synchronous sounds. That is, it seems that the effect of the sound associated to the second visual stimulus (delayed in respect the visual stimulus) provides a temporal shift of the visual perception towards the occurrence time of the associated sound (i.e., the second light is temporally ventriloquized by the second sound). Conversely, two sounds presented between the two visual stimuli worsened performance but the temporal modulation of the first sound associated to the first visual stimulus have no effect. Only the effect of the second sound seems to increase the temporal disparity between the two lights thus providing a significant benefit to report the correct visual stimuli sequence. The stimulus reliability alone seems not to justify the results of the Morein-Zamir et al's study (2003).

Temporal dominance of the sound can be overcome by providing even an auditory illusion in which a visual event affects the perceived duration of an accompanying sound. Schutz and Lipscomb (2007) discovered an illusion in which a visual event affects the perceived duration of an accompanying sound. They made several videotapes of a professional percussionist playing single notes. The performer played each note using long and short gestures. Participants were asked to judge the durations of sounds and to ignore visual information when judging the duration produced with these two gestures. In the absence of visual information, participants judged the durations of the notes to be equal. However, when they heard the sound while watching the video, they judged notes produced by long gestures to be longer than notes produced by short gestures. The authors concluded that although longer gestures do not produce tones with longer acoustic durations, they do create an illusion in which tones sound longer because of audiovisual integration. In this last study, the temporal dominance of the sound is biased by the visual stimulation that seems to spatially ventriloquize the 'duration of the sound' along the duration of the long movement. In this case stimulus reliability and stimulus appropriateness provides different expectations. It seems that the reliability of different stimuli is not provided only by the mere stimulation per se. How our brain can decide that the more appropriate modality (i.e., the acoustic modality in the temporal domain) is less reliable than the less appropriate (i.e., the visual modality)? How our brain assigns the weight to each modality? Different factors like the prior knowledge (i.e., experience) of the stimulation can be highly relevant. This could be the reason why different results have been obtained with similar experimental design that used different stimuli and might also accounts for the high variability between participants. For instance, the aforementioned audiovisual illusion (Schutz & Lipscomb, 2007) in which the length of the gesture used to produce a sound altered the perception of that sound's duration is at odd with the temporal ventriloquism (e.g., Shams, Kamitani & Shimojo, 2000) in which auditory superior temporal acuity prevails. The stimulus reliability hypothesis does not explain behavioural bias toward the least reliable modality and one of the reasons may be because it does not contemplates the effect of the previous knowledge (i.e., experience of causality, prior probability) that seems to exert a fundamental role in multisensory integration. All the aforementioned studies can be predicted by one or more (or a combination) of de described models. But any of the models seems, at least partially, to predict the outcome of all the studies.

More than thirty years ago Welch and Warren (1980) reviewed the literature on intermodal bias with the aim to find a model able to explain the mechanisms of such bias in crossmodal integration. Welch and Warren's idea of the 'unity assumption' considered a very complex model in which different factors were contemplated: stimulus properties (i.e., modality appropriateness), modality characteristics (i.e., the way in which each sensory information is acquired, transmitted and finally perceived from a biological structural point of view), prior participant knowledge and task instructions (i.e., primary and secondary attentional factors in the model). All this variables are supposed to have a significant effect on whether the perceiver will perceive a unitary event, the extent of this unity binding, and thus the resulted outcome (i.e., direction of the perceptual bias).

The effects of high-level decision components seem to play a fundamental role in multisensory integration outcome. In Bertelson and Aschersleben (2003) study, the role of the sound was significant only when acoustic and visual stimuli were presented in the same spatial position supporting the hypotheses that the influence of the sound on vision emerged only when the two modalities were perceived as a unique crossmodal event. In Aschersleben and Bertelson (2003) study, the resulted bias direction was modulated by the modalities to be ignored. In Schutz and Lipscomb (2007) study, the temporal dominance of the sound was biased towards the duration of the visual gesture supporting the prior knowledge of causality in visual acoustic events.

In other studies as, for example in Wallace and colleagues' study (Wallace et al., 2004) it has been found that the spatial bias of auditory stimuli towards the light was highly modulated by the participant's perception (judgment) of a unique audiovisual event (i.e., the assumption of unity). More precisely, participants performed an auditory localisation task in which they were also asked to report whether they perceived the auditory and visual stimuli to be perceptually unified. The auditory and visual stimuli were delivered at a variety of spatial (0°, 5°, 10°, 15°) and temporal (200, 500, 800 ms) disparities. Result show that regardless of their disparities, whenever participants reported perception of unity, the acoustic stimulus was biased towards the visual stimulus position. Perceptual unity occurred even with substantial spatial (i.e., 15°) and temporal (i.e., 800 ms) disparities. In contrast, when the stimuli were perceived as not unified, auditory localisation was often biased away from the visual stimulus even if acoustic and visual stimuli share the same spatial position.

Another study in which the cognitive component seem to exert a fundamental role in audiovisual interaction is that of Lippert et al. (2007) discussed earlier in which by varying the stimulus onset asynchrony between sound and target in the crossmodal conditions (i.e., −400, −150, 0, +150 and +400 ms), the authors were able to eliminate the visual detection enhancement effect provided by the sound in the synchronous spatially congruent condition. That is, the absence of a stable relation between sound and visual stimulus led to the perception of independent stimulations from the two modalities. Therefore, the audiovisual stimulation was not perceived as a unique crossmodal event, thus eliminating even the enhancement in the synchronous crossmodal condition. The synchronous crossmodal condition provided visual detection enhancement in a previous experiment in which crossmodal stimulation was always synchronous. The authors suggest that the participants' belief about the relation between sound and visual targets is a crucial factor that determines the resulting response. Thus, also in this case, cognitive factor seems to have a prominent role on sensory interaction. Not only, surprisingly, participants' belief about the relation between the two stimuli has been changed very fast. In one experimental section, the lack of a constant relation between visual and acoustic stimuli (i.e., the sound could have been randomly delivered before or after the visual stimulus) deleted the crossmodal sensitivity (*d'*) enhancement previously founded. This result indicates a very fast cognitive plasticity modified by the relation between sensory modalities as highlighted by the participants' change in the decision criterion (*logβ*).

In general, it is undeniable that both perceptual and cognitive factors play a crucial role in such a complex interaction between modalities. Studying how people resolve these conflict situations (i.e., how people weight the various sensory cues when they disagree) helps scientists to understand the rules that normally govern multisensory integration. An example may be the study of Sanabria and colleagues (Sanabria, Spence & Soto-Faraco, 2007) in which they measured the perceptual sensitivity (*d'*) in two different tasks regarding the direction of auditory apparent motion streams presented in noise. In the critical conditions, a visual motion distractor, moving either leftward or rightward, was presented together with the auditory motion. The results demonstrated a significant decrease in sensitivity in reporting the direction of the auditory targets in the crossmodal conditions as compared to the unimodal baseline conditions in which there was not visual distractor. In addition, they observed significant shifts in response criterion (*c*), which were dependent on the relative direction of the distractor apparent motion. Due to the dissociation between criterion direction and performance sensitivity, the authors claim that perceptual and decisional factors are both involved in their crossmodal task but the two components seems to be highly independent. The presence of trials in which the visual stimulus was not informative (i.e. in 50% of the trials the visual stimulus direction was incongruent with the auditory apparent motion direction to be detected) eliminates the crossmodal enhancement also in the congruent condition like found by Lippert and colleagues (2007). Result of Sanabria and colleagues' (2007) study supports the high influence of cognitive factors on participants' response.

Another interesting example is an audiovisual illusion produced when a single flash of light is presented interposed between two brief auditory stimuli separated by 60-100 ms (Shams et al., 2000). Participants typically report perceiving two flashes. On the basis of their results, Shams et al. suggested the 'discontinuity hypothesis' stating that the influence of the discontinuous modality (i.e., the sounds) on the continuous modality is stronger than vice versa. That is, because audition has much better temporal resolution and the fact to have two temporally very close sounds, the sound discontinuity produce a fission (i.e., separation) of the unique (i.e., continuous) visual stimulus into the illusory perception of the occurrence of two visual stimuli. Interestingly, by providing two visual stimuli temporally close and only one sound between them, participants did not report the fusion (i.e., unification) illusion of the two visual stimuli in one unique visual perception. Shams and colleagues' claim was that the modality appropriateness hypothesis does not explain their results because the more 'appropriate' modality (i.e., the sound in this case) did not lead the direction of the crossmodal interaction both in fission and fusion.

By using the same experimental paradigm used by Shams et al. (2000), Andersen and colleagues (Andersen, Tiippana & Sams, 2004) were able to find also a fusion illusion despite it was weaker than the fission one. They modulate the intensity of the sound (i.e., sound reliability) in two different blocks and found that the effect of the sound induced visual illusion was stronger when the beeps were at a clearly audible level than when they were near subjects' auditory threshold, in accordance with the information reliability hypothesis. However, Andersen et al. (2004), by manipulating task instructions, were able to reverse the effect of the illusion by obtaining a visual induced auditory illusion. That is, the authors found an effect of the visual modality on the acoustic modality by asking participants to count the beeps instead of the flashes. Andersen and colleagues' results highlight the contribution of the instruction (i.e., count the flashes or beeps) on the direction of the resulted illusory bias. Indeed, the effect of the instructions on participants' response is contemplated only in the unity assumption hypothesis.

Another particular and very popular example of an illusion in the language comprehension domain (i.e., in terms of meaning of a spoken syllabus) is called the

McGurk Effect (McGurk & MacDonald, 1976). Watching a speaker's face helps us understand what is being said because integrating congruent sight (i.e., lip movements) and sound of speech enhances the brain activity that underlies speech perception (Sumby & Pollack, 1954; Sams, Aulanko, Hämäläinen, Hari, Lounasmaa, Lu & Simola, 1991). In McGurk and McDonald's classical example of this effect, lip movements (i.e., a video of a face uttering /ga/) and sound of speech (i.e., a voice saying /ba/) are not congruent. This results in an auditory percept of hearing /da/. Language is a high-level cognitive function and a particular field of study in which multisensory integration of visual and acoustic stimuli play a fundamental role. Different factors are involved to language comprehension. That is, the experience with the speech word include the knowledge of its sound, the meaning, and the lip movements needed to uttering it together with the emotional content that a moving face induce in the listener. Moreover, visual and acoustic information seem to interact ad early level of perception ((Kayser et al., 2005; 2007; Sams et al., 1991) leading to a comprehension bias effect of vision on audition. Tuomainen and colleagues (Tuomainen, Andersen, Tiippana & Sams, 2005) trained subjects to categorize sine-wave speech tokens in two arbitrary categories but subjects were not aware of the speech like nature of the stimuli. Then the authors used audiovisual stimuli consisting of the sine-wave speech tokens dubbed onto the face uttering the speech tokens and found a weak McGurk effect in the incongruent audiovisual stimulation. Afterwards, subjects were trained in perceiving the sine-wave speech tokens as speech and to categorize them phonetically. When the same crossmodal task was then tested the authors found a very strong McGurk effect. Tuomainen and colleagues (2005) interpreted this result as "evidence of a speech specific mode of audiovisual perception which can be manipulated by cognitive factors". Again, instructions and prior knowledge about speech stimuli (i.e., cognitive factors) seems to have a fundamental role in crossmodal interactions.

Bayesian models seem to adapt to different crossmodal paradigms and predict very well interactions between modalities. This model encompassing the prior knowledge of the perceiver in addition to the reliability of the single modalities is a mathematical formalization very close to what is stated by the unity assumption model and seems to be a very good predictor of the resulting bias. A model based on Bayesian inference has been proposed by Sato and colleagues (Sato, Toyoizumi & Aihara, 2007). In this model, the Bayesian observer adopts the maximum a posteriori approach to estimate the physically delivered positions or timings of presented stimuli (i.e., perceptual reliability), while adaptively changes the inner representation of the Bayesian observer in terms of experience (i.e., cognitive factors). By means of this Bayesian model, the authors were able to reproduces perceived spatial frame shifts due to the audiovisual adaptation known as the ventriloquism aftereffect (Recanzone, 1998). In this adaptation phenomenon, after a prolonged exposure to simultaneous audiovisual stimuli with spatial disparity, the spatial unity perception shifts toward the presented disparity. Bayesian models seem to be very promising for the formalization of multisensory interaction mechanisms.

Taken together results of the aforementioned studies highlight the complexity of the network to process different information and the articulation of different factors involved in the process of bind or segregate different cues coming from the same or different modalities. The above reported studies provide evidences that both low-level perceptual (i.e., stimulus reliability) and high-level cognitive factors (i.e., prior experience and instructions) are involved in crossmodal integrations. Furthermore, cognitive factors and task instructions seem to characterize crossmodal sensory bias mostly in case of uncertainty, that is, when different modalities provide incongruent cues. Thus, maximum likelihood estimation and stimulus reliability hypothesis by lacking in encompassing cognitive factors seem not to be the best formalization to describe multisensory interaction.

Introduction to the studies

The introduction purposed to provide a multisensory general overview by describing structural and behavioural cognitive aspects of multisensory perception. The discussion will be now restricted to the visual and acoustic interaction field by considering the role of eye movements auditory spatial representation. Since eye position plays a fundamental role in visual and acoustic spatial discrimination (e.g., Knudsen & Brainard, 1995), the aim of the first study presented in this dissertation will be focused to disentangle controversial results of two previous studies (see Kopinska & Harris, 2003; Pavani, Husain & Driver, 2008). Pavani and colleagues found a significant effect of voluntary eye movements in an acoustic localisation task. By contrast, Kopinska and Harris (2003) found no effect of eye movements. Indeed, the two studies have different methodological aspects. For instance, Pavani et al. used a setup constituted by external speakers (i.e., free field sounds) in different positions while Kopinska and Harris used intracranial sounds delivered through headphones. To possibly disentangle these two controversial results and clarify the role of eye movements in a spatial acoustic discrimination task, in the study reported here participants were presented with different type of sounds (i.e., free field vs. intracranial sounds) in two experiments by using the same experimental procedure (within subject design).

The second study will investigate a recent described audiovisual illusion. Hidaka and colleagues (Hidaka et al., 2009) demonstrated that a visual stimulus blinking at a fixed location is perceived to move laterally when its flash onset is synchronized to an alternating left-right sound. The authors called this effect ‘sound induced visual motion’ (SIVM) an illusory effect in which the auditory signal can induce motion perception of a static visual stimulus. Hidaka and colleagues (2009) asked to the participants to report the motion direction of the visual stimulus (left or right) and measured the participant's response bias towards the position of the sound. Starting from Hidaka and colleagues' study, in addition to measure the perceptual response bias, the aim of the reported study was to check whether eye movements landing positions towards the illusory visual position were also biased by the sound.

Finally the last study will present data of three experiments purported to investigate the possibility of using auditory information to induce a visual detection improvement in patients with deteriorated visual functions not caused by brain injuries (i.e., patients suffering from low vision). The enhancing effect of a sound paired with a visual stimulus already emerged in other studies with normal sighted participants (Andersen & Mamassian, 2008; Doyle & Snowden, 2001; Hairston et al., 2003a; Marks et al., 1986; Frassinetti et al., 2002a) and patients with visual deficit (Frassinetti et al., 2002b, 2005). Results of three experiments in which spatial and temporal disparities between audiovisual stimuli has been manipulated will be described.

Eye movements and sounds

Background introduction

Humans and animals in general quickly learn to ignore irrelevant stimuli (e.g., sounds) in their environment, while ‘motivation to survive’ avoids missing those stimuli cues that provides information about dangerous situations (e.g., predators). Beyond the evolution of this important behaviour there are subtle mechanisms in which different sensory information interact to provide an appropriate outcome like, for instance, to gaze towards the external spatial position of an acoustic cue. Indeed, auditory and visual modalities are strongly linked in spatial localisation processes (e.g., Dufour, Després, & Pebayle, 2002). Along the primary sensory pathways, signals coding the spatial location of auditory, visual and somatosensory targets are based on distinctly different coordinate systems. For instance, the location of a sound source can be deduced from differences in monaural spectral cues, in sound arrival time (i.e., interaural time difference) and pressure level (i.e., interaural intensity difference). Acoustic localisation cues are affected by the position of the head and external parts of the ears (Blauert, 1997 for review) and thus acoustic location is processed in head centred coordinates. Visual stimuli location instead, is processed into retinotopic coordinates (Warnking et al., 2002). These different coordinate systems converge firstly in the superior colliculus (SC). The SC is also a key structure in binding different sensory modalities and for this reason accurate sound localisation can also enhance responses to visual scenes by means of multisensory integration (e.g., Calvert, et al., 2004; Knudsen & Brainard, 1995; Spence & Driver, 2004; Stein & Meredith, 1993). Indeed, the SC is structured with different layers to which converge visual, auditory and somatosensory information.

Neurons in the superficial layers are purely visual while those in the deep layers are often characterized by bimodal visual-acoustic or visual-somatosensory inputs but may be even trimodal. Visual, acoustic and somatic spaces are topographically represented in the intermediate and deep layers (Middlebrooks & Knudsen, 1984). Multisensory neuron's receptive fields (RFs) overlap, and for this reason multisensory cells respond to stimuli from the same external region of space despite the sensory modality input. Moreover, RFs are arranged to provide a functional map of the external space so that SC's multisensory neurons RFs are in 'spatial register' (Meredith & Stein, 1990). SC has also a pivotal role in saccade generation (e.g., Stein & Clamann, 1981). Neurons in the deeper layers of the SC are visuo-motor as they exhibit small visual bursts before the actual motor bursts. They are also topographically organized: the neurons discharge before a saccade directed to a specific region of the visual field which is in register with the visual fields of the visual neurons situated in the superior layers. This organization pattern characterizes the SC 'premotor map' allows transducing different sensory cues into common motor responses with the goal to gaze towards the spatial position provided by one or more sensory information that take place in the surrounding environment. Studies concerning different modalities highlight their relation with SC's multisensory motor neuron involved in saccade generation like for instance movements of cat's pinnae (Stein & Clamann, 1981), and saccades towards somatosensory targets (Groh & Sparks, 1996b). Jay and Sparks (1984) highlighted a relation between modalities to constitute a common external spatial map of the environment at the level of the SC. The authors found that the auditory receptive fields in the SC shifted with changes in eye position even when the head and ears remain stationary. Their hypothesis was that this might be a mechanism allowing auditory and visual maps to remain in register. More recently, modulations of neural responses to sound location as a function of eye position have been confirmed in the SC (Groh, Kelly & Underhill, 2003; Groh, Trause, Underhill, Clark & Inati, 2001; Zwiers, Versnel, Van Opstal, 2004).

Beyond the high relevance of the convergence of different sensory information to the SC cortical structures are also involved in constituting a common spatial frame of reference for different sensory modalities. Guo and Li (Guo & Li, 1997) found an effect of eye position on half of the neurons in the primary visual cortex (V1) of the monkey suggesting that such modulation is used to constitute a head-centred frame of reference from a retinotopic input. A modulation of neural responses to sound locations as a function of eye position has also been described within the auditory cortex (Werner-Reiss, Kelly, Trause, Underhill & Groh 2003). Translation of visual stimuli to a head-centred frame of reference may be at the basis to the convergence of visual and auditory stimuli to a shared spatial map. Indeed, with each eye/head movement, stationary objects in the world change position on the retina and thus to perceive the world as stable retinotopic coordinates as to be remapped (e.g., Merriam, Genovese, Colby, 2007; Nakamura & Colby 2002). Russo and Bruce (1994) provided evidences that in primate frontal eye fields, the neural response related to saccades towards auditory target is modulated by direction of gaze starting position.

Stricanne and colleagues (Stricanne, Andersen & Mazzoni, 1996; see also Bremmer, Pouget, Hoffmann, 1998, for similar findings) provided evidences that neurons of the lateral intraparietal association area (LIP) of the monkey, known to be involved in planning saccades towards visual stimuli, seem to process spatial cues independent to the modality involved (i.e. either visual or auditory). Indeed, this was one of the first studies to demonstrate that the auditory receptive fields of LIP neurons can be coded in an eye-centred reference frame. The authors suggest that area LIP may be involved in transformation of auditory cues in an eye-centred reference frame for visuo-motor responses. Middlebrooks and colleagues (Middlebrooks, Clock, Xu &

Green, 1994) found that single neurons in the cat's anterior ectosylvian sulcus cortical area (AES) can code for sound locations throughout 360° of azimuth.

Beyond the physical evidence regarding a common frame of reference between different sensory modalities and the relation with eye movements, extensive literature highlights the relation between visual attention (e.g., Chelazzi, Biscaldi, Corbetta, Peru, Tassinari, & Berlucchi, 1995; Crovitz & Daves, 1962; Hoffmann & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995; Posner, 1980) and auditory attention (Rorden & Driver, 1999).

In light of this background, to deepening the understanding of the relation provided by common mechanisms between eye movements and audition, two experiments will be presented.

In the first study, the effect of a forced eye movement in an acoustic discrimination task was investigated. The aim of this study was to try to disentangle results of two studies that report conflicting results. In particular, Pavani and colleagues (2008) did found a significant effect of a forced eye movement in an acoustic spatial discrimination task. More precisely, while eccentrically fixating a point at 25 degree on the left or on the right, participants heard a first sound. The position of this sound was kept as target for a subsequent comparison. In between the two sound participants were either asked to keep the same eccentric fixation (i.e., static eccentric condition) or to switch to the opposite side (i.e., eye movement condition, from left to right or vice versa). Then, after around 2.5 seconds (thus after that the eventual eye movement was terminated) a second sound was delivered either in the same external spatial position or in a different position (i.e., 10 degrees on the left or right in respect the position of the target). Participants’ task was to decide whether the first and the second sound came from the same external spatial position or not (i.e., same/different task). Results highlight a significant decrement of the performance when participants made an eye movement but only when the apparent motion of the sound (i.e., different trials in which the sound moved from one spatial position to another) was in the opposite direction in respect to that of the eye movement. Conversely, Kopinska and Harris (2003) by using a different experimental paradigm reported no evidence of an eye movement effect in coding the sound location. They asked participants to remember the location of an auditory target presented intracranially through headphones, and then move their eyes, their head and eyes, or their body (beneath a stationary head). Sounds' spatial virtualization was obtained by varying the binaural intensity difference (for instance, a sound in front of the participant at 0° of azimuth was obtained by providing the same intensity to the two ears). Afterwards, participants were asked to repositioning (by adjusting the binaural intensity difference) an auditory probe to match the remembered location of the previous target. Sound localisation was modulated by head-on-body position, but they did not found effect of eye movements alone. However, it is worth noting that in their study Kopinska and Harris (2003) presented all sounds intracranially through headphones. Thus, difference in spatial sounds' cues (i.e., free field sounds vs. intracranial sounds) and difference in the two tasks paradigms highlight the need to disentangle such difference in the results. To this aim, all participants in the presented study underwent with both the experimental conditions. That is, the same experimental procedure used by Pavani and colleagues (2008) was used with both free field and intracranial sounds lateralised by varying only the interaural intensity difference like in Kopinska and Harris' (2003) study.

In the second study related to eye movements and sounds the aim was to further understand the relation between perception and action. More precisely, the purpose was to verify how the perception of an illusory movement (i.e., left /right) of an actual static visual stimulus induced by an alternating sound (i.e., left /right) might affect eye movements towards the position of the visual stimulus. In Hidaka and colleagues'

(2009) work the authors provide evidences of the existence of a spatial illusory visual movement provided by a moving sound. The authors called this effect sound induced visual motion (SIVM). While fixating a point participants were provided with a flickering static vertical bar on the visual field periphery. In some trials, the onset of the flickering bar was synchronously coupled with an alternating sound (left/right) provided through headphones. In this conditions participants' perception of the static bar position was illusorily biased towards the actual side of the sound (i.e., left/right movement). Results highlight the effect of the illusory SIVM effect on participants' judgment of the visual stimulus position towards the side (i.e., left or right) of the last occurring sound.

The aim of this study was to replicate Hidaka and colleagues' (2009) findings by asking participants to report a left/right judgment about the spatial position of the bar. Moreover, to check the relation between perception and action in such illusory condition, the same participants underwent a second experiment in which he/she was asked to gaze towards the position of the last 'illusory' moving bar while eye movement were recorded.

Results of these two experiments will be discussed to deepening the knowledge of the relation between eye movements and sound localisation.

Eye movements and spatial acoustic cognition

Thi work was done in collaboration with Alessio Fracasso and Professor Francesco Pavani.

Introduction Experiment 1

Experience plays a critical role in establishing and maintaining congruent visual-auditory associations, so that the different sensory cues associated with targets that can be both seen and heard are synthesized appropriately. For stimulus location, visual information is normally more accurate and reliable and provides a reference for calibrating the perception of auditory space (King, Parsons & Moore, 2000). The relation between vision and audition and the transformation between one and another spatial coordinates system relies on complex mechanisms (for a review on this topic see King, 2009, for a review on this topic). In humans, acoustic localisation cues refer to a head-centred reference frame because of interaural time and intensity difference is related to ears' position. Therefore, any possible change of head position will affect auditory localisation. In the present study, it will be investigate whether changes of just the position of the eye may also affect auditory spatial cognition.

Accurate eye movements towards sounds require a coordinate transformation of the acoustic target into eye-centred motor commands, which necessitates information about eye position in the head (Jay & Sparks, 1984, 1987). Furthermore, in everyday life, eye and head positions change continuously, both relative to the target sound and to each other. To ensure accurate acoustic orienting of eyes and head, the motor system should account for these changes. A few studies used double-step visuo-auditory paradigms, in which participants were instructed to make a combined eye-head movement towards an auditory target, after a previous eye-head shift towards a visual target (Goossens & Van Opstal, 1999; Vliegen, Van Grootel, & Van Opstal, 2004). Regardless of whether the target sound was presented before an eye-head movement

(Goossens & Van Opstal 1999; Vliegen et al. 2004), or during the first eye-head shift towards the visual target (Vliegen et al. 2004), participants were accurate to point with head-gaze movements towards the sound. Presumably, the intervening eye and head movements were on average fully compensated for also when the sound was presented during the eye-head movement. This result suggests that the brain is capable of constructing a stable representation of auditory space across eye and head movements, at least when planning and executing sound-directed spatial motor actions (i.e., a motor-related representation). Vliegen and colleagues (2004) suggest that the human auditory system is able to process online the dynamically varying acoustic cues that result from head movements. Results show that the eyes and head are both driven by a common gaze-displacement signal, created by integrating head-centred target location with head orientation in space, with the goal to construct a stable representation of the target in world coordinates. This signal is subsequently translated in their own appropriate reference frame in two different motor commands for the eyes and the head to accurate localisation motor responses.

In contrast with these results of motor response to auditory stimuli (i.e., eye movements), psychophysical studies in humans that have examined the consequences of deviated static eye position (i.e., orbital deviation) on sound localisation highlight the presence of miss-localisation of the sound spatial position. Lewald and colleagues (e.g., Lewald, 1997, 1998; Lewald & Ehrenstein, 1996a, b) provide useful information regarding the effects of statically deviated eye position on auditory space perception. When eyes were directed to one eccentric side, participants reported consistent shifts of apparent sound location in the opposite side on the order of 2-3 degrees of azimuth in magnitude. Results are consistent through various types of presented sounds (e.g., dichotic or free field) and different measures of sound localisation (e.g., by pointing to sounds or with the verbalization of the position in respect to the egocentric midline or a central visual reference). For instance, in a study in which participants were instructed to aim at a central sound source using a hand-controlled swivel-pointer rotating on a central axis, participants pointed further to the right of the actual location of the sound when their eyes were directed in a fixed manner towards the left (Lewald, 1998). On average, the mean sound localisation error in respect to the left or right of the body midline was 2.7 degrees to the right when eye position was 45 degrees to the left, and 3.5 degrees to the left when eye position was 45 degrees to the right (Lewald, 1998). In some cases interpretation of related findings on effects of static eye deviation can be complicated by the possibility that eye deviation might affect some reference point that is required for the requested judgment (e.g., subjective midline, or central visual reference; cf. Lewald & Ehrenstein, 2000), rather than affecting only sound perception per se. Indeed, Doufour and colleagues (2002) provided evidences that despite the type of stimulus used as reference (i.e., a visual or an auditory stimulus positioned straight ahead to the participant fixation) acoustic localisation improves significantly in respect to the ‘no reference’ condition. Nevertheless, when taken together, these studies do indicate that static eye deviation towards one particular eccentric side tends to shift perceived sound location slightly in the opposite direction, particularly when eye position deviates substantially from sound sources located near the medial plane (e.g., Lewald, 1998; Lewald & Ehrenstein, 1998a, b).

In another study, Kopinska and Harris (2003) measured the lateralization perception of dichotic auditory targets in a contest in which eye and head position changed dynamically during the task. They asked participants to remember the location of an auditory target presented intracranially through headphones by modulating the interaural intensity difference. After presentation of the first sound, participants were asked to move their eyes, or their head and eyes, or their body (beneath a stationary head), before repositioning the memorized sound location in the head by adjusting the binaural level difference. Results show that sound localisation was modulated by head-on-body position (i.e., head orientation on the body, and body orientation with the head fixed), but no by the eye movement alone. The authors concluded that acoustic stimuli are expressed in a body-centred frame of reference.

Pavani and colleagues (2008) examined further the role of eye movements in an acoustic task. The authors presented sounds pairs in free field while participants had to maintain the head fixed and fixate 25 degrees either to the left or right of the egocentric midline. Each trial started with the eccentric fixation on one of the two sides and then the first sound was delivered. Only after the offset of the first sound participants were requested to either keep the same eccentric fixation or to move their eyes to the opposite eccentric fixation. The second sound was delivered with a 2.4 seconds SOA, thus allowing participants to easily complete the eye movement before the second sound stimulus was presented. The second sound was played either from the same speaker as the first (i.e., same condition) or from a different speaker (i.e., different condition). The participant’s task was to produce a same/different response. Results showed that performance significantly decreased in all the eye movement conditions (i.e., same or different). Moreover, performance decreased further in the ‘different’ trials when the apparent motion of the sound occurred in the direction opposite the eye movement, with respect to the when the apparent motion of the sound and eye movement directions corresponded. This study described a highly specific effect from intervening eye movement during an acoustic task that required a same/different perceptual comparison (across a delay) rather than a spatial motor response towards the sound spatial location. As mentioned earlier, evidences in the literature report an acoustic spatial miss-localization in the direction opposite to the deviated gaze (e.g., Lewald & Ehrenstein 1998a, b; Lewald 1998). For this reason, Pavani at al. (2008) expected to found a worst performance when the sound apparent motion moved in the same direction of the eye movement. Indeed, in this case, the representation of the actual spatial disparity (i.e., 10 degrees in the experiment) is reduced by the summation of the miss-localization provided by the deviated gaze in the two opposite fixations. This, in turn, render the judgement more difficult in respect the condition in which the sound moves in the direction contrary to the eye movement because in this case the representation of the actual disparity increased. Contrary to their prediction, Pavani and colleagues (2008) found a greater reduction when the sound apparent motion moved in the opposite direction in respect the intervening eye movement. The authors explained their results in terms of an effect of the eye movement on the internal representation of the first auditory stimulus location.

The aim of the present study is to disentangle the role of eye movement in spatial acoustic perception firstly because of the contrasting results reported in the previously described studies (i.e., Kopinska & Harris, 2003; Pavani et al., 2008). One reason could be that different paradigms (e.g., pointing motor response vs. repositioning the remembered sound position or the way the sounds were presented, that is, free field vs. intracranially through headphones) have been used. Second, only Kopinska and Harris study isolated eye from head movements, however, all sounds were presented intracranially through headphones. As suggested by Pavani and colleagues (2008) it is possible that an impact of an eye movement on sounds spatial representation might be more pronounced for free field auditory stimuli than for intracranial auditory stimuli.

As mentioned in the introduction of the present study, the relation between vision and audition and the transformation between one and another spatial coordinates system relies on complex mechanisms in which experience seems to play a critical role in establishing and maintaining congruent visual-auditory associations (King, 2009). Our everyday experience is highly associated with ecological external acoustic cues that share the same spatiotopic map since low-level structures as the superior colliculus which is also involved in saccade generation (Jay and Sparks, 1984, 1987). Despite the lack in the auditory domain, there is evidence of peri-saccadic bias effects on visual cognition (e.g., Ross, Morrone, Goldberg & Burr, 2001), and trans-saccadic effects when pointing to visual targets after an eye movement either with the arm or with the gaze (Henriques & Crawford 2000; Henriques, Klier, Smith, Lowy & Crawford, 1998).

In light of this background the next study will try to disentangle different results obtained in Pavani et al. and Kopinska and Harris' studies. To this aim, the same experimental procedure used by Pavani and colleagues (2008) was used with both free field and intracranial sounds like used by Kopinska and Harris (2003). All the participants underwent with both the experimental conditions (free field vs. intracranial sounds) in two different experimental sections. Due to evidence of the effect of the deviated static gaze founded also for dichotic acoustic stimuli (e.g., Lewald & Ehrenstein 1996a, b; Lewald 1997, 1998) is reasonable hypothesise an eye movement effect with intracranial sounds (if any), when the apparent motion of the sound is in the direction of the saccade.

Material and methods

Participants

Twelve participants took part in the study (2 left-handed, 6 female; mean age 27 years; range from 19 to 46 years). Participants, mostly students of the university of Trento, reported normal hearing and had normal or corrected-to-normal vision. Participants were naïve as to the purpose of the experiment and varied in their previous experience with psychophysical testing procedures. The experiment was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki (most recently amended in 2008, Seoul), as well as the ethical guidelines laid down by the University of Trento. All participants gave their informed consent prior to their inclusion in the study.

Apparatus and stimuli

The apparatus consisted of a semicircular plastic structure (64 cm length) covering around 52 degrees of visual angle when positioned at 70 cm from the participants. Five loudspeakers were mounted on this structure at ±20, ±10 (negative values indicate that the loudspeakers was on the left of participant's body midline, whereas positive values indicate that the loudspeakers on the right of participant's body midline) and 0 degrees (see Figure 1, p. 50). Furthermore, at the two ends of the semicircular structure at ±25 degrees were mounted two light emitting diodes (i.e., two-colour LEDs, with a diameter of 0.4 mm) that could be switched between green and red colour. The two LEDs acted as eccentric left and right fixation positions. The entire semicircular structure was covered with an acoustically transparent black curtain to avoid the visual information about the number and position of the speakers, while allowing vision of the LEDs. A laptop PC (Dell Precision M6300) and a Matlab script (The MathWorks, Inc.) were used to deliver the stimuli and collect participant's response. Participants provide the responses by pressing the two main buttons of a mouse connected to the laptop. The same setup was used both in the condition with free field sounds (i.e., by using the 5 external speaker mounted on the semicircular structure) and in the condition in which the sounds were delivered intracranially through headphones (in this case, only the two external fixation at ±25 degrees were used). Both free field and intracranial auditory stimuli consisted of the presentation of 251 ms white noise burst (60 dB as measured from the participants' head position). A sine wave

(25Hz) was superimposed to the white noise signal to obtain a 5 peaks of white noise burst to increase the spatial cues to sound localisation (i.e., 15 ms of sound signal and 44 ms of silence). By using the integrated sound card of the laptop computer connected to an external loudspeaker the sounds' signal were amplified. The amplified signal was then switched between 5 relays (NEC MR62-4.5 USB) by using the digital outputs of a data acquisition boards (National Instruments, NI USB 6259). The same data acquisition board was used to control the two LEDs at the eccentric fixations. By activating the desired relay, the sound signal was played by the associated speaker. Speakers were round-shaped (5 cm diameter of Mylar; Pro Signal ABS-210-RC range 350-20000 Hz, 8 Ω, 1 W RMS Power).

The sounds provided through headphones were lateralised by varying the intensity of the two headphones' channels. For example, to obtain a sound source position on the centre (at the position 0 degrees), the same sound intensity was delivered to the two ears (i.e. 58 dB). To lateralise the other four experimental conditions (i.e., ±10 and ±20 degrees) intensity on one ear was decreased while on the other ear intensity was increased by the same amount (see Table 1, p. 54). To obtain the right parameters for sounds delivered intracranially through headphones, a pilot study to equal the difficulty for the two tasks (i.e., intracranial vs. free field sounds) was run (see Pilot study p. 53 for more details).

Participants' eye movements were recorded by means of an eye tracker (EyeLink 1000 Desktop Mount, SR Research).

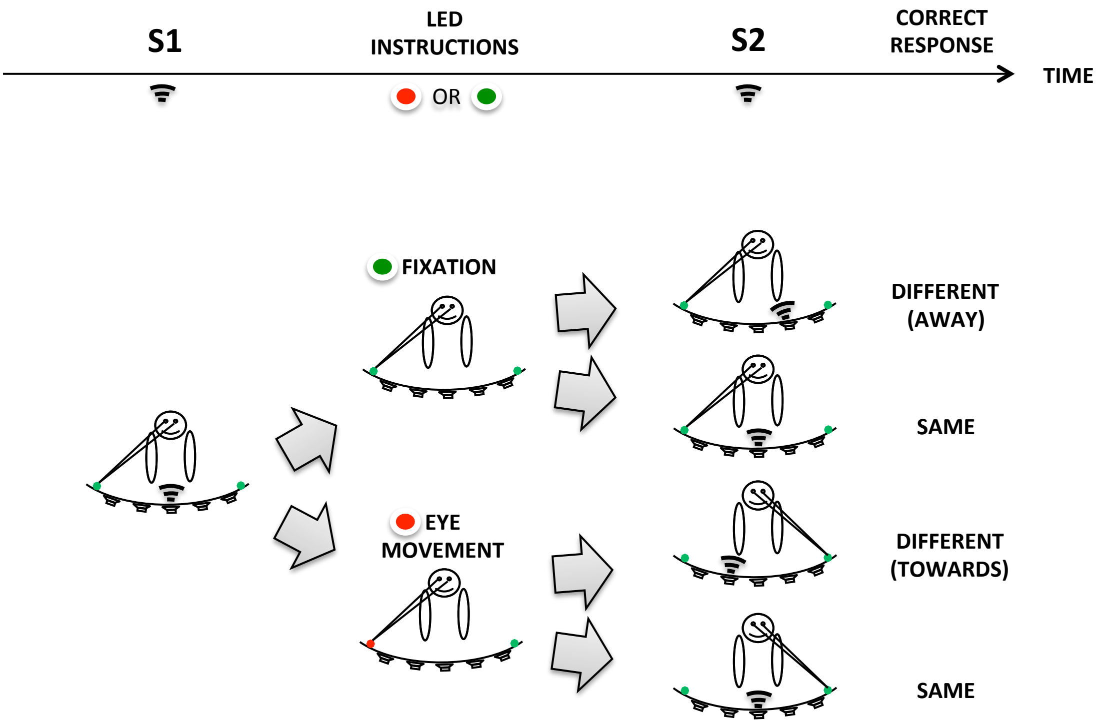

Figure 1 Schematic representation of a trial. In the reported example the first sound (S1) is played in central spatial position. Then after S1 offset, the colour of the fixation LED instruct participants whether to keep the gaze either on the same previous fixation or gaze towards the opposite fixation. At this point, the second sound (S2) is played either in the same or a different spatial location. In the example, the correct participants' possible responses are reported for the 'same' or both the 'different' conditions. The two possible 'different' conditions are related to the direction of the apparent movement of the sound either 'away' from, or 'towards' to, the initial fixation position. The same procedure was used with intracranial sounds but the sounds were delivered through headphones.

Procedure and experimental design

Participants sat at 70 cm in front of the semicircular structure in a dimly lit room (average luminance 30 cd/m²). At the start of each block, participants were asked to undergo a brief calibration procedure to record the two fixation positions indicated by the two LEDs at the two ends (i.e., left and right) of the setup that have to be fixated during the experiment. During this calibration, participants were asked to keep their head straight ahead towards the centre of the apparatus, rest their chin on a support and move only their eyes first towards the left fixation at -25 degrees and then towards the fixation on the right at 25 degrees. Participants were instructed to maintain fixation on each LED until the experimenter recorded eyes position. Each fixation position coordinates were obtained using the average position recorded during 1000 ms of eye recording. Afterwards, a tolerance rectangle of 4 by 4 degrees of visual angle was computed around each fixation position using the obtained average fixation coordinates (i.e., X and Y), to allow for small deviations from the expected positions.

During the actual experimental task, each trial started with the two LEDs turned on with a steady green colour. Then, one of the two green LEDs was turned off for 1000 ms and then turned on again to instruct the participant to gaze towards that fixation position. Participants had to maintain the fixation on that LED until other instructions were provided. At this point, if the participant was actually looking towards the indicated fixation the target sound (i.e., S1) was played. This control was achieved by using an online procedure to check whether the effective participant's gaze position was inside the tolerance rectangle during the 500 ms before S1 onset. S1 could be delivered from any of three different spatial positions at -10, 0 and 10 degrees. Then, 1000 ms after the offset of S1, the participant was either instructed to maintain the same fixation position or to move the eyes towards the fixation point on the other side of the apparatus. To instruct participants to maintain the same fixation, the green LED (that the participant had to fixate) was turned off for 100 ms and then turned on again. By contrast, to instruct the participants to move their eyes to the opposite side, the green LED was switched to the red colour for 100 ms and then green again. After further 1400 ms, the test sound (i.e., S2) was delivered only if the participant was able to follow the instructions (i.e., by moving the eyes or by keeping the same fixation). If the participant failed to follow the instructions the trial was repeated from the beginning. S2 had the same physical characteristics (e.g., duration, number of peaks) as S1 and was delivered from either the same spatial position as the target sound (i.e., the same physical speaker) or at ±10 degrees with respect the location of S1 (i.e., the closest speaker to S1 on the left or right). It is worth noting that due the online control gaze procedure and the very wide temporal interval between S1 and S2, sounds were always delivered when participants had a stable fixation and never when they were moving the eyes. After the offset of S2, the participants' task was to report whether the sound came from the same spatial position as S1 or not by pressing the left or the right button of the mouse (i.e., same/different task). Participants handled the mouse horizontally with one hand while responses were provided with the other hand. The hand used to provide the response was balanced between participants in a way that half of the participants used the left index to respond 'same', the test sound originated from the same position as the target sound and the left middle finger for the 'different' response, while the other half used the right middle finger for the 'same' response and the right index finger to respond 'different'. Participants were also stressed to concentrate their judgment on spatial cues thus avoiding just intensity discrimination while enhancing the spatial discrimination task.

The experimental design comprised 36 different conditions: 3 different target (S1) positions (-10, 0, +10 degrees), 3 different test (S2) positions (same as S1, 10 degrees left, 10 degrees right), 2 gaze conditions (fixation or eye movement) and 2 starting sides (left or right). Initial fixation side of each block was randomised between left or right fixation but, when an eye movement to the other side was requested, the subsequent trial had to start by fixating the position on which the eyes were moved to. That is, if the first trial started with the fixation on the left side, and then an eye movement was requested towards the right side, the subsequent trial always had to start with the fixation on the right side and vice versa. This was thought to limit the occurrence of the shifting of the gaze positions from one fixation to the other as adopted by Pavani and colleagues (2008).

To verify that the procedure was clear, before the actual experimental section, the participants were asked to undergo a brief practice (12 trials).

The same procedure was used with the sounds delivered intracranially through headphones. That is, participants had to look towards the same apparatus while wearing the headphones, thus, calibration, instructions to maintain the fixation or move the eyes (i.e., LEDs colours coding) and response collection were the same.

The whole experiment was divides in two experimental sections, each constituted by 4 blocks for each of the two sound conditions (i.e., free field and intracranial sounds) and run in two different days. Each block consisted of 72 trials (i.e., 36 different conditions repeated twice presented in random order) for a total of 288 trials. These consisted of 96 ‘same trials’ (S1 and S2 presented from the same speaker) and 192 ‘different trials’ (S1 and S2 presented from different speakers). S1 and S2 position were unpredictable on every trial, and the sequence of ‘same’ or ‘different’ trials was randomised between participants. Each experimental session (i.e., 288 trials) lasted around 45 minutes with a rest between blocks.

Pilot Study

To compare performance on free field and intracranial sounds when participants were asked either to maintain fixation or to perform an eye movement, the two tasks were preliminary equated in terms of sensitivity to provide a stable baseline condition and avoid possible confounds in terms of differences between the two sounds conditions. To this aim, a pilot study was run on 7 participants aimed to find the right parameters from which the correspondent value in dB for each ear was derived and used to lateralize intracranial sounds’ spatial positions.

The procedure was identical to the one described above with the following exceptions. The pilot experiment comprised two blocks of 72 trials, one with free field sound and the other with intracranial sounds (task order was randomised across participants). In this pilot study only performance in trials at the two eccentric fixations were equated. The overall proportion of correct response (i.e., same and different responses) was computed and compared between the two experimental setups (i.e., free field vs. intracranial sounds).

The chosen step that equated performance on the two tasks was around 0.8 dB/degree and was kept fixed for the intracranial condition for all participants in the experiment (see Table 1 p. 54).

	-20°	**-10°**	**0°**	**10°**	**20°**
Right (dB)	44	51	58	65	72
Left (dB)	72	65	58	51	44

Table 1 Values obtained in the pilot study referred to the sound intensity (dB sound pressure) for the spatial 'virtualization' of intracranial sounds in relation to the 5 experimental spatial positions.

Data analysis

Participants' performance in detecting changes in auditory spatial location both for 'fixation' and 'eye movement' conditions where examined.

As in Pavani et al. (2008) study, different location trials in which S2 appeared at a location towards initial fixation than S1 itself (i.e., S2 shifts towards initial fixation), and different location trials in which S2 appeared at a position further away from initial fixation than S1 (i.e., S2 shift away from initial fixation) were analysed separately.

The dependent measure was sensitivity (*d'*). Sensitivity was computed using signal detection procedures modified for same/different designs (Macmillan &

Creelman, 1991). Sensitivity was calculated based on the hit rate for detecting a sound location change (i.e., 'different' response produced in a true different location trial), separately for each specific S1 location (i.e., at -10, 0 and 10 degrees), and the false alarm rate for the corresponding S1 location (i.e., 'different' response produced when the two sounds were in fact presented from the same position).

Results

A within participants' analysis of variance (ANOVA) was performed on the d' data, with three factors namely: gaze condition (i.e., fixation or eye movement), direction of auditory change on different trials (i.e., S2 shifting towards, or away from, initial fixation) and sound condition (i.e., free field vs. intracranial sounds) as variable. The results showed a significant difference between gaze conditions F (1, 11) = 20.4, p < .001 indicating that there was a cost in the participants' ability to respond caused by worse overall performance when the eyes moved compared to when they remained fixed. A significant interaction between direction of auditory change and sound conditions was also found F (1, 11) = 15.6, p < .01 indicating that direction of auditory change influenced participants performance differently in the two sound condition. The graphic representation (see Figure 2 p. 56) shows that for free field sounds there is a cost in the different trials when S2 moves toward the initial fixation, whereas for intracranial sounds the effect is surprisingly reversed. Finally, the ANOVA showed a three level interaction F (1, 11) = 11.7, p < .01 meaning that the 2 way interaction between direction of auditory change and gaze condition differs in the two levels of sound condition in terms of magnitude.

The lack of a significant effect for the main factor sounds condition provides further evidences that the two tasks (i.e., free field vs. intracranial sounds) were indeed equated.

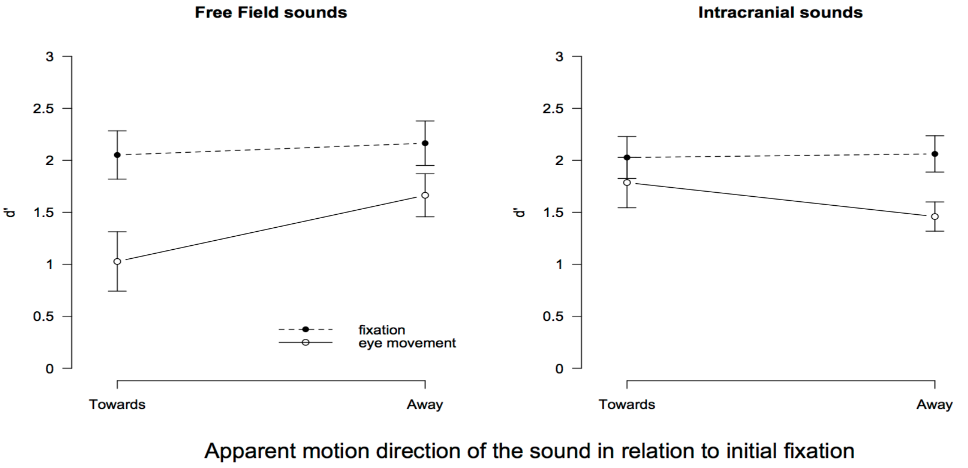

Figure 2 Averaged participants' performance (*d'*) in the two sound conditions in relation to the two 'different' conditions of the apparent motion direction of the sound (i.e., away/towards) and the gaze conditions (i.e., fixation/eye movement). On the left, the performance in free field sounds condition. On the right, the performance in the intracranial sounds condition.

To better highlighting the results showed by the previous analysis of variance, two further ANOVAs were run on the two different sound conditions. The first ANOVA on free field sounds showed a significant difference between gaze condition F (1, 11) = 13, p < .01, confirming that there was a cost in the participants' ability to respond when an eye movement was performed. A significant difference has also emerged between directions of auditory change F (1, 11) = 6, p < .05, highlighting the cost of an eye movement independently from the direction of the auditory change (i.e., towards or away from the initial fixation). Furthermore, the interaction between direction of auditory change and gaze conditions was significant F (1, 11)= 15.6, p < .01 indicating that the cost in the participants' ability to respond when an eye movement was performed was significantly larger when the sound moves towards than when it moves away from the initial fixation. A post-hoc t-test revealed that in comparison to the ‘fixation’ condition, an aye movement affects participants’ response significantly in both the direction of the auditory change (towards, *d’*(fixation) = 2.05 vs. *d’*(eye movement) = 1.03, p < .05; away, *d’*(fixation) = 2.16 vs. *d’*(eye movement) = 1.66, p < .05). The difference between directions of the auditory change in the eye movement condition was also significant (*d’*(towards) = 1.03 v.s. *d’*(away) = 1.66, p < .05).

In line with results of Pavani and colleagues (2008) an eye movement significantly impairs participants’ performance when the sound moved towards the initial fixation. However, unlike Pavani and colleagues’ results, in the present study a significant cost of the eye movement in participants’ performance was found also when the sound moved away from the initial fixation.

The ANOVA on intracranial sounds showed a significant difference between gaze condition F (1, 11) = 12.9, p < .01, confirming that there was a cost in the participants’ ability to respond when an eye movement was performed. However, no difference between directions of the auditory change has been found (i.e., towards vs. away from the initial fixation, p = .16). The interaction between direction of auditory change and gaze conditions was significant F (1, 11)= 6.4, p < .05 indicating a significant cost in the participants’ ability to respond when an eye movement was performed in the condition in which the sound moved away from the initial fixation. This results contrasts with previous findings of Kopinska and Harris (2003) in which no effect of eye movement was emerged in an acoustic task with intracranial sounds.

Despite to the lack in the literature regarding quantitative effects of an eye movement neither on the spatial localisation with free field sounds, nor for intracranial sounds, some information may be exploited from the visual domain. Different studies report peri-saccadic effects on visual spatial representation (e.g., Ross et al. 2001), as well as studies showing that directional biases may be observed when pointing to visual targets after an eye movement either with the arm or with the gaze (e.g., Henriques & Crawford 2000; Henriques et al., 1998). Henriques and colleagues (1998) measured this trans-saccadic error in pointing or gazing towards the spatial location previously seen, in the opposite direction of the eye movement with a magnitude related to the amplitude of the saccade. The mean reported error is around 13% of the saccade amplitude. In light of this framework, one hypothetical explanation of the different cost emerged in the 'towards' conditions between the two sound conditions, may derive by the summation of the deviated gaze and eye movement effects.

Due to the close relation between the visual and acoustic systems in spatial representation, the first assumption is that a trans-saccadic effect might take place also for auditory spatial representation with a similar magnitude as that observed for the visual spatial representation. If the first assumption is true, the trans-saccadic miss-localisation on auditory representation in the present experiment should be around 6 degrees considering the requested 50 degrees eye movement. The second assumption is that for intracranial sounds there is not (or is very low) trans-saccadic effect on auditory representation. Although there is no specific evidence in the literature, trans-saccadic remapping of sounds may particularly apply for more ecologically acoustic stimuli that have to share together with the visual information a common spatial reference frame (see King, 2009 for a review on this topic). If the two aforementioned assumptions hold, the hypothesized major cost of an eye movement on the acoustic spatial representation in the 'away' condition in case of the sole deviated eye effect is reversed. That is, the higher miss-localisation effect of an intervening eye movement with respect of the deviated gaze, would provide a higher cost in the 'towards' condition. Furthermore, this effect would be present mostly in the free field sound condition as pointed out by the lack of a significant trans-saccadic effect in the 'towards' condition with intracranial sounds. For example, when the participant look to the left fixation (i.e., -25°) and a S1 is presented in the central position (i.e., 0°) the perceived position of S1 is shifted at +2° because of the static deviated gaze effect. Then, the participant is instructed to make an eye movement towards the fixation at +25° (i.e., rightwards eye movement). The trans-saccadic displacement of S1 representation is shifted in the opposite direction of the eye movement and for this reason now S1 is represented at -4°. In the 'away' condition, S2 is presented at +10° but is perceived at +8° because the static deviated gaze on the new fixation. In this case the difference between S1 and S2 positions is 12°. In the 'towards' condition instead, S2 is presented at -10° but because of the effect of the static gaze deviation on the right, S2 is miss-localised toward left at -12°. In this case the difference between the two perceived spatial positions is 8°. At the end, S1 and S2 spatial representations are closer in the 'towards' (i.e., 8°) than in the 'away' (i.e., 12°) condition and for this reason is more likely that participants respond incorrectly 'same' when actually is a 'different' trial in the 'towards' conditions.

Discussion

The aim of the present study was to disentangle contrasting results in the literature on the effects of dynamic changes in eye position on sound localisation. Compared to the deviated static gaze condition performance, the present study provides evidence of the cost of an intervening eye movement in a task in which two sound source positions have to be judged (same/different task) when the stimuli were delivered in free field. The results confirmed the findings of Pavani and colleagues' (2008) in which a significant performance decrement emerged when the sound moved in the opposite direction with respect the intervening eye movement. In the present study, the cost of an eye movement when the sound moved towards the initial fixation was significantly larger with respect that measured when the sound moved away from the initial fixation. Interestingly, a significant cost of an eye movement emerged also in the ‘away’ condition in comparison to the gaze ‘fixation’ condition.

When the acoustic stimuli were presented intracranially, results of the present study reveal a significant trans-saccadic cost in the ‘away’ condition and it does not differ significantly from that in the ‘away’ condition with free field sounds. This results contrast with previous findings of Kopinska and Harris (2003) in which no effect of eye movement was found in an acoustic task with intracranial sounds. However, it is worth noting that three main differences between the present and Kopinska and Harris’ study might have contributed to provide different conclusions. In fact the apparatus, the experimental design and the task’s response differed substantially. More precisely, in Kopinska and Harris’ study, participants were presented with an intracranial sound lateralized by varying the binaural intensity. After an eye movement they had to repositioning the memorized sound location in the head by adjusting the binaural level difference.

The present study has been designed to disentangle the differences between the experimental procedure used in Pavani et al.’s (2008) and Kopinska and Harris’ (2003) studies. The results highlight the presence in both sound conditions of the cost of an eye movement with the same magnitude in the same experimental condition (i.e. when the sound moved away from the initial fixation).

It is important to note that the trans-saccadic effect on sound localisation in the ‘away’ conditions emerged with both sounds conditions (i.e., free field and intracranial sounds) is compatible with the localisation bias of sound sources with static deviated eyes (Lewald and Ehrenstein 1996a, b; Lewald 1997, 1998). These studies provide evidence that for either type of sounds, when participants are asked to localize a centrally presented sound while fixating a peripheral target, usually they tend to report the perceived position of the sound biased towards the opposite direction of the current static deviation. As pointed out by Pavani and colleagues (2008), the static deviation of the eyes would predict a decrease in sensitivity on those trials in which S2 apparently moves in the same direction of the eye movement, thus away from the initial fixation.

For free field sounds, the trans-saccadic magnitude of the cost in the 'towards' condition is significantly larger with respect that in the 'away' condition. For sound presented intracranially instead, the 'towards' condition do not differ in relation to the gaze condition. A possible explanation for the cost of an intervening eye movement emerged in the 'toward' condition only with free field sounds seems to be attributable to a trans-saccadic effect which disrupts the auditory spatial representation of the target sound (i.e., S1). Evidence in the literature shows a trans-saccadic effect on visual spatial representation (e.g., Henriques & Crawford 2000; Henriques et al., 1998). Despite the lack of evidences of such effect in the auditory domain, the assumption is that a trans-saccadic error on auditory spatial representation might particularly apply with free field sounds. Thus, the results with intracranial sounds seem to be explained mainly by the effect of the deviated gaze (i.e. the significant effect only in the 'away' condition). For free field sounds the effect of the static deviated gaze seem to be overcompensated by a trans-saccadic effect on auditory spatial representation of the target sound (S1). The remapping of S1 spatial representation determine a cost in both the direction of auditory change conditions with a greater magnitude of performance decrement in the 'towards' condition (see example at the end of Results p. 55 for more details).

The hypothesis of a higher trans-saccadic effect on free field auditory spatial representation is supported by evidence in the literature of a close relation between visual and auditory spatial representation since the level of the superior colliculus (e.g., Middlebrooks & Knudsen, 1984; King, 2009). The superior colliculus is also a key structure involved in saccade generation (Groh & Sparks, 1996a; Jay & Sparks, 1984, 1987; Stein & Klamann, 1981). Moreover, Pavani and colleagues (Pavani, Husain, Ladavas & Driver, 2004) suggested that visual and auditory spatial cognition in spatial neglect might be more closely tied for extrapersonal space (i.e., free field sounds).

Results of the present study supports Pavani and colleagues' (2008) findings and provide evidence of a significant cost in participants' judgment irrespectively to the sound apparent motion with free field sounds. Furthermore, results extend the eye movement effect also to acoustic spatial representation with intracranial sounds thus contrasting with Kopinska and Harris' (2003) results. As suggested by Pavani and colleagues (2008), it is likely that the auditory sound location (S1) held in memory might be 'remapped' eye-centrically further to the left when a rightward eye movement follows it, or further to the right after a leftward eye movement. The higher magnitude of the eye movement remapping effect seems to compensate and revers the effect of the static deviated gaze. Eye movements can thus influence auditory spatial cognition especially for free field sounds and for judgments that do not require a spatial motor response.

Eye movements in the sound induced visual motion illusion

Thi work was done in collaboration with Alessio Fracasso and Professor David Melcher.

Introduction Experiment 2

Vision provides the most appropriate/precise cues in optimal viewing circumstances regarding stimulus location several studies have demonstrated that the visual modality dominates over the auditory modality when dealing with spatial attributes of multisensory stimuli (e.g., Battaglia, Jacobs & Aslin, 2003; Bertelson & Aschersleben, 1998). The phenomenon of ventriloquism (Howard & Templeton, 1966), in which movements of the mount of a puppet create the illusion that the performer's voice is shifted towards the puppet's moving mouth, is a classical example of how vision influences audition over spatial disparity. Interestingly, crossmodal ventriloquism as described above, does not appear to be confined to situations involving a static visual and auditory event. For instance, adaptation in response to the apparent motion of moving visual stimuli provides a motion aftereffect in the auditory modality as shown by Kitagawa and Ichihara (2002) highlighting the great influence of vision over audition in audiovisual tasks. Multisensory contributions to motion perception are most pronounced when moving stimuli in different sensory modalities occur at the same spatial locations and at approximately the same time following the classical principles of multisensory interaction governing superior colliculus multisensory neurons (e.g., Stein & Meredith, 1993). For instance, in a typical crossmodal dynamic capture study visual and auditory apparent motion streams are presented. Each stream consists of pairs of stimuli presented in sequence from two different spatial locations. The participants' task involves trying to discriminate the direction of motion of the target sensory modality whilst simultaneously trying to ignore the apparent motion of the distractor stimuli presented in the task-irrelevant sensory modality. The results of numerous studies have now shown that although direction discrimination performance is near perfect when the two (i.e., the target and the task-irrelevant) streams are congruent, auditory motion is captured by the direction of visual motion when the direction of the two streams is congruent. Crucially, the direction of the visual task-irrelevant stream leads participants to incorrectly report the direction of the auditory target stream on a certain proportion of trials. However, the sensory bias magnitude is greatly reduced when the motion signals did not share common paths or were not presented at the same time (see Soto-Faraco & Kingstone, 2004, for a review). As with static ventriloquism, this dynamic ventriloquism effect can occur for other modality combinations, for instance, with visual motion capturing tactile motion, and auditory and tactile motion capable of capturing one another (Occelli, Spence, & Zampini, 2009; Sanabria, Soto-Faraco, & Spence, 2005; Soto-Faraco, Spence & Kingstone, 2004a, b).

Recent findings suggest that multisensory interactions involving dynamic stimuli might occur at early processing stages that involve the way in which motion stimuli are experienced. In an fMRI study it has been found a modulatory effect of auditory motion on visual motion cortical area (MT) responses (Alink, Singer, & Muckli, 2008). This might be one of the neurophysiological sources of the inverse ventriloquist effect. Indeed, as suggested by the appropriateness and stimulus reliability models (Ernst & Banks; 2002; Alais & Burr, 2004, see also p. 21 of the present dissertation), as vision provides less detailed spatial cues (i.e., visual cues no more provide the most reliable spatial information), vision will lose his spatial biasing effect (i.e. its ventriloquist effect; Alais & Burr, 2004).

Cognitive factors (i.e., prior knowledge/decisional factors), as proposed by the unity assumption model, may also play an important role in multisensory interactions of motion information (Meyer & Wuerger, 2001; Sanabria, Lupiáñez & Spence, 2007; Sekuler, Sekuler, & Lau, 1997). If one look on a computer screen at two identical objects (e.g. two black discs) moving toward each, overlapping, and then moving apart, two different events can be perceived. Either the objects can be perceived to follow the same trajectory (i.e. to cross each other's trajectory), or else to reverse their direction (i.e. to collide and bounce back). The situation can be made less ambiguous by presenting a sound at the exact moment of overlap (Sekuler et al., 1997). In that case, the number of time that people report that the two objects are bouncing off one another significantly increases. This audiovisual effect clearly shows that auditory information (not necessarily a realistic collision-like sound) can influence the perception of a visual event when presented at a significant moment. In this case, the presence of a sound in the same physical visual stimulation often might provide the perception of a collision, highlighting the presence of a high-level cognitive modulation due to the previous experience (i.e., when two object bounce together the impact produces a sound). Our brain has to integrate spatial and temporal congruent cues into the same physical object while disambiguate incongruent cues. As suggested by the aforementioned models, in case of incongruent cues, the brain seems to keep as reference the most reliable sensory modality. However, the most reliable modality is not always the most appropriate to disentangle the physical incongruence, thus the perceived outcome might be an illusory perception. For this reason, a well discriminable visual stimulus can illusorily alter the perception of an acoustic source (that provides less reliable spatial cues) as in the ventriloquism effect.

Auditory bias effect has also been reported for well discriminable visual stimuli (Shams et al. 2000) in the temporal domain, were temporal acoustic cues are most reliable. Shams and colleagues results highlights the auditory bias effect by inducing the illusory visual perception of two flashes when actually only one flash was provided coupled with two beeps.

In a recent study Hidaka and colleagues (2009) demonstrated that when visual cues are degraded, a moving sound might induce illusory visual motion perception of a static stimulus. This phenomenon has been called sound induced visual motion (SIVM). In the SIVM, a flickering static bar is perceived to illusorily move left and right when its onset is synchronized with an alternating left-right sound source. The SIVM is clearly observed when the bar is presented at least at 10 degrees from a fixation point. Indeed, in the peripheral visual field visual acuity is less reliable than in the fovea and in line with stimulus reliability hypothesis, in such situation auditory information prevails on vision in a way to induce visual motion perception.

In another recent study, Teramoto and colleagues (Teramoto, Hidaka & Sugita, 2010) highlighted the aftereffect of sounds on visual motion perception providing evidences that the aftereffect produced by the visual modality on the auditory modality (Kitagawa & Ichihara, 2002) can be reversed. Teramoto and colleagues (2010) demonstrated that sounds containing no motion or spatial cues could induce illusory visual motion perception for static visual objects. In an adaptation phase, observers were exposed to visual apparent motion for 3 min. In that phase, two white circles were presented in alternation side by side and two different tones (e.g., A and B) were synchronously associated to each visual circle's onset. In the test phase, a white circle was physically presented twice in the same spatial position. Results show that when the onset of the first circle was synchronized to the tone 'A' and the onset of the second circle with the tone 'B', the circle appeared to move in the same direction of the adaptation phase.

In light to the aforementioned evidences, the aim of the second study of the present dissertation is to verify whether participants' visuo-motor behaviour might be biased by the SIVM illusion (as emerged for perception). More precisely, starting from Hidaka and colleagues' (2009) experimental paradigm the purpose is to check how the

SIVM illusion might affect eye movements towards the position of the illusory visual stimulus. Since the fundamental work of Milner and Goodale (1995) the dissociation between perception and action systems is still debated. This study aims to deepening the knowledge on the relation between perception and action by comparing participants' perceptual and visuo-motor responses in presence of the SIVM illusion.

Material and Methods

Participants

Twelve participants (7 female; mean age of 25 years; range from 20 to 46 years) took part in the study. Participants were all students of the university of Trento, reported normal hearing and had normal or corrected-to-normal vision. The experiment was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki (most recently amended in 2008, Seoul), as well as the ethical guidelines laid down by the University of Trento. All participants gave their informed consent prior to their inclusion in the study and were naïve as to the purpose of the experiment.

Apparatus and stimuli

Participants sat at a table in a dimly lit room (average luminance 40 cd/m²), 60 cm from an LCD 22-inch screen (HP Compaq LA 2205 WG at 60Hz) used for presenting the visual stimuli. Their head movements were fully restrained by an adjustable chin rest, which included two lateral, adjustable stands to hold the head rigidly near the temples, and thereby keep head-direction constant. Visual stimuli were characterized by a static flickering white vertical bar (high 10 and width 0.4 degrees of visual angle) with 400 ms duration and 500 ms of stimulus onset asynchrony (SOA) synchronized with an auditory stimulus (a white noise burst, with 75 ms duration) provided through headphones (Sony MDR-XD200). The bar might be presented at three different eccentricities (15, 16 and 17 degrees of visual angle) with respect to the fixation point that might be eccentrically positioned either on the left or right of the display midline (at 4 degrees of visual angle). Fixation point and flickered bar were always on the opposite sides in relation to the display midline (see Figure 3 p. 69). Both, eccentricities values and fixation point side were balanced and varied randomly across trials. The bar might flicker 5 or 6 times (i.e., repetitions) in a balanced fashion across trials. The number of repetitions provided the direction of the illusion for the data analysis (left vs. right). In crossmodal trials, the sound moved from one ear to the other synchronously with the flickering bar for inducing the visual illusion (i.e., the SIVM) in the direction of the sound movement. However, a preliminary pilot study revealed that there was not a clear one-to-one matching between sound direction and bar illusory motion direction. That is, on some trials a left to right sound direction could induce a coherent left to right bar motion, while on other trials the same left to right sound could induce an opposite right to left bar motion. This alternating stochastic direction of the illusion could not be controlled a priori and would have pose a serious problem in the analysis phase. To overcome this burden, a physical displacement of the bar (i.e., physical inducer) was implemented at the start of each trial to match the direction of the apparent motion of the sound. More precisely, from the first to the second bar onset, the spatial position of the bar was physically displaced (i.e., 1 degree of visual angle) towards the side were the first sound would have been played. The first sound was presented simultaneously with the second bar onset. It is important to note that subsequent repetitions of the bar would have been presented always in the same spatial position, without any further displacement. In this way it was possible to match the first bar physical displacement with the first sound apparent direction and to establish a unique expected participant's response for each trial, according to the bar (sound) repetitions (5 or 6, see above).

Irrespectively to the task type (i.e., perceptual or visuo-motor) an eye tracker (EyeLink 1000 Desktop Mount, SR Research) has been used to control participants' gaze on fixation during each trial and/or to record eye motor response at the end of each trial in the visuo-motor task.

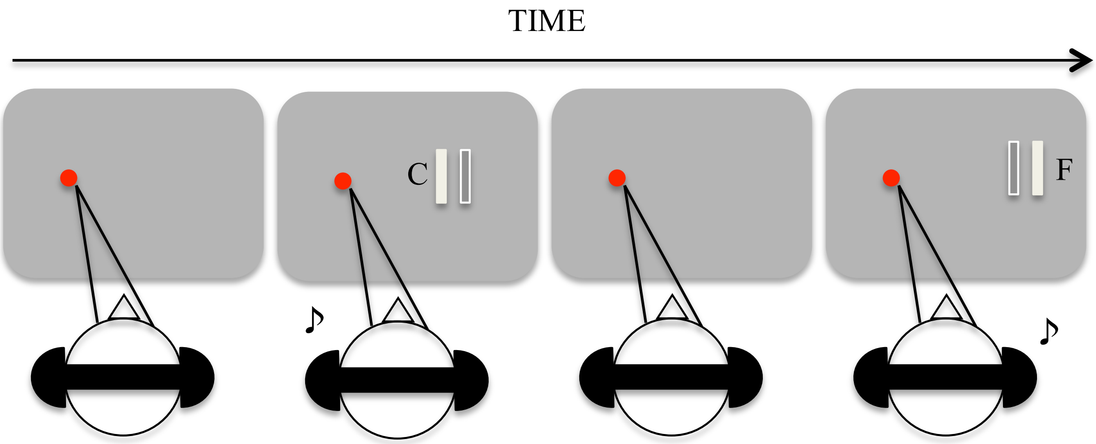

Figure 3 Schematic representation of a trial. In the example, fixation is 4° on the left side of the screen midline. Participants fixated the red dot while the bar flickered peripherally at one of the different eccentricities (i.e., 15°, 16° or 17°) in respect the fixation. The empty bar indicated the actual static position while the white bar represents the perceived illusory position related to the sound side. 'C' and 'F' stand respectively for 'close' and 'far' position in relation to the fixation point. At the end of the bar's flickering repetitions (i.e., 5 or 6), participants were asked to either press the left/right mouse button to report the last perceived direction of the last bar movement (perceptual task), or to gaze towards the last perceived position of the bar (visuo-motor task).
Demo at http://www.rikkyo.ne.jp/web/souta_hidaka/SIVM.html.

Procedure and experimental design

Participants were asked to undergo a calibration procedure for the eye tracker system at the start of each block. Then, participants were instructed to maintain their fixation on the fixation point until the end of the trial. When the participant was actually looking to the fixation, the trial started. The repetitions of the bar (i.e., 5 or 6) were all statically presented in the spatial position corresponding to the eccentricity condition value (i.e., 15, 16 or 17 degrees). In the perceptual task, after the offset of the last repetition of the flickering bar, participants were asked (by showing a message on the screen) to press a left or right mouse button to report the last perceived movement direction of the bar. Participants provided the response by using their dominant hand. In the visuo-motor task, in which an eye movement was requested, the fixation point disappeared 50 ms before the offset of the last presented bar (see Figure 3 p. 69). The participants were instructed to move their eyes towards the last perceived position of the bar at the offset of the fixation point. The trial was repeated whenever the participants moved their eyes before the fixation offset (or before the presentation of the last bar in the perceptual task).

The experimental design was constituted by 12 different conditions namely, 3 eccentricity positions of the bar (i.e., 15, 16, or 17 degree), 2 fixation positions (i.e., 4 degrees either on the left or on the right from the screen midline) and 2 flickering repetition values (i.e., 5 or 6 repetitions). Conditions were repeated 4 times on each experimental block. Four blocks with a perceptual response and four blocks with the eye motor response constituted each session (i.e., 192 trials). Each participant underwent sessions with or without sound (i.e., sound condition) in two different days (i.e., two sections with a within participant experimental design). Response conditions within the same day (i.e. perceptual or visuo-motor) and sound/no-sound sessions sequence in different days were balanced across participants. To verify that the procedure was clear, before the actual experimental section the participants were asked to undergo a brief test (12 trials).

Data analysis

For the perceptual task, proportions of ‘far’ responses were computed for sound and no sound conditions. Performance for the two different fixation positions and for the three different eccentricities was averaged. In this way, the left/right coding of participants’ response for the position of the last perceived bar has been coded in relation to the distance between the last position of the illusory perceived bar and the fixation position. More precisely, leftwards movement of the last bar with fixation on the right and rightwards movement of the last bar with fixation on the left were clustered as ‘far’ (because the bar was suppose to be perceived to move away from fixation). Conversely, leftwards movement of the last bar with fixation on the left and rightwards movement of the last bar with fixation on the right were clustered as ‘close’ (because the bar was suppose to be perceived to move towards the fixation, see Figure 3 p. 69 for an example). The same procedure has been used for the analysis of eye movements in the visuo-motor task. In this case, the difference between the executed eye movement amplitude and the expected eye movement amplitude (i.e., the actual eccentricity between the fixation and the position of the static bar) was computed for all experimental conditions. This variable transformation was applied in order to obtain comparable values, independent of the actual bar eccentricity. It is important to note that the ‘far’ and ‘close’ coding of trials holds either for the ‘sound’ or the ‘no sound’ conditions. Even in the absence of sound, the presence of the direction specific physical inducer allowed to derive an expected perceptual as well as motor response, according to the bar repetitions.

Results

An analysis of variance (ANOVA) on the proportion of ‘far’ response in the perceptual task with factor distance (far vs. close) and sound (sound vs. no-sound) show a significant difference between ‘far’ and ‘close’ clusters F (1,11) = 90.9, p < .001, indicating that there was a greater number of responses in the ‘far’ cluster than in the ‘close' cluster. The visual illusion direction was indeed in the direction of the auditory apparent motion as found by Hidaka and colleagues (2009). Furthermore, a significant interaction between the two factors has been found F (1,11) = 25.8, p < .001, indicating that the presence/absence of the sound influenced participants performance differently in the two levels of the distance condition. That is, the visual illusion was greater when the sound was presented as emerged in the graphic representation reported in Figure 4 (p. 73). The main effect of sound condition was not significant (F (1,11) = .01, p = .9).

The ANOVA of the expected landing position of eye movements in the visuo-motor task with factor distance (i.e., far/close) and sound (i.e., sound/no-sound) show a significant difference between ‘far’ and ‘close’ clusters F (1,11) = 13.7, p < .01, indicating that the eye landing position was indeed biased by the direction of the visual illusion. That is, when the direction of the last repetition of the bar was perceived to move away from the fixation, participants overshot the actual static position of the bar. Vice versa, when the direction of the last repetition of the bar was perceived to move towards the fixation, participants undershot the actual static position of the bar. Note that participants had a tendency to undershoot the requested saccadic amplitude, as shown clearly by the negative values on the plot summarizing visuo-motor results. This general tendency can be explained by the nature of the task, participants started to move their eyes after the last bar offset and were actually performing a blind saccade. However, neither a main effect of sound condition nor interaction between sound and distance conditions is emerged. This result indicates that the visual illusion effect is present also in the condition without the sound as for the perceptual response task. The graphic representation (see Figure 4, p. 73) shows the magnitude of the visuo-motor response in the two sound conditions.

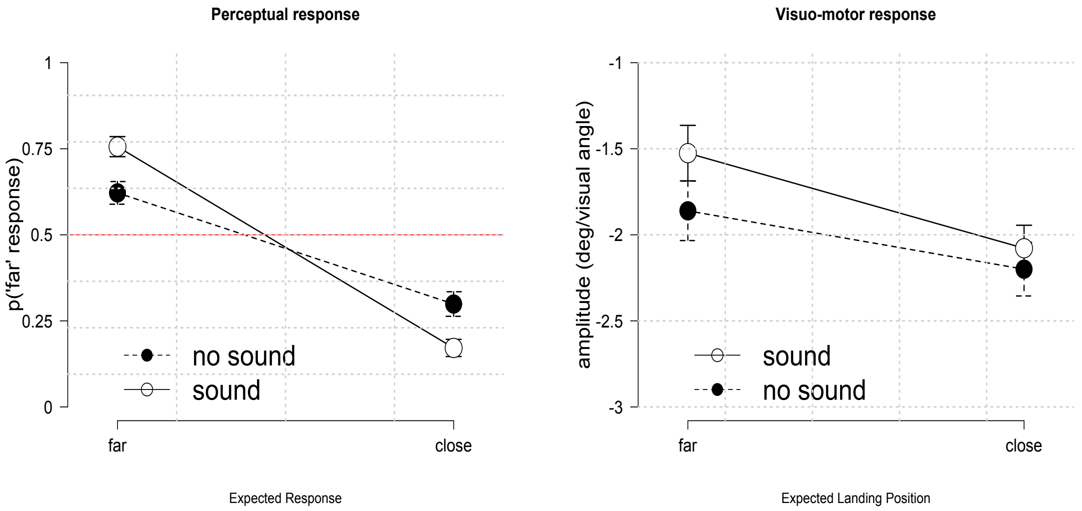

Figure 4 On the left side, the perceptual response. The graph shows the averaged proportion of responses (far) in respect the two clusters of directions (far/close) in the two sound conditions (sound/no-sound). On the right side, the visuo-motor response. The graph shows the averaged amplitude of the expected landing position. Reported values represent the difference between the eye movement landing position and the effective eccentricity of the bar.

A further analysis considering perceptual effects of the sound factor show that the magnitude of the illusory effect in the two conditions is correlated across participants (R = .32, t = 2.2, p = .05). That is, the magnitude of the illusion effect in the crossmodal condition increase as the magnitude of the illusion effect in the absence of the sound gets larger. Moreover, despite the high variability between subjects in the illusory magnitude in both sound conditions, it is worth nothing the stronger magnitude of the illusion in the crossmodal condition compared to that in the unimodal condition (i.e., average of 44.9 %, the value of the regression line intercept, see Figure 5, p. 74).

Finally a marginally significant linear relation across participants' (R = .37, t = 1.96, p = .07) of the magnitude effect of the illusion in the two different response conditions (i.e., perceptual and visuo-motor) has been found (see Figure 5, p. 74). This result indicates that the magnitude of the illusory effect in the visuo-motor task increased as the magnitude of the illusory effect increased in the perceptual task. Thus, it seems possible to predict the amount of visuo-motor illusory effect from the magnitude of the perceptual illusion (and vice versa) in support to a close relation between perception and action systems.

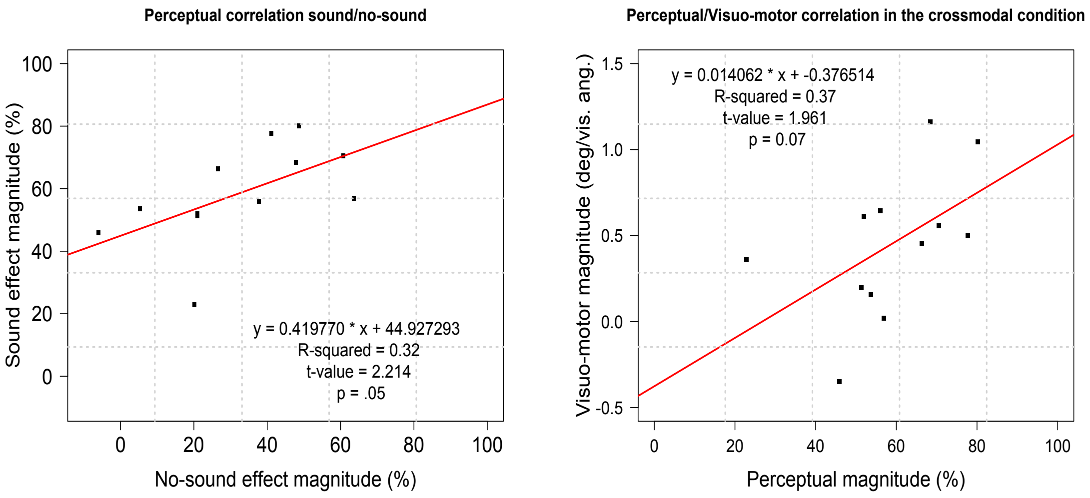

Figure 5 Scatter plot of correlations for the 12 participants. On the left, the correlation between the SIVM magnitude in the two sound conditions (sound/no-sound) for the perceptive tasks. On the right, the correlation between the SIVM magnitude in the perceptive and visuo-motor tasks in presence of the sound.

Discussion

The aim of this study was to verify whether participants' oculomotor behavior might be biased by the SIVM illusion. Starting from Hidaka and colleagues' (2009) experimental paradigm, the purpose was to investigate whether the perception of an illusory visual left/right movement of a static bar induced by an alternating left/right sound affects eye movements landing position when requested to gaze towards the position of the bar. The present study provide evidences of a positive correlation between perceptual and action systems in the SIVM illusion. The marginally significant correlation of the illusion in the two different response conditions (i.e., perceptual and visuo-motor) in presence of the sound shows the predictability of the amount of visuo-motor illusory effect from the magnitude of the perceptual illusion and vice versa. Due to the high variability of the illusory effect between participants it is likely this effect could reach significance by collecting more data.

In the perceptual task, despite a significant modulation of participants' response also in absence of the sound, the magnitude of the illusion was significantly stronger in presence of the sound (i.e., average of 44.9 % stronger) supporting Hidaka and colleagues' (2009) results. The analysis of the expected landing position of eye movements shows that the visuo-motor system is significantly affected by the illusion, in the same direction as in the perceptual response task. Indeed, only a main effect of expected landing position emerged but neither effect of sound condition nor interaction has been found, indicating that the motor response is affected either with or without the sound. This could be due by a number of factors but a possible explanation could rely on the nature of the adopted measure. Eye movement amplitude is a continuous variable, moreover participants were asked to perform a saccade towards an offset target (the bar), thus increasing the variability in saccadic landing positions Furthermore, a between participants positive significant correlation in the perceptual task in respect to the sound conditions has been found. That is, the magnitude of the illusion effect in the crossmodal condition increase as the magnitude of the illusion effect in the absence of the sound gets larger providing evidences of the subjectivity effects of the illusion.

One explanation for the modulation of participants' response in absence of the sound in both the response modalities may depend on the effect of the physical inducer (see Apparatus and stimuli section for more details, p. 67) used to match the direction of the visual illusion with the direction of the apparent motion of the sound. Indeed, without the inducer, the SIVM did not always match the acoustic apparent motion direction and it could not have been possible to have an accurate measure to the direction of the illusion. However, it is important to note that the illusory effect driven by the mere presence of the physical inducer in the absence of the sound is sufficient to influence either motor behaviour or perceptual judgments in a consistent way.

The distinction between perception and action seem to be supported by Milner and Goodale (1995) conclusions. The authors suggest that the ventral pathway provides the visual information required for the object recognition (i.e., vision for perception) while, the dorsal pathway processes the visual information necessary for the control of movements (i.e., vision for action) highlighting the dissociation between the two systems. However the two visual systems hypothesis has been challenged by, the view that the ventro-dorsal pathway has both motor and perceptive functions has been further confirmed by the discovery of mirror neurons (i.e., sensorimotor neurons) in both, the ventral premotor cortex and in the rostral part of the inferior parietal lobe of monkeys (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fogassi, Fadiga & Rizzolatti, 2002). Mirror neurons seem to code specific motor acts and not single movements. Moreover, they fire both when the monkey performs a given action and when it sees someone performing a similar action. The action and perception properties of mirror neurons suggest that their characterizing mechanism process the visual content of an action and its corresponding motor execution, playing a fundamental role in understanding the meaning of actions performed by others (Rizzolatti, Fogassi, & Gallese, 2001).

In the present study it is not possible to make any claim about the neurophysiological substrate of the tasks involved however, results seem to support the correlation between perception and action systems providing psychophysical evidence of such relation.

Evidences of the effects of the reliability of each sensory modality in multisensory interaction have been provided in both neurophysiological (e.g., Stein & Meredith, 1993), behavioural (e.g., Alais & Burr, 2004; Ernst & Banks, 2002) and neuroimaging studies (Noesselt et al., 2010 for a review). Thus, the alternation of sound location can induce an illusory visual motion where the visual sensory signal is less reliable (i.e., at large retinal eccentricities at which visual acuity degrades) highlighting an inverse ventriloquist effect. Results of the present study confirmed the previously found perceptive illusory effect of the SIVM and provide evidence of a correlation with motor response highlighting the involvement of multisensory integration in such visual illusion and its relation with stimulus reliability in both the cognitive domains.

Multisensory interaction and low vision

Background introduction

Visual deficits may be caused by different pathologies or traumatic events that can endanger structures at different levels of visual information processing. They are firstly classified in relation to the level in which the visual pathway is damaged. When the peripheral visual structures are highly compromised (i.e., structures of the eye, the optic nerve or the optic chiasm) the resulted visual deficit is classified as a peripheral visual deficit (i.e., low vision). In all the other cases, when the damage is related to the post-chiasmatic visual pathway, the resulted visual deficit is classified as a central visual deficit as the deficit is related to cortical or subcortical structures (e.g., hemianopia, neglect).

Low vision

Low vision is a condition of permanent reduction of visual field and visual acuity, not correctable by glasses, due to several eye diseases, varying in severity and nature (World Health Organization, 2009). Low vision is very heterogenic between patients because for the many different pathologies' aetiology with the associated visual field type and locus of the deficit. The possible visual deficit may cover the entire visual field with some spared parts or it may also be focalized mostly in the central or in the peripheral visual field in relation to the pathology. Moreover the visual acuity in the spared/damaged visual field areas may also be characterized by very defined boundaries so that very close areas can show very different levels of visual acuity.

Before to a brief description of the most common low vision pathologies, it is helpful to summarily understand how the visual information are converted in electric impulses that can be interpreted by subcortical and cortical structures.

Eye Anatomy

The eye is a spherical structure about an inch in diameter (see Figure 67, p. 80). It has a clear protuberance on the front side, which is the cornea. Beyond the cornea there are other three tissue layers respectively the sclera, a protective layer that connects to the transparent cornea at the front of the eye, the choroid, a vascular layer that is continuous with the ciliary body and the iris on the front side of the eye, and finally, the retina, a light-sensitive tissue constituted by different types of photoreceptors and that cover all the back wall of the inner eye. The eye is a very complex structure but it may be thought to work as a camera. An external object reflects light rays that enter the eyes through the cornea a transparent external membrane that covers the eye. From the cornea the light has to pass through a round hole, the pupil. The size of the pupil can vary by means of the dimension of the iris (the coloured portion of the eye that surrounds the pupil) moved by the ciliary muscles. In this way, the amount of light that pass through the pupil is maintained constant to avoid damaging photoreceptors on the retina. The next step of the light is to pass through the lens that bends the light rays by focusing them on the retina. The retina contains millions of light-sensing nerve cells called rods and cones. Cones are concentrated in the centre of the retina, in an area called the macula, a part of the fovea that provides the highest resolution. Indeed, in bright light conditions, cones provide clear, sharp central vision and detect colours and fine details useful for instance for reading. Rods are located outside the macula and extend towards the outer edge of the retina where are maximally concentrated and allow the eyes to detect motion and to see in dim light and at night. These cells in the retina convert the light into electrical impulses conveyed through the optic nerve that sends these impulses to the primary visual area in the brain where the image starts its path that will end with the awareness of what we are seeing.

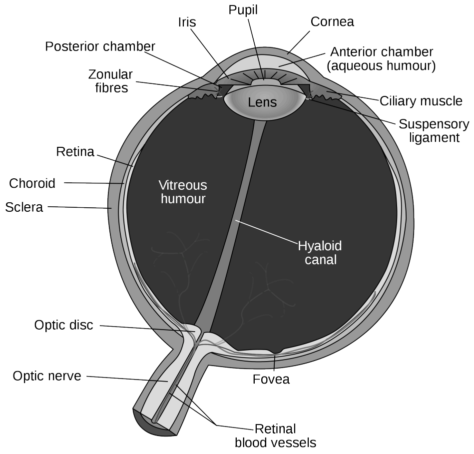

Figure 6 Schematic diagram of the vertebrate eye. Image: Courtesy Wikipedia

Cataract

The most common cause of vision impairment and blindness worldwide is cataract. Since as most of visual diseases, cataract is usually associated with ageing, the magnitude of the problem in western countries seems to be destined to grow in the next years, due to the progressive increase of the percentage of elderly people (World Health Organization, 2009). Cataract is caused by a progressive lens opacisation that modifies the light flow on the retina (see Figure 7, p. 85 for an example). In addition to ageing, different reasons may lead to the loss of the lens transparency as for instance pathologies like the diabetes, prolonged use of drugs or medicines and family history.

Cataract is relatively easily to tract in developed countries by means of surgery and the substitution of the lens.

Macular Degeneration

Macular degeneration (MD) is a visual pathology that progressively destroys the macula, the central portion of the retina, impairing central vision. It rarely causes blindness because only the centre of vision is affected. Macular pathologies impair the ability to see straight ahead clearly and very often make it difficult to read, drive, or perform other daily activities that require fine central vision (see Figure 7, p. 85 for an example). Macular degeneration might be caused by different factors like, for example, the family history (i.e., genes), high blood pressure, high fat intake, smoking, or just ageing. There are different forms of macular degeneration depending of the aetiology.

The most common form of inherited juvenile macular degeneration is the Stargardt's disease (e.g., Weleber, 1994). It is usually diagnosed under the age of twenty. It causes a progressive loss of central vision and, in the early stages patients may experience difficulty with reading and seeing in dim lighting. Other common symptoms include blurriness and distortion. Patients with early Stargardt's disease appear to have simple macular degeneration. When the disease progresses, lipid rich deposits accumulate in the retinal pigment epithelium, a tissue that lies between the retina and the choroid. In advanced Stargardt's disease, the buildup of lipid deposit causes atrophy of the macula because of the lack of photoreceptor cell nutrition.

Differently from Stargardt’s disease, age-related macular degeneration is a visual pathology related to ageing (AMD; e.g., De Jong, 2006; Varma, Fraser-Bell, Tan, Klein, & Azen, 2004). Nowadays, with the increase of life expectancy in developing countries, AMD is getting more common. Usually, it affects people after the age of 60 with a rate of about 10% of elderly between age 65 and 75, increasing to 30% after age 75.

There are two types of AMD: the dry type, and the more frequent wet type. In both dry and wet AMD, the person may notice a blind spot. In the wet AMD, there is a rapid loss of central vision while in the dry form this effect occurs slowly. The wet form is caused by the fast proliferation of new blood vessels leak fluid under the macula that is raised from its normal position thus distorting vision. An early symptom of wet AMD is that straight lines appear undulate.

Despite the aforementioned pathologies are the most common other causes seems to be related to life style and type of nutrition. Smoking and obesity raise the likely of macular degeneration to occur. Furthermore, race and gender factors highlight differences of macular degeneration incidence. Caucasians are much more likely to lose vision from macular degeneration than African Americans are (Varma et al. 2004), while women tend to be mostly affected than men.

Glaucoma

Glaucoma is a common eye disease that can endanger the optic nerve and other parts of the eye, causing the loss of vision or even blindness. The common cause of glaucoma is the rise of the fluid pressure inside the eyes because of slowed fluid drainage from the eye (e.g., Salmon, 1999). It usually affects people after age 40 and it is the second cause of blindness after the cataract. Other causes of glaucoma may be ageing, family history (i.e., genes) and other form of optical affection that can involve the optic nerve. In all cases, the progress of the optic nerve damage is related with the progressively reduction of the visual field starting from the periphery (see Figure 7, p. 85 for an example). There are mainly three types of glaucoma, all linked to the draining defect of the aqueous humor flow. The most common is the open-angle glaucoma, in which the eye pressure rises because the aqueous humor cannot freely drain through the Schlemm's channel. In the acute angle-closure glaucoma, instead, it is the iris that reduces the aqueous humor flow thus raising the internal eye pressure. It is a less common form of glaucoma but it can rapidly impair vision. The last type of glaucoma is the congenital form in which there is a congenital defects in the aqueous humor flow.

The ocular hypertension is an increase in the pressure in the eyes that is above the range considered normal with no detectable changes in vision or damage to the structure of the eyes (e.g., Gordon et al., 2002). The term is used to distinguish people with elevated pressure from those with glaucoma, for whom ocular hypertension is the cause of their pathology. Not all people with ocular hypertension will develop glaucoma because it can be controlled, however, there is an increased risk of glaucoma to occur.

Retinits Pigmentosa

Retinitis pigmentosa is a inherited visual pathology in which abnormalities of the photoreceptors (i.e., mainly the rods but also cones) in the retina lead to progressive visual loss (e.g., Bird, 1995). It may manifest with ageing but also in young people with a degenerative cycle that can endure also years. People affected by retinitis pigmentosa first experience defective dark adaptation (i.e., night blindness), then a progressive reduction of the visual field (i.e., tunnel vision; see Figure 7, p. 85 for an example) and eventually loss of central vision. Retinitis pigmentosa may be inherited in a number of different ways, more than 30 different genes are known to cause it.

Diabetic Retinopathy

Diabetic retinopathy is a common complication of diabetes affecting the blood vessels in the retina (e.g., Frank, 1995). It begins without any noticeable change in vision, but often there are extensive changes in the retina visible to an ophthalmologist. There are mainly two classes of diabetic retinopathy, non-proliferative and proliferative. Non-proliferative retinopathy is the earlier stage in which there may be hemorrhages (i.e., bleeding retinopathy) in the retina with leakage of blood causing deposits of protein (i.e., exudative retinopathy) in the retina. As a consequence in both cases, the retina does not receive enough nutrition. The second stage is the gravest. In proliferative retinopathy new abnormal vessels develop in the retina and grow towards the centre of the eye (i.e., towards the macula; see Figure 7, p. 85 for an example) producing very often fibre tissue that may contract itself and let large scars on the retina. The contraction of this fibre tissue may even cause the underlying retina to detach (i.e., retinal detachment).

Figure 7 Examples of different low visual impairments (column to the right) in respect to the normal vision (column to the left). On the right column starting from the top, myopia, cataract, age related macular degeneration, diabetic retinopathy, glaucoma and retinitis pigmentosa. Images: Courtesy NIH National Eye Institute.

Several studies have investigated unisensory mechanisms of visual perception in low vision individuals. However, previous research has focused on the visual modality and, more precisely, on reading performances (Cheong, Legge, Lawrence, Cheung, & Ruff, 2008), fixation stability (Falkenberg, Rubin, & Bex, 2007), colour recognition (Naïli, Despretz, & Boucart, 2006), eye movements (Crossland & Rubin, 2006) and visual search paradigms (Liu, Kuyk, & Fuhr, 2007).

The studies presented here are the first attempt to investigate possible multisensory interactions in low vision disease.

Introduction to the experiments

The environment we continuously interact with provides a large amount of sensory information, which is processed by our different senses. These inputs are bound together by our brain, and used to construct spatial maps of the external world through the process commonly known as 'multisensory integration' (e.g., Alais et al., 2010). A considerable body of empirical research investigates the three fundamental principles of sensory interaction (e.g., see Calvert et al., 2004, for a review; Stein & Meredith, 1993; Wallace et al., 1992). Spatial and temporal coincidences of sensory modalities that have to be bound together seem to be at the basis of neuronal mechanisms of integration. Integrative effects have been reported in audiovisual (e.g., Frassinetti et al., 2002a, b, 2005; Hairston et al., 2003a, b; Harrington & Peck, 1998; Hughes, Reuter-Lorenz, Nozawa & Fendrich, 1994; Teder-Sälejärvi, Di Russo, McDonald & Hillyard, 2005) and visuotactile (e.g., Forster, Cavina-Pratesi, Aglioti & Berlucchi, 2002) detection tasks, as well as in the case of auditory-somatosensory localisation judgments (e.g., Caclin, Soto-Faraco, Kingstone & Spence, 2002; Occelli, Spence & Zampini, 2011). However, at the behavioural level, these principles seems to be more flexible because of the contribution (i.e., modulation) of high-level polisensory areas involved in the complex cognitive perceptual-decisional network (see Klemen, Chambers, 2011, for a review). Indeed, a more recent electrophysiological study provide evidences that neuronal response enhancement can be observed even when the stimuli are spatially misaligned in their external positions (Wallace & Stein, 2007). Studies from humans indicate that multisensory effects occur over spatial and temporal separations (Spence & Squire, 2003; Wallace et al., 2004) and are not limited to stimulus presentations within the same hemispace. The principle of inverse effectiveness is another principle of multisensory integration. It states that the enhancement in the stimulus detection is proportionally greater when unimodal stimuli are less effective (e.g., Rowland & Stein, 2008; Stein et al., 1996). In a recent fMRI study Noesselt and colleagues (Noesselt et al., 2010) provided evidence of this proportional enhancement effect. A number of audiovisual behavioural studies have reported crossmodal enhancement in relation to stimulus intensity (Andersen & Mamassian, 2008; Frassinetti et al., 2002a, b, 2005; Hairston et al. 2003a, b; Marks et al.,1986). In particular, a series of studies pointed out the facilitatory effect of auditory stimuli in visual detection tasks, performed by either neurologically-intact people with normal vision (with masked subthreshold visual stimuli; Frassinetti et al., 2002a) in conditions of induced myopia (Hairston et al., 2003a) or brain-damaged patients with visual deficit (Frassinetti et al., 2002b, 2005). Frassinetti and colleagues (2005) for instance, showed that in patients affected by hemianopia or neglect, the audiovisual interaction could improve visual perception in the damaged/neglected visual hemifield (i.e., where visual stimuli presented in isolation were less effective) consistently with the principle of inverse effectiveness (e.g., Stein & Meredith, 1993). Moreover, the visual detection enhancement emerged only when auditory and visual stimuli originated from the same spatial position (i.e., crossmodal congruent condition), supporting the spatial rule of multisensory integration (e.g., Stein & Meredith, 1993).

In accordance with the aforementioned evidence, the next three studies reported here aim for the first time to investigate possible multisensory interactions in low vision, by using auditory information to induce a visual detection improvement. The following studies were designed to investigate whether co-occurring sounds, not carrying any information about the presence of the visual target (note that sounds could also be presented alone), could improve visual target detection in low vision individuals (i.e., yes/no task). In the first study, the effect of spatial disparity between synchronous audiovisual crossmodal stimuli was examined. In the second and third study instead, the temporal disparity between spatially coincident audiovisual crossmodal stimuli was manipulated. The hypothesis is that the simultaneous and colocalized presentation of task irrelevant sounds should produce the best benefit for detecting visual stimuli presented in those spatial positions where visual acuity is highly compromised (according to the principle of inverse effectiveness; e.g., Stein & Meredith, 1993). By contrast, participants' performance should not be improved for those spatial positions where visual acuity is not impaired. Moreover, in line with spatial and temporal principles of multisensory integration (Stein & Meredith, 1993; Wallace et al., 1992), as the spatial or temporal disparity between the two modalities increases, the enhancement produced by the disparate sound has to decline.

Audiovisual interaction in low vision: Spatial disparity factors

Thi work was done in collaboration with Valeria Occelli PhD

Introduction Experiment 3

The aim of this first experiment was to verify whether a sound synchronously coupled with a visual stimulus might improve the detection of the visual stimulus in respect to the condition in which the visual stimulus was provided alone. Furthermore, it was aimed to check whether the presumed enhancement might depend to the spatial disparity between the two sensory modalities. Indeed, as stated by the principle of the inverse effectiveness (e.g., Stein & Meredith, 1993), in the conditions in which visual information is sufficient for the task demands (i.e., the visual positions in which the visual deficit is less pronounced) coupling the visual stimulus with a sound should not provide any benefit. Indeed, Frassinetti and colleagues (2005), for instance, showed that in patients affected by hemianopia or neglect, the audiovisual interaction could improve visual perception in the damaged/neglected visual hemifield (i.e., where visual stimuli presented in isolation were less effective) but not in the spared visual hemifield.

To this aim, crossmodal conditions might vary in spatial disparity while a unimodal visual condition was kept as visual performance baseline. Moreover, a unimodal acoustic condition was thought to act as catch trial (i.e., the 'yes' response when actually any visual stimulus was delivered).

Material and methods

Participants

Thirty-one low vision participants (14 female; mean age of 55 years; range from 19 to 82 years; two left handed and two with no hand preference; see Table 2, p. 90 for details) took part in the study. Participants were mostly members of the "Unione dei Ciechi e degli Ipovedenti" (Italian Association for Blindness and Low Vision) of Trento. All the experiments reported in this study were conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki (most recently amended in 2008, Seoul), as well as the ethical guidelines laid down by the University of Trento. All participants gave their informed consent prior to their inclusion in the study and were naïve as to the purpose of the experiment.

Patient	Age	Visus	Visual pathology	Duration
1	30	1/10	Stargardt Disease	20
2	65	1/10	Macular Distrophy	15
3	25	1/20	Acute Maculopathy	20
4	64	1/20	Corneal Opacity (Left eye OFF)	54
5	66	1/20	Maculopaty, Acute Glaucoma (Left eye OFF)	11
6	51	1/10	Glaucoma, Keratoconus	18
7	70	1/10	Acute Degenerative Maculopathy	34
8	66	1/10	Maculopathy	23
9	55	1/10	Macular Distrophy	25
10	56	1/10	Restricted Maculopathy, Optical Neuritis	7
11	52	1/20	Bilateral Congenital Glaucoma	47
12	69	1/20	Retinal Degeneration, Angioid Streaks	16
13	73	1/10	Exudative Maculopathy	5
14	57	1/20	Stargardt Disease, Inverted Retinitis Pigmentosa	37
15	72	1/20	Myopic Choroiditis, Incipient Cataract	22
16	72	1/10	Retinal scar, Strabismus, Cataract	64
17	73	1/10	Albinism, Acute Myopia	5
18	24	1/10	Nistagmus, Degenerative Retinopathy	24
19	72	1/20	Chorioretinopaty, Angioid Streaks	42
20	60	1/10	Maculopathy	33
21	68	1/20	Myophic Choroiditis	24
22	68	1/10	Diabetic Retinopathy	20
23	25	1/20	Optic Chiasm Aneurysm	3
24	59	1/10	Maculopathy	6
25	54	1/20	Corneal Opacity	24
26	39	1/10	Maculopathy	29
27	61	1/10	Myophic Maculopathy	19

Table 2 Participants' age and information about clinical pathology.

Apparatus and stimuli

The setup was a semicircular plastic structure (130 cm length) covering around 112 degree of visual angle and positioned at 70 cm from the participants. Eight LEDs (light emitting diodes) and eight speakers were mounted on this support, resulting in a symmetric arrays of 8 overlying acoustic and visual positions at 8, 24, 40 and 56 degrees of visual angle in either hemifields (see Figure 8, p. 92). All the semicircular structure was covered with an acoustic permeable black curtain so that only the 8 LEDs were visible. A laptop pc (Dell Precision M6300) and a Matlab script (The MathWorks, Inc.) have been used to deliver the stimuli and collect participant's response. A keyboard was positioned in front of the participants to allow them to provide the responses.

Auditory stimuli consisted of the presentation of a 100 ms white noise burst (80 dB as measured from the participants' head position). Sounds were played by using the integrated sound card of the laptop computer connected to an external loudspeaker to amplify the signal. The signal was then switched between 8 relays (N4100F-2) by using the digital outputs of an Interface Board Module (Velleman Extender USB VM140) to activate the desired speaker. Speakers were round-shaped (5 cm diameter of Mylar; Pro Signal ABS-210-RC range 350-20000 Hz, 8 Ω, 1 W RMS Power). A between-trials balanced random amplitude modulation of the generated signal was introduced (values of 85, 90, 95 or 100% of the whole signal amplitude) to compensate for the minor speakers' difference in propagating the acoustic stimuli. Visual stimuli consisted of the presentation of a 100 ms (i.e., same duration as the auditory stimuli) green visual targets (LED, Avago Technologies model HM65-Y30DD). The luminance of each LED associated to each speaker has been calibrated to 80 cd/m² set in a dark environment. LEDs were oval-shaped with a diameter of 5 mm (0.4 degrees of visual angle) and a viewing angle of 100° (i.e. the angle from which the 80 cd/m² luminance was maintained constant), so that the visual stimulus has a constant luminance despite its position on the setup (i.e. minor difference in the direction towards the participant's position). The onset of visual and auditory stimuli was synchronized by using a digital oscilloscope (Agilent Technologies MSO 6054A).

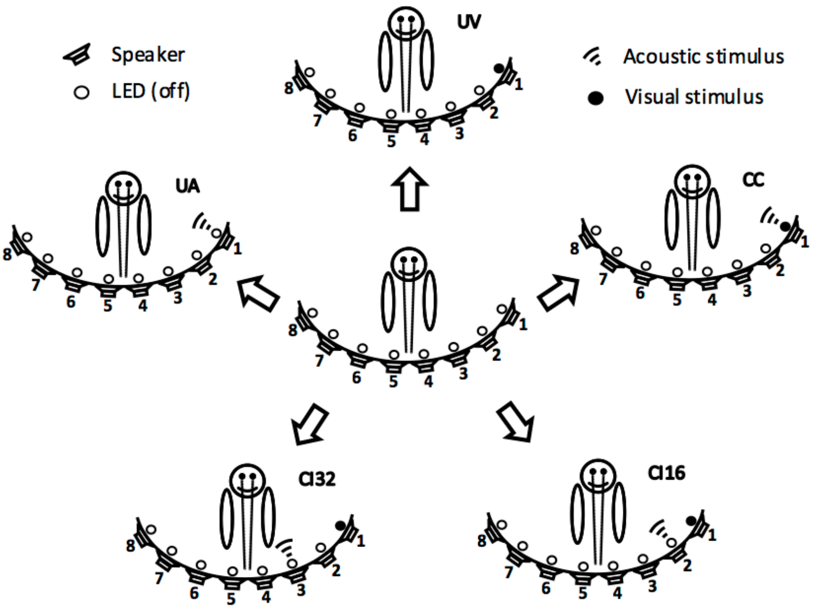

Figure 8 Schematic representations of the experimental apparatus. The viewing distance is 70 cm. Each speaker is placed symmetrically at an eccentricity of 8, 24, 40 and 56 degrees from the centre, in either hemifield. In the central panel, the initial trial is represented with LEDs and loudspeakers turned off. In the surrounding panels, the five conditions are represented in a clockwise orientation starting from the top: unimodal visual (UV), crossmodal congruent (CC), crossmodal incongruent with 16 degrees of audiovisual disparity (CI16), crossmodal incongruent with 32 degrees of audiovisual disparity (CI32), and unimodal acoustic (UA; i.e., catch trial).

Procedure and experimental design

Participants sat at 70 cm in front of the semicircular structure in a dimming room (average luminance 40 cd/m²). The chosen ratio between LED (80 cd/m²) and room luminance calibration was thought to reduce any possible flash light reflection. Participants were asked to keep their head and their eyes as steady as possible by looking straight ahead to the central position of the apparatus. The experimenter sat in front of the participant (behind the apparatus) to check whether head and eyes were always in the requested (constant) position. Each trial started automatically after the participant response. It is worth noting that crossmodal trials were always synchronous thus reducing the possibility that the sound acted as a cue for any possible eye movement that could affect visual detection performance.

Participants were requested to detect the presence - not the spatial position - of the visual stimuli and ignore the sound. The participants were not informed about the number and spatial locations of the speakers. For each trial, participants were presented with five different conditions: Visual stimulus alone (i.e., unimodal visual condition, UV), acoustic stimulus alone (i.e., unimodal acoustic condition, UA or catch trials), and simultaneous presentation of a visual and auditory stimulus (i.e., crossmodal condition). Crossmodal condition were three, namely; trials in which the visual and the auditory stimuli were presented from the same spatial position (i.e., crossmodal congruent condition, CC) and trials in which the two different modalities were presented from different spatial positions (i.e., crossmodal incongruent condition, CI) either at 16 (i.e., CI16) or 32 (i.e., CI32) degrees of visual angle.

The participants were instructed to use both the hands to press two buttons on a keyboard placed in front of them (‘F’ key for ‘Yes, I saw the light’ and ‘J’ key for ‘No, I did not see the light’) using, respectively, the left and the right index finger. Response mapping was not counterbalanced between participants. The whole experimental section was divided in 8 blocks. Each block consisted of 72 trials (i.e., 8 UV trials, one for each visual position; 24 UA trials/catch trials, 3 for each of the 8 acoustic positions; and 8 CC trials, visual and acoustic stimuli presented synchronously from each of the 8 positions). Finally, there were 32 CI trials (4 for each visual position) in which acoustic stimuli were presented to the left or to the right of the visual stimulus at either 16 or 32 degrees of disparity. The proportion of conditions in the experimental design has been chosen for different reasons: to respect the same design used by Frassinetti and colleagues (2005) and to shortening as much as possible the duration of the experimental section because of the very high patients' mean age. Each participant was presented with 576 trials in total.

To verify that the procedure was clear, before the actual experimental section, the participants were asked to undergo a brief test with the same experimental conditions (i.e., 10 to 20 trials randomly chosen from the sequence of a block). The whole experimental section lasted around 90 minutes with a rest between each of the 8 blocks.

Data analysis

Participants' performance was analysed by computing the proportion of 'yes' responses. In relation to the 8 spatial positions, 50 values have been obtained namely: 8 values for the UA condition (catch trial), 8 values for the UV condition, 8 values for the CC condition, and 26 different values for each CI condition corresponding to left and right performance at 16° or 32° of disparity. Indeed, CI values were 26 instead of 32 because of the 'lack' of some positions. For example, for the visual spatial position at -56° on the left visual hemifield, there were no sounds on the left neither at 16° (i.e., at -72°) nor at 32° (i.e., at -88°). In this case, conditions on the right at 16° (i.e., at -40°) and 32° (i.e., at -24°) were presented two times. Next, CI positions were clustered and averaged according to the side (i.e., right or left) and the distance of the sound from the reference visual position (i.e., 16° and 32°), giving rise to four CI values for each participant (i.e., CI16L, CI32L, CI16R and CI32R). Finally, proportion values for UA, UV, CC, CI16 and CI32 (the last two were calculated by averaging the values for left and right sides) related to each of the eight spatial positions have been ordered starting from the lowest to the highest as measured by the performance in the UV condition (i.e., ascending order).

After ordering the data in such described way, the first position (i.e., the one in which the proportion of correct responses in the UV condition was the lower, that is, the most impaired visual position) could have been either in the periphery or in the center, depending on the participant’s visual deficit (see Table 3, p. 96 for details). Four participants have been discarded from the subsequent analysis because they had a very high UV performance in the first spatial position (i.e., in the most impaired spatial position the UV performance was above 95% of correct responses).

	Unimodal visual performance p(y) and spatial position							
Patient	**Most Impaired positions (MI)**				**Less Impaired positions (LI)**			
1	.01 (5)	.01 (6)	.13 (7)	.38 (4)	.88 (1)	.99 (2)	.99 (3)	.99 (8)
2	.38 (5)	.63 (4)	.88 (8)	.99 (1)	.99 (2)	.99 (3)	.99 (6)	.99 (7)
3	.01 (3)	.13 (4)	.13 (5)	.25 (6)	.38 (8)	.63 (1)	.75 (2)	.99 (7)
4	.01 (1)	.25 (2)	.75 (8)	.88 (5)	.99 (3)	.99 (4)	.99 (6)	.99 (7)
5	.50 (3)	.63 (4)	.88 (1)	.88 (5)	.88 (8)	.99 (2)	.99 (6)	.99 (7)
6	.01 (1)	.01 (2)	.13 (3)	.99 (4)	.99 (5)	.99 (6)	.99 (7)	.99 (8)
7	.75 (5)	.88 (2)	.88 (3)	.99 (1)	.99 (4)	.99 (6)	.99 (7)	.99 (8)
8	.75 (5)	.99 (1)	.99 (2)	.99 (3)	.99 (4)	.99 (6)	.99 (7)	.99 (8)
9	.25 (8)	.88 (7)	.99 (1)	.99 (2)	.99 (3)	.99 (4)	.99 (5)	.99 (6)
10	.88 (1)	.88 (3)	.88 (6)	.88 (7)	.88 (8)	.99 (2)	.99 (4)	.99 (5)
11	.25 (1)	.63 (6)	.63 (7)	.75 (8)	.99 (2)	.99 (3)	.99 (4)	.99 (5)
12	.01 (5)	.13 (3)	.13 (4)	.13 (6)	.25 (7)	.88 (1)	.88 (2)	.99 (8)
13	.88 (3)	.88 (5)	.99 (1)	.99 (2)	.99 (4)	.99 (6)	.99 (7)	.99 (8)
14	.63 (4)	.63 (8)	.75 (2)	.88 (1)	.88 (5)	.99 (3)	.99 (6)	.99 (7)
15	.01 (5)	.13 (3)	.25 (4)	.38 (6)	.88 (1)	.88 (2)	.99 (7)	.99 (8)
16	.38 (1)	.38 (5)	.63 (3)	.63 (4)	.63 (8)	.75 (6)	.75 (7)	.88 (2)
17	.88 (2)	.88 (5)	.99 (1)	.99 (3)	.99 (4)	.99 (6)	.99 (7)	.99 (8)
18	.50 (4)	.99 (1)	.99 (2)	.99 (3)	.99 (5)	.99 (6)	.99 (7)	.99 (8)
19	.25 (4)	.25 (5)	.25 (6)	.50 (3)	.88 (7)	.99 (1)	.99 (2)	.99 (8)
20	.13 (4)	.99 (1)	.99 (2)	.99 (3)	.99 (5)	.99 (6)	.99 (7)	.99 (8)
21	.01 (1)	.01 (5)	.13 (4)	.13 (6)	.88 (2)	.88 (7)	.99 (3)	.99 (8)
22	.25 (1)	.25 (6)	.25 (8)	.38 (7)	.50 (5)	.63 (2)	.75 (3)	.75 (4)
23	.01 (6)	.01 (7)	.01 (8)	.50 (5)	.75 (1)	.75 (3)	.99 (2)	.99 (4)
24	.50 (5)	.75 (1)	.75 (3)	.88 (6)	.99 (2)	.99 (4)	.99 (7)	.99 (8)
25	.01 (7)	.01 (8)	.13 (6)	.50 (4)	.50 (5)	.75 (2)	.88 (3)	.99 (1)
26	.01 (5)	.25 (4)	.38 (6)	.99 (1)	.99 (2)	.99 (3)	.99 (7)	.99 (8)
27	.13 (1)	.25 (5)	.88 (6)	.88 (8)	.99 (2)	.99 (3)	.99 (4)	.99 (7)

Table 3 For each participant, the spatial positions have been ordered as a function of the performance in the Unimodal Visual (UV) condition. 'MI' indicates the cluster of the most impaired four positions; 'LI' the cluster of the least impaired four positions.

Results

A first analysis was conducted to assess whether there was a difference between side of the sound (left or right) and disparity (16 or 32 degrees) with respect to the visual stimulus position in the CI condition. Greenhouse-Geisser correction was applied to the within participants analysis of variance (ANOVA) with factors disparity and side of the sound. The results showed a significant difference between spatial disparity 16°

(proportion of 'yes' responses, p(y) = .75) and 32° (p(y) = .71), F (1, 26) = 28.2, p < .001, indicating that the integration enhancement effect decreases as a function of the increasing spatial disparity between the sound and the visual stimuli (see Figure 9, p. 97). This result is in line with the spatial rule of multisensory integration (e.g., Stein & Meredith, 1993). Neither the effect of side of the sound (left side: p (y) = .74 vs. right side: p (y) = .72), F (1, 26) = 1.12, p = .30, nor the interaction between side of the sound and spatial disparity, F (1, 26) = .09, p = .76, were significant.

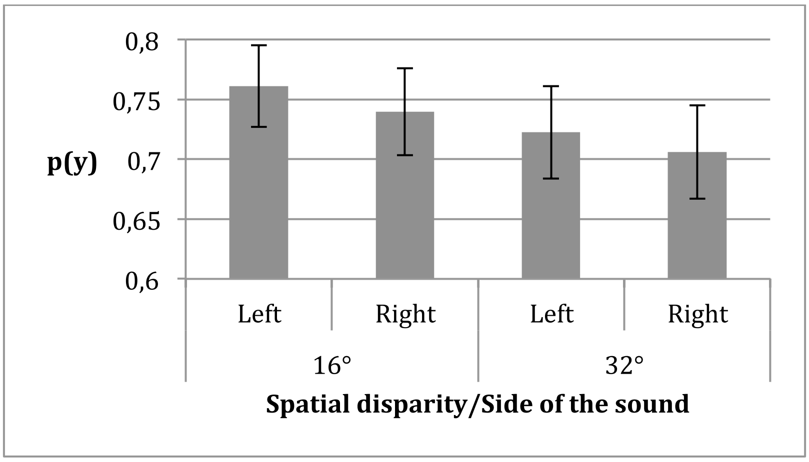

Figure 9 Mean proportions of 'yes' responses, for each sound-LED spatial disparity (i.e., 16 and 32 degrees) and side of presentation (i.e., sounds presented to the left or to the right of the LED). Performance for the two sound disparities (averaged for left and right side) differs significantly (p < .001).

Given that there was no difference between incongruent conditions as a function of the side of the sounds (i.e., on the left or on the right of the visual stimulus), mean proportion of correct responses were collapsed for the left and right incongruent positions (i.e. for each spatial position, CI16L with CI16R and CI32L with CI32R were averaged) to obtain only two different CI values for each participant in relation to the spatial disparity (i.e., CI16 and CI32). Then, proportion of correct responses corresponding to the four most impaired positions (i.e., MI positions, namely the first four ordered positions) and the proportion of correct responses for the least-impaired/spared four positions (i.e., LI positions, namely the ordered positions from the fifth to the eighth) were clustered for each participant and condition (see Table 3, p. 96 for an example for the UV performance). A within participants ANOVA with the factors cluster of positions (i.e., MI vs. LI) and condition (i.e., UV, CC, CI16 and CI32) revealed a significant effect of condition, F (3, 78) = 4.5, p = .006, a significant effect of cluster of positions F (1, 26) = 48.6, p < .001, and a significant interaction between condition and cluster of positions, F (3, 78) = 3.2, p = .03. As expected, results show a difference between the stimulation conditions, while the emerged interaction show that this differences may be present only in one of the two clusters of positions (i.e., MI vs. LI, see Figure 10, p. 98).

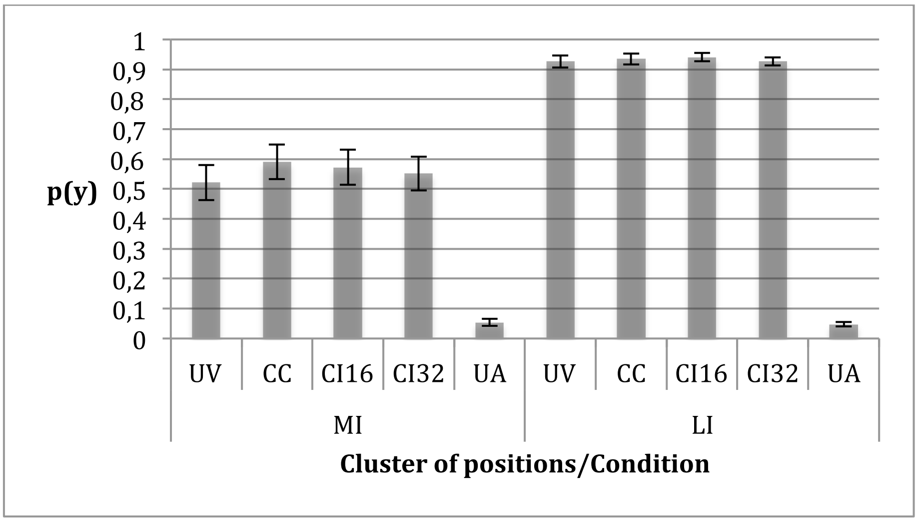

Figure 10 Mean proportions of 'yes' responses, reported for each cluster of positions (i.e., most impaired four positions, MI; less impaired four positions, LI), and each experimental condition (i.e. unimodal visual, UV; crossmodal congruent, CC; crossmodal incongruent at 16 and 32 degrees of disparity, respectively CI16 and CI32; unimodal acoustic, UA). Performance comparisons UV vs. CC and CC vs. CI32 differ significantly (p < .05). Difference between UV and CI16 is marginally significant (p = .06). Proportions of UA (catch trials) in the two clusters of positions (i.e., MI vs. LI) did not differ significantly.

The post-hoc comparisons pointed out a significant difference in the MI cluster of positions between UV and CC (p(y) = .52 vs. p(y) = .59, p = .03) indicating that the CC condition provided a significant performance improvement with respect the UV condition. As expected, a sound in the same spatial position of the visual stimulus provides a significant improvement in the visual detection task, but, this improvement is no longer present when the sound is at 32° from the visual stimulus (UV, p(y) = .52 vs. CI32, p(y) = .55, p = .46). Surprisingly, a sound at 16° from the visual stimulus also provided a marginally significant enhancement in the visual detection task performance as compared to the unimodal visual condition (UV, p(y) = .52 vs. CI16, p(y) = .57, p = .06). Furthermore, the performance was significantly better in the congruent position than in the one with a audiovisual disparity of 32° (CC, p(y) = .59 vs. CI32, p(y) = .55, p = .015). Finally, performance at 16° was not different from the one observed in the congruent position (CC, p(y) = .59 vs. CI16, p(y) = .57, p = .32), indicating that for low vision patients there could be visual acoustic integration also with such relatively wide disparity. No difference between conditions has been found in the LI cluster of positions (for all comparisons p = 1; UV, p(y) = .93, CC, p(y) = .93, CI16, p(y) = .94 and CI32, p(y) = .93). Taken together, these results support the principle of inverse effectiveness, given that the multisensory enhancement has been found in the MI cluster of positions (i.e., where visual stimuli are less reliable) and not for the LI cluster of positions (i.e., where stimuli were highly effective; see, e.g., Frassinetti,et al., 2002a, b, 2005; Hairston, et al., 2003a; Noesselt et al., 2010, for similar results).

Finally, responses to catch trials were analysed to examine whether participants were less able to ignore the sounds presented alone in the MI than in LI cluster of positions. A pairwise comparison t-test did not reveal any significant difference (MI, p(y) = .053 and LI, p(y) = .047, t(26) = .86, p = .39).

Discussion

Results of the present study show that a synchronous sound presented from the same spatial position significantly enhances the performance of low vision individuals in a yes/no visual detection task as compared to the condition where the visual stimulus was presented in isolation. Moreover, a significant acoustic crossmodal effect is observed for the most impaired visual positions (i.e., MI cluster), but not for those in which the visual sensory signal is still reliable (i.e., LI cluster). That is, a significant enhancement is observed in the spatial positions in which the unimodal visual performance is mostly deteriorated. This result is in line with previous behavioural studies highlighting the role of the visual stimulus reliability in visual acoustic crossmodal tasks (Frassinetti, et al., 2002a, b, 2005; Hairston, et al., 2003a; see Noesselt et al., 2010, for both behavioural and neuroimaging evidence).

The visual detection performance in the crossmodal congruent condition and at 16 degrees of disparity does not differ significantly, while when the sound is presented at 32 degrees of disparity performance does not significantly differ from that in the unimodal visual condition. Therefore, the absence of any difference between unimodal visual condition and audiovisual pairs separated by 32 degrees of disparity shows that the enhancement effect cannot be attributed to a general unspecific alerting effect induced by the mere presence of auditory stimuli on bimodal trials (e.g., Posner, 1978, 1980).

To our surprise, the results showed a marginally significant performance enhancement as compared to the unimodal visual condition also in case of a sound disparity of 16 degrees, regardless of whether the visual stimulus was central or peripheral. This pattern is, however, partially consistent with the performance observed by Frassinetti and colleagues in neglect patients without hemianopia (Frassinetti, et al., 2002b, 2005), who showed an enhancement of visual detection when the sound was at 16 degrees in the peripheral visual field. The authors argued that this effect could be due to the presence of an attentional deficit that may enlarge the size of the area where the crossmodal integration occurs. However, Frassinetti and colleagues (2002a) found an analogous effect also in neurologically-intact people with normal vision, in which attentional deficit can be likely excluded. These authors have explained this spatial disparate enhancement effect by referring to electrophysiological evidence, showing that auditory receptive fields in multimodal neurons are larger than visual receptive fields (Middlebrooks & Knudsen, 1984).

A second possible explanation for the enhancement effect of the sound at 16 degrees may refer to the inverse ventriloquism effect, in which the sound leads the fusion process by providing a most reliable spatial cue. For instance, Alais and Burr (2004) asked participants to localize the spatial position of a visual stimulus or a sound (i.e., left/right judgment) presented either in unimodal or crossmodal fashion. The authors found that when the visual stimulus is well discriminable, an acoustic stimulus has no influence on the performance. By contrast, the more the visual stimulus is blurred, the more participants' judgment is biased towards the source position of the sound. This interesting result suggests that the spatial ventriloquism effect seems to be determined by the contribution of both modalities, in a simple model of optimal combination, in which the most reliable sensory signal will provide the highest contribution (e.g., Ernst & Banks, 2002). Thus, the marginally significant effect of the sound found also at 16 degrees in the present study may be the result of a lack in the reliability of the visual information that constitute, together with the sound, an unique crossmodal event (e.g., Hairston et al. 2003b). As a consequence, multisensory enhancement may be observed not only when the low reliable visual stimulus and the sound's spatial position coincide, but also when the disparity between the two stimuli is around 16 degrees. Wallace and colleagues (Wallace et al., 2004) found that the audiovisual interaction effects were modulated by the participant's perception (i.e., judgment) of the unity of the event. In that study, localisation bias (i.e., sound localisation towards the visual stimulus) and reports of perceptual unity occurred even with substantial spatial (i.e., 15 degrees) and temporal (i.e., 800 ms) disparities. Similarly, one could suggest that the multisensory enhancement in the CC and CI16 emerged in the present study because low vision participants perceived acoustic and visual stimuli as being originated from the same event. By contrast, the sound at 32 degrees of disparity does not provide a significant enhancement, supporting the existence of a spatial boundary in which visual acoustic fusion effect breaks off (cf. Wallace et al., 2004). Other studies have nevertheless reported substantial localisation biases with similarly large disparities (e.g., Bermant & Welch, 1976; Bertelson & Radeau, 1981) despite values emerged in this study seem large in comparison with some previous reports of crossmodal interaction (e.g., Lewald & Guski, 2003; Lewald et al., 2001).

Irrespectively of the causal explanation for the effect of the disparate sound on visual detection highlighted in the present study, the enhancement seems to be consistent throughout the visual field. In fact, the MI cluster of positions is constituted by positions either in the peripheral or in the central visual field in relation to the patients' pathology. Thus, it is unlikely that it has been determined just by the effect of those particular positions in the periphery (like found by Frassinetti et al. 2002a in healthy participants), where acoustic RFs have temporal borders which extend to the peripheral space more than the nasal borders (i.e., RFs are elongated towards the periphery; Middlebrooks & Knudsen, 1984; Stein et al., 1989).

In the present study, audiovisual stimuli were always presented simultaneously. However, in line with Wallace and colleagues study (Wallace et al., 2004), the marginal enhancement at 16 degrees of disparity emerged in the present study might be due to the fact that low vision participants perceived acoustic and visual stimuli as being originated from the same event. Furthermore, as in Wallace et al. (2004) it has been demonstrated that sensory integration can take place also between stimuli that are not temporally coincident, but which fall within the 'temporal window' of integration (Meredith et al., 1987; Spence & Squire, 2003). Therefore, in the next experiment the aim is to investigate whether the behavioural enhancement effect found for spatial disparities in the present experiment might result for temporal disparities as well. Furthermore, it would be possible to try to define the temporal window boundaries for the audiovisual enhancement effect found in low vision participants.

Audiovisual interaction in low vision: Temporal disparity factors I

Thi work was done in collaboration with Valeria Occelli PhD

Introduction Experiment 4

Results of Experiment 3 show that a sound synchronously coupled with a visual stimulus provides a significant enhancement in detecting the presence of the visual stimulus when compared with the unisensory visual presentation. As expected, this effect is present only in the visual positions where visual information is compromised. Moreover, surprisingly, beyond the expected enhancement in the crossmodal spatially congruent condition, the enhancement effect is marginally significant also when the sound is disparate of 16 degrees in respect the visual stimulus, and this enhancement does not differ significantly from that in the spatially congruent condition. While this result does not match with some previous studies (e.g., Frassinetti, et al., 2002b, 2005), it is in line with other crossmodal studies in which spatial or temporal disparities seem to extent the spatial and temporal window of integration between modalities (e.g. Wallace et al., 2004; Meredith et al., 1987; Spence & Squire, 2003 for temporal disparities).

Thus, the aim of the present experiment was to verify the role of temporal factors in audiovisual integration in low vision individuals. More precisely, whether the enhancement effect of a sound provided in the same spatial position of the visual stimulus might depend to the temporal disparity between the two sensory modalities. Furthermore, is possible to hypothesize that, as found in Experiment 3, in the visual positions in which the visual deficit is less pronounced, coupling the visual stimulus with a sound should not provide any benefit.

To this aim, in the present experiment, crossmodal conditions might vary in temporal disparity while a unimodal condition was kept as baseline. As in Experiment 3 design, a unimodal acoustic condition was thought to act as catch trial to check the proportion of false alarms (i.e., the 'yes' response when actually any visual stimulus was delivered).

Material and methods

Participants

Fifteen low vision participants (7 female; mean age of 50 years; range from 20 to 73 years; see Table 4, p. 105 for details) took part in the study. Participants were mostly members of the "Unione dei Ciechi e degli Ipovedenti" (Italian Association for Blindness and Low Vision) of Trento.

Patient	Age	Visus	Visual pathology	Duration
1	73	1/10	Exudative Maculopathy	5
2	60	1/10	Myophic Maculopathy (Right eye OFF)	50
3	74	1/20	Acute Degenerative Maculopathy	10
4	65	1/10	Optical neuritis	17
5	42	1/10	Myophic Maculopathy	25
6	66	1/10	Maculopathy	23
7	65	1/20	Diabetic Maculopathy	13
8	39	1/10	Maculopathy	29
9	25	1/20	Acute Maculopathy	20
10	61	1/10	Myophic Maculopathy	19
11	66	1/20	Maculopaty, Acute Glaucoma (Left eye OFF)	11
12	24	1/10	Nistagmus, Degenerative Retinopathy	24
13	20	1/10	Stargardt Disease	10
14	52	1/20	Bilateral Congenital Glaucoma	47

Table 4 Participants' age and information about clinical pathology.

Apparatus and stimuli

The apparatus and the stimuli were as in the Experiment 3 (see Figure 8, p. 92) with the following exceptions. In this experiment participants provided the responses by pressing a button on a Wii Remote (the primary controller for Nintendo's Wii game console, Nintendo, Inc. Figure 11, p. 108) connected to the laptop by means of the Bluetooth connection. Furthermore, participants' fixation was controlled by means of an eye tracker (EyeLink 1000 Desktop Mount, SR Research).

An accurate calibration of the onset of crossmodal stimuli was verified by using a digital oscilloscope (Agilent Technologies MSO 6054A).

Procedure and experimental design

Participants sat at 70 cm in front of the semicircular structure in a dimly illuminated room (average luminance 40 cd/m²). The chosen ratio between LED (80 cd/m²) and room luminance calibration was thought to reduce any possible flash light reflection. Participants were asked to keep their head and their eyes as steady as possible by looking straight ahead to the central position of the apparatus. The experimenter sat in front of the participant (behind the apparatus) to check that the participant respected all the instructions. Furthermore, the experimenter had to record the fixation position to be maintained by the participant throughout the whole experimental block by mean of an eye tracker. To this aim, when the experimental block started, the experimenter verified the participant's head and gaze directions (i.e., experimental posture). When the participant's posture was correct, the experimenter started a procedure to record participant's gaze position for around 2 seconds. During this time interval, the eye tracker (with a sampling rate of 1000 Hz) recorded around

2000 gaze positions (i.e., X/Y coordinates in a bi-dimensional external space). Then, the mean gaze position X and Y coordinates of the recorded sampled points were used to obtain a single mean fixation point corresponding to the centre of the setup. Finally, a fixation confidence area was built on the obtained mean fixation point (i.e., around 3 degrees of tolerance on the left, right, up and down directions from the mean fixation point). This relatively large fixation confidence area was necessary because most of the participants had visual damage in the central visual field and thus they had to fixate something that they could not see.

During each trial, on the screen of the laptop used to run the experiment, the experimenter was able to see where the participant was fixating within the confidence area and if necessary provided the participants with suggestions in order to adjust their gaze. Once the participants' gaze was within the confidence area for at least 50 ms, the trial started automatically. Because the sound was provided always before the visual stimulus (or synchronously), the onset of the sound provided additional information regarding the spatial position of the subsequent visual stimulus. Thus, in order to avoid a cueing sound effect on gaze direction in crossmodal trials with longer SOAs (i.e. 250 and 400 ms), a further online control procedure of the participant's gaze position has been used. In fact, only when participant was able to maintain the gaze inside the fixation confidence area during the whole SOA duration, the visual stimulus was turned on. Conversely, if the participant gazed towards the sound's cued spatial positions (i.e., the spatial position in which the visual stimulus would be appeared), the visual stimulus was not delivered, and the same trial started again. This further gaze check procedures was used neither for synchronous crossmodal stimulation nor in the 100 ms SOA conditions because the SOA duration was not enough for programming and executing a saccade towards the sound's cued position.

The participants were naïve to the purpose of the experiment and not informed about the number and the spatial locations of the speakers. They were nevertheless told that in some trials visual and acoustic stimuli were not synchronized. For each trial, participants were presented with six different conditions: Visual stimulus alone (i.e., unimodal visual condition, UV), acoustic stimulus alone (i.e., unimodal acoustic condition, UA or catch trials), and the crossmodal conditions. Crossmodal conditions were four, namely; a synchronous condition with SOA equal to 0 ms, and three conditions in which the visual stimulus was delayed by100, 250 and 400 ms (SYNC, AV100, AV250 and AV400 respectively).

The participants were requested to detect the presence - not the spatial position - of the visual stimulus (i.e., yes/no task) and to ignore the sound. By using the preferred hand, the participants had to press two buttons on the Wii Remote by using the index and the thumb fingers to provide the response (respectively the ‘A’ key for ‘Yes, I saw the light’ and ‘B’ key for ‘No, I did not see the light’, see Figure 11, p. 108).

Figure 11 The remote control (Wii Remote, Nintendo, Inc.) used by participants to provide the response. In the figure, the ‘A’ button is on the top of the remote just below the thumb finger, while the ‘B’ button is opposite to the ‘A’ button on the back of the remote. The ‘B’ button was pressed with the index finger.

The whole experimental section was divided in 5 blocks and lasted around 90 minutes with a rest between blocks. Each block consisted of 96 trials (i.e., 48 different conditions repeated twice). Namely, 8 UV trials, one for each visual position; 8 UA trials, catch trials, one for each of the 8 acoustic positions, and 32 CC trials, visual and acoustic stimuli presented ones from each of the 8 positions for the four different SOAs. At the end of the whole experiment, each participant was presented with 480 trials in total.

To verify that the procedure was clear, before the actual experimental section, the participants were asked to undergo a brief test with the same experimental conditions (i.e., 10 to 20 trials randomly chosen from the sequence of a block).

Data analysis

Participants' performance was analysed by computing the proportion of 'yes' responses. As in Experiment 3, for each of the eight spatial positions, 48 values, separately for the six experimental conditions (i.e., UV, SYNC, AV100, AV250, AV400 and UA/catch trials) were obtained. Moreover, proportion values related to each condition and each of the eight spatial positions have been ordered starting from the lowest to the highest as measured by the performance in the UV condition (i.e., ascending order). After ordering the data in such described way, the first position (i.e., the one in which the performance in the UV condition was the lower, that is, the most impaired visual position) could have been either in the periphery or in the centre, depending on the participant's visual deficit (see Table 5, p. 110 for details). Finally, proportion of correct responses corresponding to the four most impaired positions (i.e., MI positions, namely the first four ordered positions) and the proportion of correct responses for the least impaired/spared four positions (i.e., LI positions, namely the ordered positions from the fifth to the eighth) were clustered for each participant and condition (see Table 5, p. 110 for an example for the UV performance).

One participant has been discarded from the subsequent analysis because his UV performance in the first spatial position was very high (i.e., in the most impaired spatial position the UV performance was above 95% of correct responses).

	Unimodal visual (UV) performance p(y) and spatial position							
Patient	**Most Impaired positions (MI)**				**Less Impaired positions (LI)**			
1	.60 (4)	.80 (2)	.80 (7)	.90 (8)	.99 (1)	.99 (3)	.99 (5)	.99 (6)
2	.01 (3)	.10 (4)	.30 (5)	.30 (8)	.50 (6)	.99 (1)	.99 (2)	.99 (7)
3	.20 (1)	.20 (8)	.40 (2)	.40 (6)	.60 (5)	.70 (7)	.80 (3)	.90 (4)
4	.40 (4)	.99 (1)	.99 (2)	.99 (3)	.99 (5)	.99 (6)	.99 (7)	.99 (8)
5	.30 (8)	.70 (1)	.80 (3)	.90 (2)	.90 (4)	.90 (6)	.99 (5)	.99 (7)
6	.10 (5)	.60 (6)	.80 (8)	.90 (2)	.90 (4)	.99 (1)	.99 (3)	.99 (7)
7	.60 (1)	.80 (2)	.80 (3)	.80 (7)	.90 (4)	.90 (5)	.90 (6)	.90 (8)
8	.01 (5)	.01 (6)	.10 (4)	.99 (1)	.99 (2)	.99 (3)	.99 (7)	.99 (8)
9	.01 (3)	.01 (4)	.01 (5)	.01 (8)	.20 (1)	.70 (6)	.90 (2)	.90 (7)
10	.01 (1)	.10 (5)	.90 (2)	.99 (3)	.99 (4)	.99 (6)	.99 (7)	.99 (8)
11	.20 (4)	.40 (5)	.50 (3)	.80 (8)	.99 (1)	.99 (2)	.99 (6)	.99 (7)
12	.60 (4)	.90 (7)	.90 (8)	.99 (1)	.99 (2)	.99 (3)	.99 (5)	.99 (6)
13	.40 (5)	.60 (8)	.70 (4)	.90 (2)	.99 (1)	.99 (3)	.99 (6)	.99 (7)
14	.01 (1)	.10 (6)	.60 (7)	.90 (5)	.90 (8)	.99 (2)	.99 (3)	.99 (4)

Table 5 For each participant, the spatial positions have been ordered as a function of the performance in the Unimodal Visual (UV) condition. 'MI' indicates the cluster of the most impaired four positions; 'LI' the cluster of the less impaired four positions.

Results

A within participants analysis of variance ANOVA with the factors cluster of positions (i.e., MI vs. LI) and condition (i.e., UV, SYNC, AV100, AV250 and AV400) revealed a significant effect of condition, F (4, 52) = 13.3, p < .0001, a significant effect of cluster of positions F (1, 13) = 21.3, p < .001, and a significant interaction between condition and cluster of positions, F (4, 52) = 9.8, p < .0001. As expected, results show a difference between the stimulation conditions, while the emerged interaction show that this differences may be present only in one of the two clusters of positions (i.e., MI vs. LI, see Figure 12, p. 108).

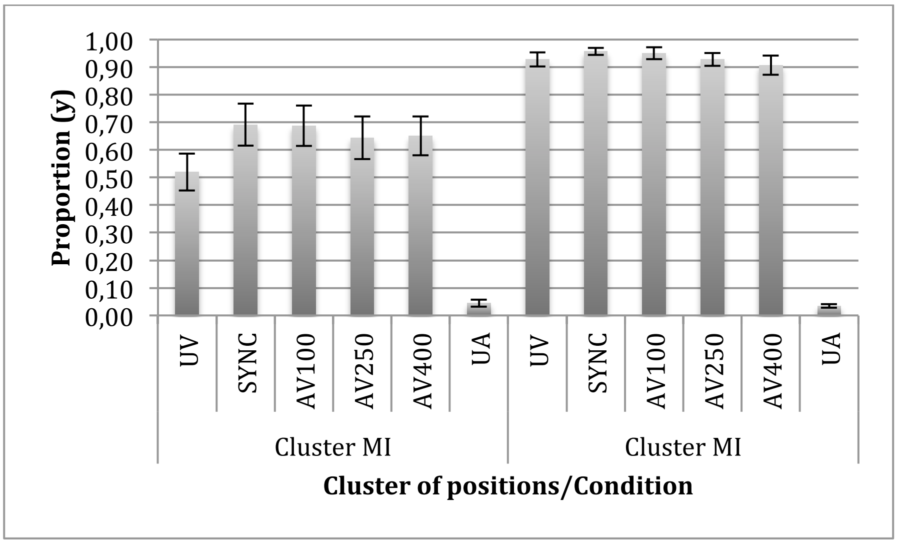

Figure 12 Mean proportions of 'yes' responses, reported for each cluster of positions (i.e., most impaired four positions, MI; less impaired four positions, LI), and each experimental condition (i.e. unimodal visual, UV; synchronous, SYNC; crossmodal temporal disparate AV100, AV250 and AV400 with the sound always leading the visual stimulus respectively by 100, 250 and 400 ms; unimodal acoustic, UA). In the MI cluster of positions performance comparisons UV versus all the crossmodal conditions differ significantly. Proportions of UA (catch trials) in the two clusters of positions (i.e., MI vs. LI) did not differ significantly.

The post-hoc comparisons pointed out a significant difference in the MI cluster of positions between UV and SYNC (p(y) = .52 vs. p(y) = .69, p < .001) indicating that the SYNC condition provided a significant performance improvement with respect the UV condition. Surprisingly, the post hoc t-test analysis on the MI cluster of positions, pointed out a significant enhancement in the visual detection performance in all the other temporally disparate crossmodal conditions. In fact in respect to the UV performance (p(y) = .52) a significant difference has been found with AV100 (p(y) = .69, p < .01), AV250 (p(y) = .64, p < .05) and AV400 (p(y) = .65, p < .01). Furthermore, all the other t-test comparisons did not show a significant difference (.4 < p < 1) indicating that the multisensory enhancement provided by the sound in the synchronous condition does not differ from that emerged in the other temporally delayed conditions. No difference between conditions has been found in the LI cluster of positions (.2 < p < 1; UV, p(y) = .93, SYNC, p(y) = .96, AV100, p(y) = .95, AV250, p(y) = .93 and AV400, p(y) = .91).

Taken together, these results support the principle of inverse effectiveness, given that the multisensory enhancement has been found in the MI cluster of positions (i.e., where visual stimuli are less reliable) and not for the LI cluster of positions (i.e., where stimuli were highly effective; see, e.g., Frassinetti, et al., 2002a, b, 2005; Hairston, et al., 2003a, b; Noesselt et al., 2010, for similar results). Finally, responses to catch trials were analysed to examine whether participants were less able to ignore the sounds presented alone in the MI than in LI cluster of positions. A pairwise comparison t-test did not reveal any significant difference (MI, p(y) = .045 and LI, p(y) = .034, t(13) = 1.13, p = .28).

Discussion

The present study provides evidence of an audiovisual integration effects in low vision individuals with crossmodal delayed audiovisual stimuli. Results show that a synchronous sound presented from the same spatial position significantly enhances the performance of low vision individuals in a yes/no visual detection task as compared to the condition where the visual stimulus was presented in isolation. Moreover, the significant enhancement is observed only in the spatial positions in which the unimodal visual performance is mostly deteriorated (i.e., only in the MI cluster of positions) highlighting the role of the visual stimulus reliability in visual acoustic crossmodal tasks (e.g., Noesselt et al., 2010). To our surprise, the results showed a significant performance enhancement as compared to the unimodal visual condition in all the temporally delayed crossmodal conditions with the sound lagging the visual stimulus by various SOAs (i.e., AV100, AV250 and AV400).

It has been demonstrated that sensory integration can take place also between stimuli that are not temporally coincident, but which fall within the 'temporal window' of integration (Meredith et al., 1987; Spence & Squire, 2003). Although these values seem large in comparison with some previous reports of crossmodal interaction (e.g., Lewald & Guski, 2003; Lewald, Ehrenstein & Guski 2001), other studies have nevertheless reported substantial crossmodal effects with similarly large disparities (e.g., Bermant & Welch, 1976; Bertelson & Radeau, 1981; Wallace et al., 2004). Wallace and colleagues found that the audiovisual interaction effects were modulated by the participant's perception (i.e., judgment) of the unity of the event. In that study, localisation bias (i.e., sound localisation towards the visual stimulus) and reports of perceptual unity occurred even with substantial spatial (i.e., 15 degrees) and temporal (i.e., 800 ms) disparities supporting results in low vision patients when spatial and temporal disparities are manipulated. As proposed by different authors (e.g., Alais & Burr, 2004; Ernst & Banks, 2002), the most reliable sensory signal (i.e., acoustic) seems to provide the highest contribution in the integration mechanisms found in low vision patients. In Experiment 3, results show a marginal significant effect also for spatially disparate stimuli (i.e., at 16 degrees). Wallace and colleagues (2004) found crossmodal integration with such disparity (i.e. 15 degrees) as well. They also found a significant crossmodal effect with temporally disparate audiovisual stimuli (i.e. even with a delay of 800 ms). The enhancement effect emerged for low vision patients in the spatial and temporal disparate conditions, may be the result of a lack in the reliability of the visual information that is perceived, together with the sound, as an unique crossmodal event as found by Wallace and colleagues' (2004) study (see also Hairston, et al., 2003b).

In this framework, the unity assumption model, in which both stimuli reliability and prior knowledge affects the participants' perception, seems to be a good candidate to explain the results. It might be that high-level multisensory association mechanisms (i.e., prior knowledge or the learned association between sound and visual stimulus) may have an influence on low-level interaction (i.e., at the perceptual level; Stein, Wallace, Stanford & Jiang, 2002). The association of the visual and acoustic modalities to a unique event provided by the information carried by the sound (in both Experiments 3 and 4 the sound was synchronous or leads the visual stimulus) may be at the basis of the multisensory enhancement observed with spatial and temporal disparities.

Lippert and colleagues (2007) demonstrated that the time necessary to learn new associations between different modalities or to disrupt old ones, seem to be very short. The authors were able to eliminate the enhancement effect provided by the sound found previously with another group of participants by varying the stimulus onset asynchrony between sound and target in the crossmodal conditions. In a first experiment, crossmodal audiovisual stimuli were always provided synchronously thus providing a significant enhancement in the visual detection. In a second experiment the sound might lead or lag the visual stimulus and the enhancement in the synchronous condition disappears. Lippert and colleagues interpreted the results as a shift in the decision criterion determined by the lack of information provided by the sound (i.e., high-level modulation determined by the learned dissociation between modalities). It is worth noting that participants in Lippert et al. study learned this dissociation very fast (i.e., in one experimental section). Furthermore, there is evidence that cat reared (i.e., adapted) with spatially or temporally disparate stimuli show integration enhancement with such disparate stimuli (Wallace & Stein, 2007). These findings indicate that our cognitive system seems to learn and adapt very fast to different associations between modalities in a useful way to optimize task demand (i.e., environmental constrains) and as a consequence, crossmodal integration enhancement take place if there is a stable relation between spatial and temporal cues of different modalities.

Indeed, the fact that all the temporal delayed crossmodal conditions provide a significant enhancement in respect the unimodal visual condition cannot rule out a general unspecific alerting effect induced by the mere presence of the auditory stimulus on bimodal trials (e.g., Posner, 1978, 1980). Thus, the aim of the next experiment is to prevent participants to construct a fixed rule of association between visual and acoustic modalities by introducing crossmodal trials in which the visual stimulus would be provided first (thus, the acoustic stimulus conveyed information would be not always present as in Lippert et al., 2007 study). If the enhancement effect found in the present experiment will disappear or change significantly, the hypothesis of a general unspecific alerting effect induced by the mere presence of the auditory stimulus on bimodal trials (e.g., Posner, 1978, 1980) could be ruled out.

Audiovisual interaction in low vision: Temporal disparity factors II

Thi work was done in collaboration with Valeria Occelli PhD

Introduction Experiment 5

Results of the Experiment 4 shows that for low vision patients, a spatially congruent crossmodal stimulation, either synchronous or with a delay between acoustic and visual stimuli, always provides a significant enhancement in the proportion of correct responses as compared to the performance for the unimodal visual condition. Surprisingly, there is a significant improvement in the performance also with a relatively large temporal disparity (i.e., in the AV400 condition, where the visual stimulus is 400 ms delayed in respect the sound).

As mentioned in the discussion of the results, this wide temporal window of integration might be determinate by the perception of a unique crossmodal event due to the constant spatial cue information and congruency provided by the sound. This might lead to the stabilization of a constant relation between the two modalities that, in turn, provides the crossmodal enhancement (e.g., cognitive high-level association and a consequent modulatory effects on low-level integration mechanisms).

It does not seem to be the mere presence of the sound to lead the participants to respond ‘yes’ in the crossmodal trials. Indeed, the proportion of catch trials (in both Experiment 3 and 4) in the two clusters of spatial positions (i.e., least impaired and the most impaired clusters, respectively MI and LI clusters) was very low, and most importantly, it did not differ significantly between the two clusters (thus, it was not related to the reliability of the visual stimulus; i.e., magnitude of the visual deficit). However, in the present experiment the aim was to prevent participants to construct a fixed rule of association between visual and acoustic modalities. To this aim, by balancing trials in which the acoustic stimulus conveyed information for the occurrence of the subsequent visual stimulus (i.e., the conditions used in the previous experiment) and trials in which this information was not present (i.e., visual stimulus first), it might be possible to reduce or eliminate the enhancement effect provided by the sound as found by Lippert and colleagues (2007).

Material and methods

Participants

Twelve low vision participants (5 female; mean age of 53 years; range from 19 to 73 years; see Table 6, p. 117) took part in the study. Participants were mostly members of the "Unione dei Ciechi e degli Ipovedenti" (Italian Association for Blindness and Low Vision) of Trento.

Patient	Age	Visus	Visual pathology	Duration
1	73	1/10	Exudative Maculopathy	5
2	72	1/10	Diabetic Retinopathy	8
3	52	1/20	Bilateral Congenital Glaucoma	47
4	55	1/10	Optical neuritis	22
5	72	1/20	Chorioretinopaty, Angioid Streaks	42
6	25	1/20	Acute Maculopathy	20
7	60	1/10	Maculopathy	33
8	70	1/10	Acute Degenerative Maculopathy	34
9	24	1/10	Nistagmus, Degenerative Retinopathy	24
10	65	1/20	Diabetic Maculopathy	13
11	39	1/10	Maculopathy	29

Table 6 Participants' age and information about clinical pathology.

Apparatus, stimuli, procedure and experimental design

The apparatus, materials, design, and procedure were as in the Experiment 3 and 4 (see Figure 8, p. 92) with the following exceptions. Two SOAs in which the visual stimulus was provided first were added (symmetrically chosen 250 and 400 ms, respectively VA250 and VA400). As in Experiment 4, to avoid a cueing sound's effect on gaze direction in the crossmodal conditions in which the sound was provided first (i.e., AV250 and AV400) an online control procedure of the participant's gaze position during the whole SOA has been used. For trials in which the visual stimulus was provided first or synchronously with the acoustic stimulus (i.e., VA250 and VA400 and SYNC) no further gaze control procedure was adopted in addition to that used to verify the fixation position at the start of each trial. Indeed, in trials in which the visual stimulus was provided first or synchronously with the acoustic stimulus, the acoustic stimulus did not provide any further information to be used by the participants to gaze towards a particular spatial position thus enhancing the detection of the visual stimulus.

On each trial, participants could have been presented (randomly and balanced in each block) with seven different conditions: a visual stimulus alone (i.e., unimodal visual condition, UV), an acoustic stimulus alone (i.e., unimodal acoustic condition, UA or catch trials), and the presentation of both visual and auditory stimuli (i.e. the five crossmodal condition with different SOAs, that is, VA400, VA250, SYNC, AV250 and AV400). The whole experimental section was divided in 10 blocks and lasted less than 2 hours with a rest between blocks. Each block consisted of 56 trials (i.e. for each visual position there were one UV trial, one UA trial, and one trial for each of the five different crossmodal conditions for a total of 56 trials for each block). At the end of the whole experiment each participant was presented with 560 trials in total.

To verify that the procedure was clear, before the actual experimental section, the participants were asked to undergo a brief test with the same experimental conditions (i.e., 10 to 20 trials randomly chosen between the conditions of one block).

Data analysis

Participants' performance was analysed by computing the proportion of 'yes' responses by following the same procedure as in the previous two experiments (i.e., 3 and 4). Thus, in relation to the 8 spatial positions, 56 values of proportions have been obtained, namely 8 values respectively for the UV, SYNC, VA400, VA250, AV250, AV400 and UA (catch trials) conditions. Finally, as in the previous experiments, proportion values related to each condition and each of the eight spatial positions have been ordered starting from the lowest to the highest as measured by the performance in the UV condition (i.e., ascending order). After ordering the data in such described way, the first position (i.e., the one in which the performance in the UV condition was the lower, that is, the most impaired visual position) could have been either in the periphery or in the centre, depending on the participant's visual deficit (see Table 7, p. 120 for details). Finally, proportion of correct responses corresponding to the four most impaired positions (i.e., MI positions, namely the first four ordered positions) and the proportion of correct responses for the least impaired/spared four positions (i.e., LI positions, namely the ordered positions from the fifth to the eighth) were clustered for each participant and condition (see Table 7, p. 120 for an example for the UV performance).

One participant has been discarded from the subsequent analysis because his UV performance in the first spatial position was very high (i.e., in the most impaired spatial position the UV performance was above 95% of correct responses).

	Unimodal visual (UV) performance p(y) and spatial position							
Patient	**Most Impaired positions (MI)**				**Less Impaired positions (LI)**			
1	.60 (4)	.70 (5)	.90 (1)	.99 (2)	.99 (3)	.99 (6)	.99 (7)	.99 (8)
2	.01 (5)	.99 (1)	.99 (2)	.99 (3)	.99 (4)	.99 (6)	.99 (7)	.99 (8)
3	.20 (1)	.20 (7)	.30 (6)	.50 (8)	.90 (4)	.99 (2)	.99 (3)	.99 (5)
4	.60 (6)	.70 (7)	.70 (8)	.80 (4)	.80 (5)	.90 (1)	.99 (2)	.99 (3)
5	.01 (4)	.01 (5)	.60 (6)	.90 (7)	.90 (8)	.99 (1)	.99 (2)	.99 (3)
6	.01 (1)	.01 (4)	.01 (5)	.01 (8)	.10 (3)	.40 (6)	.70 (7)	.90 (2)
7	.01 (4)	.70 (6)	.90 (2)	.99 (1)	.99 (3)	.99 (5)	.99 (7)	.99 (8)
8	.10 (5)	.90 (1)	.90 (8)	.99 (2)	.99 (3)	.99 (4)	.99 (6)	.99 (7)
9	.80 (4)	.99 (1)	.99 (2)	.99 (3)	.99 (5)	.99 (6)	.99 (7)	.99 (8)
10	.80 (7)	.90 (2)	.90 (8)	.99 (1)	.99 (3)	.99 (4)	.99 (5)	.99 (6)
11	.01 (4)	.01 (5)	.01 (6)	.90 (3)	.99 (1)	.99 (2)	.99 (7)	.99 (8)

Table 7 For each participant, the spatial positions have been ordered as a function of the performance in the Unimodal Visual (UV) condition. 'MI' indicates the cluster of the most impaired four positions; 'LI' the cluster of the less impaired four positions.

Results

A within participants analysis of variance ANOVA with the factors cluster of positions (i.e., MI vs. LI) and condition (i.e., UV, SYNC, VA400, VA250, AV250 and AV400) revealed a significant general effect of condition, F (5, 50) = 3.35, p < .01, a significant effect of cluster of positions F (1, 10) = 15.8, p < .01, and a significant interaction between condition and cluster of positions, F (5, 50) = 3.1, p < .05. Results show a difference between conditions, while the emerged interaction show that this differences may be present only in one of the two clusters of positions (i.e., MI vs. LI, see Figure 13, p. 121) as found in the previous two experiments.

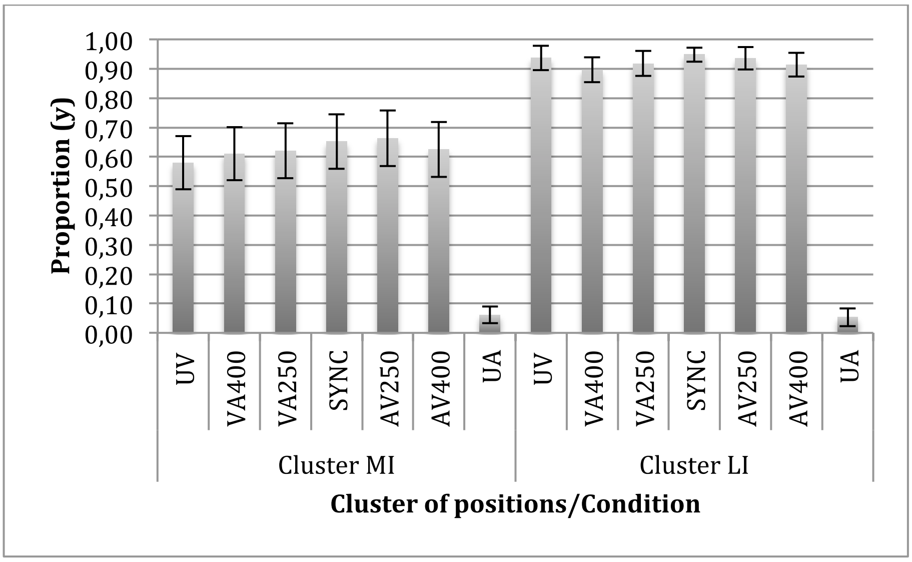

Figure 13 Mean proportions of ‘yes’ responses, reported for each cluster of positions (i.e., most impaired four positions, MI; less impaired four positions, LI), and each experimental condition (i.e. unimodal visual, UV; crossmodal temporal disparate VA250 and VA400 with the sound lagging the visual stimulus respectively by 250 and 400 ms; synchronous, SYNC; crossmodal temporal disparate AV250 and AV400 with the sound leading the visual stimulus respectively by 250 and 400 ms; unimodal acoustic, UA). In the MI cluster of positions performance comparison between UV and SYNC conditions differ significantly ($p < .05$). Proportions of UA (catch trials) in the two clusters of positions (i.e., MI vs. LI) did not differ significantly.

The post-hoc comparisons show a significant difference in the MI cluster of positions between UV and SYNC ($p(y) = .58$ vs. $p(y) = .65$, $p < .05$) indicating that the SYNC condition provided a significant performance improvement with respect the UV condition. Moreover, a marginal significant enhancement effect provided by the sound has also been found between UV and VA250 ($p(y) = .58$ vs. $p(y) = .62$, $p = .062$) and between UV and AV250 ($p(y) = .58$ vs. $p(y) = .66$, $p = .067$).

As expected results show a reduction of the enhancement effect provided by the sound found in Experiment 4 in which the difference between UV and AV250 was significant. Surprisingly, there is the same magnitude of visual enhancement (marginally significant) also in the VA250 condition when the sound is delivered 250 ms after the visual stimulus. No significant enhancement effect between UV and the two conditions in which the delay was 400 ms has been found (VA400 (p(y) = .61, p = 1 and AV400 (p(y) = .63, p = .9).

In line with the experimental hypothesis and with Lippert et al.'s (2007) findings, adding conditions in which the sound does not provide spatial cue information lead to an overall reduction in the visual detection enhancement. All the other t-test comparisons did not show any effect (.8 < p < 1) indicating that the effect provided by the sound in the crossmodal conditions did not differ. No difference between conditions has been found in the LI cluster of positions (.35 < p < 1; UV, p(y) = .94, SYNC, p(y) = .95, VA400, p(y) = .90, VA250, p(y) = .92, AV250, p(y) = .94 and AV400, p(y) = .91). As found in Experiment 3 and 4, this results support the principle of inverse effectiveness, given that the multisensory enhancement has been found in the MI cluster (i.e., where visual stimuli are less reliable) and not in the LI cluster of positions (i.e., where stimuli were highly effective; see, e.g., Frassinetti, et al., 2002a, b, 2005; Hairston, et al., 2003a, b; Noesselt et al., 2010, for similar results). Finally, responses to catch trials were analysed to examine whether participants were less able to ignore the sounds presented alone in the MI than in LI cluster of positions. A pairwise comparison t-test did not reveal any significant difference (MI, p(y) = .061 and LI, p(y) = .053, t(10) = 1.26, p = .23).

Discussion

As hypothesized, the introduction of crossmodal trials in which the acoustic stimulus did not conveyed any information because delivered after the occurrence of the visual stimulus (i.e., VA400 and VA250) seems to prevent participants to construct a stable rule of association between the two modalities. This happened despite the fact that in three out of five crossmodal conditions the sound was provided synchronously or before the visual stimulus (i.e., same crossmodal conditions used in Experiment 4 namely; SYNC, AV250 and AV400).

The results of this study shows that only the performance in the synchronous condition provided a significant effect in respect to the unimodal visual condition, while both the two temporally delayed conditions with 250 ms SOA (i.e. VA250 and AV 250) provided a marginal significant enhancement. Surprisingly, there is the same magnitude of visual enhancement (marginally significant) also in the VA250 condition when the sound is delivered 250 ms after the visual stimulus. However, this result is compatible with evidences in the literature that highlight differences in time to process visual and acoustic stimuli. In fact, due to these differences, it seems that the magnitude of the integration effect is higher when the acoustic stimulus is provided around 100 ms after the visual stimulus in respect to the synchronous condition (e.g., Lewald & Guski, 2003). Furthermore, results of Experiment 4 highlight the wide temporal window of integration in low vision patients that might be hypothesized to extend and to allow integration also when the sound lag the visual stimulus of around 250 ms (as found in the present experiment).

Unlike the results of Experiment 4 in which there was a significant detection enhancement in respect to the UV condition also in the condition in which the sound lead the visual stimulus of 400 ms (i.e., AV400), now this enhancement is no more significant. Furthermore, the magnitude of the crossmodal integration in AV400 condition did not differ significantly from that in the VA400 condition in which the sound lagged the visual stimulus of 400 ms. Indeed, for the VA400 condition, the enhancement was not expected because the visual stimulus has been provided too much early in respect to the sound (i.e., 400 ms).

A reduction of the enhancement effect provided by the sound was expected due to the fact that the lack of a constant relation between sensory modalities was hypothesized to lead to the suppression of the enhancement also in the synchronous condition. Thus, if this is the correct interpretation of the results, differences between the Experiments 4 and 5 are mainly due to the difference in the information provided by the sound and in turn, differences in participants' knowledge (e.g., decisional criterion).

Results of this study support previous finding (Wallace et al., 2004; Lippert et al., 2007) in which high-level associations between modalities seem to modulate participants' performance (e.g., the constant relation that leads to the perception of a unique event).

In conclusion, the results seem to mitigate the hypothesis of a general unspecific alerting effect induced by the mere presence of the auditory stimulus on bimodal trials (e.g., Posner, 1978, 1980) that in turn, do not explain the results.

General discussion on low vision

Taken together, the results of the three experiments with low vision patients show that a spatial and temporal coincident crossmodal audiovisual stimulation significantly improves the patients' ability to correctly detect the presence of the visual stimulus in respect the condition in which the visual stimulus is provided alone. Moreover, in all the three studies, the enhancement effect provided by the sound is significant only in the external spatial positions corresponding to the visual field positions in which patients' vision is mostly impaired (i.e., only in the MI cluster of positions). These results are in line with neurophysiological studies highlighting the three principles of multisensory interaction at the level of the superior colliculus (i.e., spatial, temporal and inverse effectiveness principles; e.g., Stein & Meredith 1993; Stein et al. 1996). Furthermore results support previous behavioural and neuroimaging results highlighting spatial and temporal constrains and the role of the visual stimulus reliability in visual acoustic crossmodal tasks (e.g., Bolognini et al., 2005; Frassinetti, et al., 2002a, b, 2005; Hairston, et al., 2003a, b; Noesselt et al., 2010).

Surprisingly, compared to the unimodal visual performance, results of Experiment 3 show a visual marginal significant detection enhancement for spatially disparate stimuli (i.e., 16 degrees of disparity), while results of Experiment 4 show a significant effect of the sound for all the temporal disparate crossmodal conditions in which the sound always lead the crossmodal stimulation (i.e., 100, 250 and 400 ms of SOA). These results suggest that for low vision patients both spatial and temporal boundaries of crossmodal integration are wider than those found in the majority of the behavioural studies reported in the literature (e.g., Lewald & Guski, 2003; Lewald, Ehrenstein & Guski 2001). Other studies have nevertheless reported substantial crossmodal effects with similarly large disparities (e.g., Bermant & Welch, 1976; Bertelson & Radeau, 1981; Frassinetti, et al., 2002a, b; Wallace et al., 2004).

One possible explanation for these results could be an inverse ventriloquism effect in which the less reliable visual stimulus is captured by the most reliable acoustic spatial/temporal cue as suggested by Alais and Burr (2004) and Ernst and Banks (2002). The inverse ventriloquism effect might be stronger in low vision patients than that observed in other less ‘ecological’ conditions (i.e., in which visual stimulus reliability has been manipulated artificially). However, results of Experiment 5 in which information provided by the sound was lowering down (because not always the sound was leading the crossmodal stimulation), highlight the effect of the manipulation of such association magnitude between modalities attributed by the participants. Indeed, a significant reduction in the enhancement magnitude between Experiment 4 and 5 has been found. However, results did not show a dramatic change as found by Lippert et al. (2007), where the reduction of sound information (i.e., weight of associations with the visual stimulus) eliminated the crossmodal enhancement in the synchronous condition. A model of optimal combination could explain this effect because, as hypothesized, in Experiment 5 the reliability of the sound was lower than in the Experiment 4. However, this effect seems to be better explained by a more complex model that contemplates both stimulus reliability and cognitive factors (e.g., prior knowledge and task instruction).

Wallace and Stein (2007) show that cats reared with spatially disparate crossmodal stimuli learn to bind them onto the same external event and this modification affects integration mechanisms at the level of the SC. Lewald and Guski (2003) found an effect of task instructions on crossmodal binding, adding evidence to a high-level cognitive modulation on participants’ response supporting the unity assumption model in which the effect of task instructions is encompassed.

Lippert et al. (2007) together with Wallace et al. (2004) findings indicate that our cognitive system seems to learn and adapt very fast to different associations between modalities in a useful way to optimize task demand (i.e., environmental constrains). Furthermore, the outcome of such adaptation may exploit crossmodal integration enhancement only if there is a stable relation between spatial and temporal cues of different modalities that lead the perceiver to bind the single modalities in a unique external event. This framework suggests that high-level association cortices seem to looking for a useful constant relation between sensory modalities to be learned (i.e. causality) and exploit it to optimize task demands (i.e., stimulus reliability and task instructions) and thus the behaviour. This in turn, is likely to exert a modulation of the integrations mechanisms governing early stages of sensory processing (e.g., at the level of the SC, see Stein et al. 2002; Wallace & Stein, 2007) by means for instance, of a feed forward connection. How fast may be this plasticity, how match well established relation might be modified or generalized and what kind of limits there are in establishing new relations is still to be deepened.

The present studies cannot provide information regarding whether the enlargement of spatial and temporal windows of integration effects found in the three experiments with low vision patients might be due to a fast learning of a new rule to bind visual and acoustic stimuli (Lippert et al., 2007; Wallace et al., 2004) or to an inverse ventriloquism effect (Alais & Burr, 2004) or a combination of the two. Further studies are needed to better clarify whether participants actually learn a new rule and eventually how fast can be built this new association rule thus shading lights on the influence of different aspects of multisensory integration in low vision patients.

Results of the three experiments highlight the low rate of catch trials and the lack of a significant difference between catch trials in relation to the magnitude of the patients' visual impairment. This result indicates that participants did not report to have seen the visual stimulus just because the presence of the sound and suggests that participants' response strategy was unrelated from the visual stimulus reliability (i.e., patients' impaired/least-impaired spatial positions). Furthermore, the lack of a significant difference between unimodal visual condition and crossmodal stimulation with 32 degrees of disparity in Experiment 3 together with the lower enhancement provided by the sound in Experiment 5 in respect Experiment 4 due to the manipulation of acoustic stimulus reliability suggest that the hypothesis of a general unspecific alerting effect induced by the mere presence of the auditory stimulus on bimodal trials (e.g., Posner, 1978, 1980) can be likely ruled out.

The results of the present study might provide useful insight for future audiovisual training rehabilitation programs for recovering visual impairments also in low vision patients. Passamonti and colleagues found that an audiovisual training could produce long lasting visual improvements in hemianopic patients (Passamonti, Bertini & Làdavas, 2009). By allowing eye movements, Passamonti and colleagues asked participants to gaze towards the spatial positions from which the sound was delivered. Participants were informed that visual stimulus would be presented later in the same positions. In this way, sounds were used as cues to inform the participants to gaze in the right spatial position thus improving visual detection. The learned visual strategy was then generalized also to the unimodal visual condition providing an improvement also for visual stimuli in isolation. Furthermore, this advantage was also transferred to other activity as visual search, reading, and in general to activity of daily life and was stable after one year follow-up.

Until now, the majority of low vision rehabilitation has focused only on the visual modality (Liu et al., 2007; Markowitz, 2006; Nilsson, Frennesson, & Nilsson, 2003). The multisensory crossmodal effect emerged in this study, provides new insights on multisensory integration mechanisms in patients with visual deficit and seems to have good chances to provide a further step forward for low vision rehabilitation.

General discussion

The findings emerged in the present dissertation bring new insights to deepen understanding the complexity of the cognitive mechanisms involved in audiovisual interaction.

Results of the first study show that the trans-saccadic effect influenced participants' performance in a different way in relation to the two sound conditions. The spatial miss-localization effect of the static deviated gaze highlighted in different studies (Zwiers et al. 2004; Bulkin & Groh 2006; Lewald and Ehrenstein 1996a, b; Lewald 1997, 1998) seems likely to be one factor to explain the results. However, it is likely that with free field sounds the spatial representation of the first sound held in memory might be 'remapped' eye-centrically. Despite the lack of the evidences of such trans-saccadic effect in the auditory domain, this possible interpretation of our results is in line with two studies that quantify this error in the visual domain (Henriques & Crawford 2000; Henriques et al., 1998).

The hypothesis is that (mainly) in the free field condition, both trans-saccadic and static deviated gaze effects interact. Indeed, ecological acoustic stimuli have to share together with the visual information a common spatial reference frame starting from low-level structures as the SC (e.g., Jay & Sparks, 1984, 1987; Meredith & Stein, 1990) involved in saccade generation (e.g., Stein & Clamann, 1981; Groh & Sparks, 1996). Eye movements can thus influence auditory spatial cognition especially for free field sounds and for judgments that do not require a spatial motor response.

Results of the second study confirm Hidaka and colleagues' (2009) perceptual participants' bias in reporting the direction of the sound induced visual motion (SIVM) illusion. Furthermore, when participants were asked to gaze towards the position of the bar (instead to report the direction), a significant effect of the expected eye landing position has been found. The correlation between perceptual and visuo-motor SIVM illusion add some evidence to the debated issue regarding the relation between perception and action (e.g., Giorello & Sinigaglia, 2007 for a review) indicating that, at list in the SIVM illusion, the two cognitive domains are affected very similarly. From this point of view, illusions might be a good field of study to disentangle the relation between perception and action because the motor response is likely to be unaffected (or affected in a minor way) by decisional processes that are not linked to the perception that guide motor responses.

Results of the three experiments of the last study with low vision patients show for the first time an auditory crossmodal enhancement on visual detection. Spatial and temporal coincident crossmodal audiovisual stimulation significantly enhances the visual stimulus detection compared to the unimodal visual condition. Furthermore, all the three studies consistently highlight the relation between the enhancement effect provided by the sound and the visual stimulus reliability. These results are in line with behavioural and neurophysiological studies highlighting the principles of multisensory interaction (i.e., spatial, temporal and inverse effectiveness principles; e.g., see Noesselt et al., 2010; Stein & Meredith 1993; Stein et al. 1996).

The marginal significant enhancement at 16 degrees of spatial disparity in the Experiment 3 and the significant improvement for all the temporal disparities (even with 400 ms SOA) in Experiment 4 suggest the presence of high-level cognitive modulation on low-level perceptual processes (Wallace et al., 2004). Results of Experiment 5 supports the latter hypothesis by showing that a significant reduction of the enhancement provided by the sound emerged in Experiment 4. Indeed introducing trials in which the sound lead (or lagged) the visual stimulus provided a reduction of sound's enhancement in line with Lippert and colleagues (2007) findings. This hypothesis of high-level cognitive modulation on low-level perceptual processes is in line with neurophysiological studies reporting SC's modulation exerted by cortical afferences (e.g., Jiang et al. 2001; Stein et al. 2002, 2009).

Evidences in the literature highlight the striking relation of visual and auditory information at the whole levels of information processing (see Klemen & Chambers, 2011 for a recent review) starting from primary sensory areas (e.g., Cappe et al. 2009; Ghazanfar et al., 2005; Kaas & Collins, 2004; Lakatos et al. 2007) and the SC (e.g., Meredith & Stein, 1990; Middlebrooks & Knudsen, 1984), which has a pivotal role in saccade generation (e.g., Groh & Sparks, 1996; Stein & Clamann, 1981). The involvement of the cortex in such complex multisensory network emerges by neurophysiological evidences showing modulation of cortical areas on the SC (e.g., Stein et al., 2002, 2009), and the presence of superadditivity in the superior temporal sulcus (Calvert et al., 2000; Stevenson et al., 2007) and left superior temporal gyrus (Foxe et al., 2002). Furthermore, in the posterior parietal cortex sensory signals are coded in common coordinate frames, such as auditory-visual or visual-somatosensory maps in eye-centred coordinates (Avillac et al., 2005; Battista et al., 1999; Cohen & Andersen, 2000; Schlack et al., 2005). Multisensory interaction has been recently found also in visual and auditory thalamus (Noesselt et al., 2010).

The high interconnection between different structures in the brain seem to be shaped the respond to the environment constrains and to provide a coherent ecological reppresentation of the external world.

The difference between free field and intracranial sounds founded in the study on auditory spatial representation might be related to the lack of most of the auditory spatial cues information of intracranial sounds. Intracranial sounds have not an ecological correspondence as that of free field sounds for the brain spatial representation. Modulation of neural responses to sound locations as a function of eye position has been described within the auditory cortex (Werner-Reiss et al., 2003), in the SC (Groh et al., 2001, 2003; Zwiers et al., 2004), and in the primary visual cortex of the monkey (Guo & Li, 1997). Translation of visual stimuli to a head-centred frame of reference may be at the basis to the convergence of visual and auditory stimuli to a shared spatial map (see King, 2009 for a review on this topic). Indeed, with each eye/head movement, stationary objects in the world change position on the retina and thus to perceive the world as stable retinotopic coordinates as to be remapped (e.g., Merriam et al., 2007; Nakamura & Colby 2002). It is possible that due to the plasticity of our cognitive system, an improvement in the conditions in which performance was mostly degraded due to the trans-saccadic effect, might be reduced by means of exercise with such 'artificial' conditions. However, the reported results highlight how differently an eye movement might affect naïve participants' auditory spatial representation in an acoustic perceptual judgement.

Results of the SIVM and those with low vision patients highlight the effect of spatial, temporal and the reliability of sensory modalities in multisensory interaction. Results of both the studies are in line with a number of studies showing that the interaction effects depend to visual stimuli reliability (e.g., Noesselt et al., 2010; Rowland & Stein, 2008; Stein et al., 1996).

Results of the experiment with low vision patients suggest that the principles governing multisensory integration are highly plastic and that there is no a strict a priori spatial relationship between different modalities for their integration. Direct experience with real multisensory events is necessary to develop and calibrate temporal register and spatial maps in different sensory modalities on SC's multisensory neurons (Ernst, & Bülthoff, 2004; Wallace & Stein, 2001, 2007). Causality is an example of an important 'rule' used by all the living being organisms to learn world physical effects. The brain seems to be able to store associations between modalities constituting a unique event.

However, a stable relation between spatial and temporal cues of different modalities is necessary because enhancement might occur. Subsequently, this memory will be more likely recalled if all, or some of, the constituents are experienced (Barraclough et al., 2005). Wallace and Stein (2007) and Lippert et al. (2007) findings indicate that our cognitive system seems to learn and adapt very fast to different associations between modalities in a useful way to optimize task demands (i.e., environmental constrains). Furthermore, different studies highlight the role of task instructions and prior knowledge on the interaction outcomes (e.g., Andersen et al., 2004; Aschersleben & Bertelson, 2003; Schutz & Lipscomb, 2007; Tuomainen et al., 2005). All these very complex cognitive mechanisms seem not to be accounted by the sole effect of stimulus reliability or optimal integration models (Alais & Burr, 2004; Ernst & Banks, 2002). The unity assumption model instead (Welch & Warren, 1980), accounts for stimulus properties, the way in which sensory information is acquired, prior participant knowledge and task instructions. Models based on Bayesian inference (e.g., Sato et al., 2007) encompassing the prior knowledge of the perceiver in addition to the reliability of the single modalities are a mathematical formalization very close to the unity assumption model and seems to be a very good predictor of the resulting integration outcome.

The audiovisual interaction observed in the SIVM seems more driven by low-level mechanisms and thus more robust to the high-level influences. To verify whether the SIVM effect is least influenced by cognitive modulation it would be possible decrease the reliability of the acoustic stimulus (for instance by adding to the reported SIVM experimental paradigm half of the trials in which the sound moves up/down vs. left/right). In this way, as found with low vision patients, it would be possible to decrease the weight of association attributed by participants to the audiovisual event, and to find a magnitude decrement in the left right direction of the SIVM illusion. It would be also interesting to check what will happen if, instead to use the headphones, acoustic stimuli would be delivered with free field loudspeakers very close to the external spatial position in which the visual stimulus is delivered. It is likely that the SIVM illusion would be stronger in magnitude due to the effect of the principle of spatial congruency.

For low vision patient instead, it would be very useful to know whether virtualized 3D sounds provided through headphones might also enhance visual stimuli detection. By using subjective head related transfer functions to virtualize the sounds it might be possible to maximise the conveyed spatial cues information. This point is not trivial due to the fact that an audiovisual rehabilitation paradigm for low vision patients might be based on ecological visual stimuli in a large field of view. In this framework, headphones would be the easiest way to cover a wide visual field of view allowing testing the patients in everyday life.

Future studies have also to be addressed to further investigate audiovisual interactions in low vision patients by using the functional magnetic resonance imaging (fMRI). It would be interesting to verify whether the crossmodal effect of a rehabilitation paradigm might induce cerebral plasticity behind patients' visual improvements and well being.

Results of the experiments here presented, far to constitute an exhaustive investigation, add some insights to still partial evidence on the considered topics. As always in the progress of science, new results open the door to new questions and highlights the limits of the experimental paradigms used in the previous investigations suggesting new way to better clarify and disentangle the results.

The cognitive mechanisms characterizing all the topics discussed in this dissertation are still inadequately explored suggesting the necessity of further investigations to provide a more extended and satisfactory assessment.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal cross-modal integration. *Current Biology*, **14**, 257-262.

Alais, D., Newell, F.N., & Mamassian, P. (2010). Multisensory Processing in Review: from Physiology to Behaviour. *Seeing and Perceiving*, **23**, 3-38.

Alink, A., Singer, W., & Muckli, L. (2008). Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *The Journal of Neuroscience*, **28**, 2690-2697.

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

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.

Andersen, T. S., Tiippana, K., & Sams, M., (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, **21**, 301-308.

Andersen, T.S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision Research*, **48**, 2537-2544.

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology*, **50**, 157-163.

Avillac, M., Ben Hamed, S., & Duhamel, J.R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. *The Journal of Neuroscience*. **27**, 1922-1932.

Avillac, M., Deneve, S., Olivier, E., Pouget, A., & Duhamel, J.R. (2005). Reference frames for representing visual and tactile locations in parietal cortex. *Nature Neuroscience*, **8**, 941-949.

Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., & Perrett, D.I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, **17**, 377-391.

Barth, D.S., Goldberg, N., Brett, B., & Di, S. (1995). The spatiotemporal organization of auditory, visual, and polysensory evoked potentials in rat cortex. *Brain Research*, **678**, 177-180.

Batista, A.P., Buneo, C.A., Snyder, L.H., & Andersen, R.A. (1999). Reach plans in eye-centered coordinates. *Science*. **285**, 257-260.

Battaglia P.W., Jacobs R.A, & Aslin R.N (2003). Bayesian integration of visual and auditory signals for spatial localization. *The Journal of the Optical Society of America A*, **20**,1391-1397.

Beauchamp, M.S., Lee, K.E., Argall, B.D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, **41**, 809-823.

Beauchamp, M.S., Yasar, N.E., Frye, R.E., & Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *Neuroimage*. **41**, 1011-1120.

Bermant, R. I., & Welch, R.B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Perceptual and Motor Skills*, **42**, 487-493.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin and Review*, **5**, 482-489.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Cross-modal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, **50**, 147-155.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, **29**, 578-584.

Bird, A.C. (1995). Retinal photoreceptor dystrophies. *American Journal of Ophthalmology*, **119**, 543-556.

Blauert, J. (1997). *Spatial hearing*. The psychophysics of human sound Localization. Cambridge MA: MIT.

Bremmer, F., Pouget, A., & Hoffmann, K. P. (1998). Eye position encoding in the macaque posterior parietal cortex. *European Journal of Neuroscience*, **10**, 153-160.

Bulkin, D.A., & Groh, J.M. (2006). Seeing sounds: visual and auditory interactions in the brain. *Current Opinion in Neurobiology*, **16**, 415-9.

Burr, D., Banks, M.S., & Morrone M.C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, **198**, 49-57.

Caclin, A., Soto-Faraco, S., Kingstone, A., & Spence, C. (2002). Tactile 'capture' of audition. *Perception & Psychophysics*, **64**, 616-630.

Calvert, G., Spence, C., & Stein, B.E. (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Calvert, G.A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology Paris*, **98**, 191-205.

Calvert, G.A., Campbell, R., & Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, **10**, 649-657.

Cappe, C., Murray, M.M., Barone, P., & Rouiller, E. (2010). Multisensory facilitation of behavior in monkeys: effects of stimulus intensity. *Journal of Cognitive Neuroscience*, **22**, 2850-2863.

Cappe, C., Rouiller, E.M., & Barone, P. (2009). Multisensory anatomical pathways. *Hearing Research*, **258**, 28-36.

Chelazzi L., Biscaldi M., Corbetta M., Peru A., Tassinari G., & Berlucchi G. (1995). Oculomotor activity and visual spatial attention. *Behavioural Brain Research*, **71**, 81-88.

Cheong, A.M., Legge, G.E., Lawrence, M.G., Cheung, S.H., & Ruff, M.A. (2008). Relationship between visual span and reading performance in age-related macular degeneration. *Vision Research*, **48**, 577-588.

Choe, C.S., Welch, R.B., Guilford, R.M., & Juola, J.F. (1975). The "ventriloquist effect": Visual dominance or response bias? *Perception & Psychophysics*, **18**, 55-60.

Cohen, Y.E., & Andersen, R. A. (2000). Reaches to sounds encoded in an eye-centered reference frame. *Neuron*, **27**, 647-652.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clinical Neurophysiology*. **113**, 495-506.

Crossland, M.D., & Rubin, G.S. (2006). Eye movements and reading in macular disease: further support for the shrinking perceptual span hypothesis. *Vision Research*, **46**, 590-597.

Crovitz, H.F., & Daves, W. (1962). Tendencies to eye movement and perceptual accuracy. *Journal of Experimental Psychology*, **63**, 495-498.

De Jong, P. T. (2006). Age-related macular degeneration. *The New England Journal of Medicine*, **355**, 1474-1485.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, **91**, 176-180.

Doyle, M.C., & Snowden, R.J. (2001). Identification of visual stimuli is improved by accompanying auditory stimuli: the role of eye movements and sound location. *Perception*, **30**, 795-810.

Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, **10**, 731-735.

Dufour, A., Després, O., & Pebayle, T. (2002). Visual and auditory facilitation in auditory spatial localization. *Visual Cognition*, **9**, 741-753.

Duhamel, J.R., Colby, C.L., & Goldberg, M.E. (1998). Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *Journal of Neurophysiology*, **79**, 126-136.

Ernst, M.O., & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**, 429-433.

Ernst, M.O., & Bülthoff, H.H. (2004). Merging the senses into a robust percept. *Trends Cognitive Science*, **8**, 162-169.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience*, **22**, 5749-59.

Falkenberg, H.K., Rubin, G.S., & Bex, P.J. (2007). Acuity, crowding, reading and fixation stability. *Vision Research*, **47**, 126-135.

Felleman, D.I., & Van Essen, D.C. (1991). Distributed hierarchical processing in the primate visual cortex. *Cerebral Cortex*, **1**, 1-47.

Fisher, G.H. (1968). Agreement between the spatial senses. *Perceptual and Motor Skills*, **126**, 849-850.

Fodor, J.A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Forster, B., Cavina-Pratesi, C., Aglioti, S.M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research*, **143**, 480-487.

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

Foxe, J.J., & Schroeder, C.E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, **16**, 419-423.

Foxe, J.J., Wylie, G.R., Martinez, A., Schroeder, C.E., Javitt, D.C., Guilfoyle, D., Ritter, W., & Murray, M.M. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *The Journal of Neurophysiology*, **88**, 540-543.

Frank, R.N. (1995). Diabetic Retinopathy. *Progress in Retinal and Eye Research*, **14**, 361-392.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002a). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, **147**, 332-343.

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., & Làdavas, E. (2005). Audio-visual integration in patients with visual deficit. *Journal of Cognitive Neuroscience*, **17**, 1442-1452.

Frassinetti, F., Pavani, F., & Làdavas, E. (2002b). Acoustic vision of neglected stimuli: interaction among spatially converging audiovisual inputs in neglect patients. *Journal of Cognitive Neuroscience*, **14**, 62-64.

Freides, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin*, **81**, 284-310.

Gallese, V., Fogassi, L., Fadiga, L., & Rizzolatti, G. (2002). Action representation and the inferior parietal lobule. In Prinz, W., & Hommel, B. (Eds.), *Attention and Performance XIX: Common Mechanisms in Perception and Action* (pp. 335-355). Oxford: Oxford University Press.

Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., & Logothetis, N.K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *The Journal of Neuroscience*, **25**, 5004-5012.

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

Giorello, G., & Sinigaglia, C. (2007). Perception in action. *Acta Biomed*, **78**, 49-57.

Goossens, H.H.L.M., & Van Opstal, A.J. (1999). Influence of head position on the spatial representation of acoustic targets. *The Journal of Neuroscience*, **81**, 2720-2736.

Gordon, M.O., Beiser, J.A., Brandt, J.D., Heuer, D.K., Higginbotham, E.J., Johnson, C.A., Keltner, J.K., Miller, J.P., Parrish, R.K., Wilson, M.R., & Kass, M.A. (2002). The Ocular Hypertension Treatment Study: Baseline Factors That Predict The Onset of Primary Open-Angle Glaucoma. *Archives of Ophthalmology*, **120**, 714-720.

Graziano, M. S. (2001). A system of multimodal areas in the primate brain, *Neuron*, **29**, 4-6.

Groh, J.M., & Sparks, D.L., (1996a). Saccades to somatosensory targets. II. Motor convergence in primate superior colliculus. *Journal of Neurophysiology*, **75**, 428-438.

Groh, J.M., & Sparks, D.L., (1996b). Saccades to somatosensory targets. III. Eye-position-dependent somatosensory activity in primate superior colliculus. *Journal of Neurophysiology*, **75**, 439-453.

Groh, J.M., Kelly, K.A., & Underhill, A.M. (2003). A monotonic code for sound azimuth in primate inferior colliculus. *Journal of Cognitive Neuroscience*, **15**, 1217-1231.

Groh, J.M., Trause, A.S., Underhill, A.M., Clark, K.R., & Inati, S. (2001). Eye Position Influences Auditory Responses in Primate Inferior Colliculus. *Neuron*, **29**, 509-518.

Guo, K., & Li, C.Y. (1997). Eye position-dependent activation of neurones in striate cortex of macaque. *Neuroreport*, **8**, 1405-1409.

Hairston, W.D., Laurienti, P.J., Mishra, G., Burdette, J.H., & Wallace, M.T. (2003a). Multisensory enhancement of localization under conditions of induced myopia. *Experimental Brain Research*, **152**, 404-408.

Hairston, W.D., Wallace, M.T., Vaughan, J.W., Stein, B.E., Norris, J.L., & Schirillo, J.A. (2003b). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, **15**, 20-29.

Harrington, L.K., & Peck, C.K. (1998). Spatial disparity affects visual-auditory interactions in human sensorymotor processing. *Experimental Brain Research*, **122**, 247-252

Henriques, D.Y.P., & Crawford, J.D. (2000). Direction-dependent distortions of retinocentric space in the visuomotor transformation for pointing. *Experimental Brain Research*, **132**, 179-194.

Henriques, D.Y.P., Klier, E.M., Smith, M.A., Lowy, D., & Crawford, J.D. (1998). Gaze-Centered Remapping of Remembered Visual Space in an Open-Loop Pointing Task. *The Journal of Neuroscience*, **18**, 1583-1594.

Hidaka, S., Manaka, Y., Teramoto, W., Sugita, Y., Miyauchi, R., Gyoba, J., Suzuki, Y., & Iwaya, Y. (2009). Alternation of sound location induces visual motion perception of a static object. *PLoS One*, **4**, e8188.

Hillis, J.M., Ernst, M.O., Banks, M.S., & Landy, M.S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, **298**, 1627-1630.

Hoffman, J., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, **57**, 787-795.

Holmes, N.P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Neuropsychologia*, **45**, 3340-3345.

Holmes, N.P. (2009). The principle of inverse effectiveness in multisensory integration: Some statistical considerations. *Brain Topography*, **21**, 168-176.

Howard, I.P., & Templeton, W.B. (1966). *Human spatial orientation*. New York: Wiley.

Hughes, H.C., Reuter-Lorenz, P.A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 131-153.

Jay, M.F., & Sparks D.L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature*, **309**, 345-347.

Jay, M.F., & Sparks, D.L. (1987). Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. *Journal of Neurophysiology*, **57**, 35-55.

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

Jiang, W., Jiang, H., Rowland, B.A., & Stein, B.E. (2007). Multisensory orientation behavior is disrupted by neonatal cortical ablation. *Journal of Neurophysiology*, **97**, 557-562.

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

Kaas, J., & Collins, C.E. (2004). The resurrection of multisensory cortex in primates: connection patterns that integrates modalities. In G.A. Calvert, C. Spence, & B.E. Stein (Eds.), *The handbook of multisensory processes* (pp. 285-293). Cambridge, MA: MIT Press.

Kadunce, D.C., Vaughan, J.W., Wallace, M.T., Benedek, G., & Stein, B.E. (1997). Mechanisms of within- and cross-modality suppression in the superior colliculus. *Journal of Neurophysiology*, **78**, 2834-2847.

Kaufman, L. (1974). Sight and mind. New York: Oxford University Press.

Kayser, C., Petkov, C.I., Augath, M., & Logothetis, N.K. (2005). Integration of touch and sound in auditory cortex. *Neuron*, **48**, 373-384.

Kayser, C., Petkov, C.I., Augath, M., & Logothetis, N.K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. *The Journal of Neuroscience*, **27**, 1824-1835.

King, A.J. (2009). Visual influences on auditory spatial learning. *Philosophical Transaction of the Royal Society B*, **364**, 331-339.

King, A.J., Parsons, C.H., & Moore, D.R. (2000). Plasticity in the neural coding of auditory space in the mammalian brain. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 11821-11828.

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, **416**, 172-174.

Klemen, J., & Chambers, C.D. (2011). Current perspectives and methods in studying neural mechanisms of multisensory interactions. *Neuroscience & Biobehavioral Reviews*. In Press.

Knudsen, E.I., & Brainard, M.S. (1995). Creating a unified representation of visual and auditory space in the brain. *Annual Review of Neuroscience*, **18**, 19-44.

Kopinska, A., & Harris, L.R. (2003). Spatial representation in body coordinates: evidence from errors in remembering positions of visual and auditory targets after active eye, head, and body movements. *Canadian Journal Experimental Psychology*, **57**, 23-37.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, **35**, 1897-1916.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., & Schroeder, C.E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, **53**, 279-292.

Lewald, J. (1997). Eye-position effects in directional hearing. *Behavioural Brain Research*, **87**, 35-48.

Lewald, J. (1998). The effect of gaze eccentricity on perceived sound direction and its relation to visual localization. *Hearing Research*, **115**, 206-216.

Lewald, J., & Ehrenstein, W.H. (1996a). Auditory-visual shift in localization depending on gaze direction. *NeuroReport*, **7**, 1929-1932.

Lewald, J., & Ehrenstein, W.H. (1996b). The effect of eye position on auditory lateralization. *Experimental Brain Research*, **108**, 473-485.

Lewald, J., & Ehrenstein, W.H. (1998a). Influence of head-to-trunk position on sound lateralization. *Experimental Brain Research*, **121**, 230-238.

Lewald, J., & Ehrenstein, W.H. (1998b). Auditory-visual spatial integration: a new psychophysical approach using laser pointing to acoustic targets. *Journal of the Acoustical Society of America*, **121**, 230-238.

Lewald, J., & Ehrenstein, W.H. (2000). Visual and proprioceptive shifts in perceived egocentric direction induced by eye-position. *Vision Research*, **40**, 539-547.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, **16**, 468-478.

Lewald, J., Ehrenstein, W.H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioural Brain Research*, **121**, 69-79.

Lippert, M., Logothetis, N.K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, **1173**, 102-109.

Liu, L., Kuyk, T., & Fuhr, P. (2007). Visual search training in subjects with severe to profound low vision. *Vision Research*, **47**, 2627-2636.

Logothetis, N.K., Pauls, J., Augath. M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, **412**, 150-157.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neuroscience*, **28**, 264-271.

Macmillan, N.A., & Creelman, C.D. (1991). *Detection theory: a user's guide*. Cambridge University Press, New York.

Markowitz, S.N. (2006). Principles of modern low vision rehabilitation. *Canadian Journal of Ophthalmology*, **41**, 289-312.

Marks, L.E., Szczesiul, R., & Ohlott, P. (1986). On the cross-modal perception of intensity. *Journal of Experimental Psychology: Human Perception and Performance*, **12**, 517-534.

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

Meredith, M.A., & Stein, B.E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, **56**, 640-662.

Meredith, M.A., & Stein, B.E. (1990). The visuotopic component of the multisensory map in the deep laminae of the cat superior colliculus. *The Journal of Neuroscience*, **10**, 3727-3742.

Meredith, M.A., Nemitz, J.W., & Stein, B.E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, **10**, 3215-3229.

Merriam, E.P., Genovese, C.R., & Colby, C.L. (2003). Spatial updating in human parietal cortex. *Neuron*, **39**, 361-373.

Meyer G.F., & Wuerger, S.M. (2001). Cross-modal integration of auditory and visual motion signals. *Neuroreport*, **12**, 2557-2560.

Middlebrooks, J.C., & Knudsen, E.I.A. (1984). Neural code for auditory space in the cat's superior colliculus. *The Journal of Neuroscience*. **4**, 2621-2634.

Middlebrooks, J.C., Clock, A.E., Xu, L., & Green, D.M. (1994). A Panoramic Code for Sound Location by Cortical Neurons. *Science*, **264**, 842-844.

Milner, A.D., & Goodale, M.A. (1995). The visual brain in action. Oxford: Oxford Press.

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

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Brain Research and Cognitive Brain Research*, **17**, 154-163.

Moschovakis, A.K., & Karabelas, A.B. (1985). Observations on the somatodendritic morphology and axonal trajectory of intracellularly HRP-labeled efferent neurons located in the deeper layers of the superior colliculus of the cat. *The Journal of Comparative Neurology*, **239**, 276 -308.

Mottonen, R., Krause, C.M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Research Cognitive Brain Research*. **13**, 417-425.

Murray, M.M., Molholm, S, Michel, C.M., Heslenfeld, D.J., Ritter, W., Javitt, D.C., Schroeder, C.E., & Foxe, J.J. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex*. **15**, 963-974.

Naïli, F., Despretz, P., & Boucart, M. (2006). Colour recognition at large visual eccentricities in normal observers and patients with low vision. *Neuroreport*, **17**, 1571-1574.

Nakamura, K., & Colby, C.L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 4026-4031.

Nilsson, U.L., Frennesson, C., & Nilsson, S.E.G. (2003). Patients with AMD and a large absolute central scotoma can be trained successfully to use eccentric viewing, as demonstrated in a scanning laser ophthalmoscope. *Vision Research*, **43**, 1777-1787.

Noesselt, T., Tyll, S., Boehler, C.N., Budinger, E., Heinze, H.J., & Driver, J. (2010). Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *Journal of Neuroscience*, **41**, 13609-13623.

Occelli, V., Spence, C., & Zampini, M. (2009). Assessing the effect of sound complexity on the audiotactile cross-modal dynamic capture task. *Quarterly Journal of Experimental Psychology*, **63**, 694-704.

Occelli, V., Spence, C., & Zampini, M. (2011). Audiotactile interactions in front and rear space. *Neuroscience and Behavioral Reviews*, In Press

Passamonti, C., Bertini, C., & Làdavas, E. (2009). Audio-visual stimulation improves oculomotor patterns in patients with hemianopia. *Neuropsychologia*, **47**, 546-555.

Pavani, F., Husain, M., & Driver, J. (2008). Eye-movements intervening between two successive sounds disrupt comparisons of auditory location. *Experimantal Brain Research*, **189**, 435-449.

Pavani, F., Husain, M., Làdavas, E., & Driver, J. (2004). Auditory deficits in visuospatial neglect patients. *Cortex*, **40**, 347, 365.

Pavani, F., Murray, M., & Schroeder, C.E. (2007). Rethinking mind, brain and behaviour through a multisensory perspective. *Neuropsychologia*, **45**, 467-468.

Peck, C.K. (1987). Visual-auditory interactions in cat superior colliculus: their role in control of gaze. *Brain Research*, **420**, 162-166.

Posner, M.I. (1978). *Chronometric Explorations of Mind*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

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

Radeau, M., & Bertelson, P. (1987). Auditory-visualinteractionand the timing of inputs: Thomas (1941) revisited. *Psychological Research,* **49,** 17-22.

Recanzone, G.H. (1998). Rapidly induced auditory plasticity: the ventriloquism aftereffect, *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 869-875.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Natural Review of Neuroscience*, **2**, 661-670.

Rock, J., & Victor, I. (1964). Vision and touch: an experimentally created conflict between the two senses. *Science*, **143**, 594-596.

Rorden. C., & Driver, J. (1999). Does auditory attention shift in the direction of an upcoming saccade? *Neuropsychologia*, **35**, 357-377.

Ross, J., Morrone, M.C., Goldberg, M.E., & Burr, D.C. (2001). Changes in visual perception at the time of saccades. *Trends Neuroscience*, **24**, 113-121.

Rowland, B.A., & Stein, B.E. (2008). Temporal profiles of response enhancement in multisensory integration. *Frontiers in Neuroscience*, **2**, 218-224.

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

Russo, G.S., & Bruce, C.J. (1994). Frontal eye field activity preceding aurally guided saccades. *Journal of Neurophysiology*, **71**, 1250 -1253.

Salmon, J.F. (1999). Predisposing factors for chronic angle-closure glaucoma. *Progress in Retinal and Eye Research*, **18**, 121-132.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O.V., Lu, S.T., & Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, **127**, 141-145.

Sanabria, D, Lupiáñez J, & Spence C (2007). Auditory motion affects visual motion perception in a speeded discrimination task. *Experimental Brain Research*, **178**, 415-412.

Sanabria, D., Soto-Faraco, S., & Spence, C., (2005). Assessing the effect of visual and tactile distractors on the perception of auditory apparent motion. *Experimental Brain Research*, **166**, 548-558.

Sanabria, D., Spence, C., & Soto-Faraco, S. (2007). Perceptual and decisional contributions to audio- visual interactions in the perception of apparent motion: a signal detection study. *Cognition*, **102**, 299-310.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian Inference Explains Perception of Unity and Ventriloquism Aftereffect: Identification of Common Sources of Audiovisual Stimuli. *Neural Computation*, **19**, 3335-3355.

Schlack, A., Sterbing-D'Angelo, S.J., Hartung, K., Hoffmann, K.P., & Bremmer, F. (2005). Multisensory space representations in the macaque ventral intraparietal area. *The Journal of Neuroscience*. **25**, 4616-4625.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception*, **36**, 888–897.

Sekuler, R., Sekuler, A.B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, **385**, 308.

Senkowksi, D., Saint-Amour, D., Höfle, M., & Foxe, J.J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage*. **56**, 2200-2208.

Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, **7**, 269-284.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, **408**, 788.

Slutsky, D., & Recanzone, G.H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, **12**, 7-10.

Soto-Faraco, S., & Kingstone, A. (2004). Multisensory integration of dynamic information. In G.A. Calvert, C. Spence, & B.E. Stein (Eds.), *The handbook of multisensory processes* (pp. 49-68). Cambridge, MA: MIT Press.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004a). Congruency effects between auditory and tactile motion: Extending the phenomenon of cross-modal dynamic capture. *Cognitive, Affective & Behavioral Neuroscience*, **4**, 208-217.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004b). Cross-modal dynamic capture: Congruency effect in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance*, **30**, 330-345.

Spence, C., & Driver, J. (Eds) (2004). *Crossmodal Space and Crossmodal Attention*. Oxford University Press, Oxford, UK.

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology*, **13**, 519-521.

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

Stein, B.E. (1998). Neural mechanism for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research*, **123**, 124-135.

Stein, B.E., & Clamann, H.P. (1981). Control of pinna movements and sensorimotor register in cat superior colliculus. *Brain, Behavior and Evolution*, **19**, 180-192.

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

Stein, B.E., & Stanford, T.R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*. **9**, 255-266.

Stein, B.E., London, N., Wilkinson, L.K., & Price, D.D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *Journal of Cognitive Neuroscience*, **8**, 497-506.

Stein, B.E., Meredith, M.A., Huneycutt, W.S., & Mc Dade, L. (1989). Behavioral indices of multisensory integration: orientation of visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience*, **1**, 12-24.

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

Stein, B.E., Wallace, M.W., Stanford, T.R., & Jiang, W. (2002). Cortex governs multisensory integration in the midbrain. *Neuroscientist*, **8**, 306-314.

Stevenson, R.A., Geoghegan, M.L., & James, T.W. (2007). Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Experimental Brain Research*. **179**, 85-95.

Stricanne, B., Andersen, R.A., & Mazzoni, P. (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *Journal of Neurophysiology*, **76**, 2071-2076.

Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound arrival time. *Nature*, **421**, 911.

Sumby, W., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, **26**, 212-215.

Teder-Sälejärvi, W.A., Di Russo, F., McDonald, J.J., & Hillyard, S.A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience*, **17**, 1396-1409.

Teramoto, W., Hidaka, S., & Sugita, Y. (2010). Sounds Move a Static Visual Object. *PLoS ONE*, **5**, e12255.

Thurlow, W.R, & Jack, C.E. (1973). Certain determinants of the “ventriloquism effect”. *Perceptual and Motor Skills*, **36**, 1171-1184.

Tuomainen, J., Andersen, T.S., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. *Cognition*, **96**, B13-22.

Varma, R., Fraser-Bell, S., Tan, S., Klein, R., & Azen, S.P. (2004). Prevalence of age-related macular degeneration in Latinos: the Los Angeles Latino eye study. *Ophthalmology*, **111**, 1288-1297.

Vliegen, J., Van Grootel, T.J., & Van Opstal, A.J. (2004). Dynamic sound localization during rapid eye-head gaze shifts. *The Journal of Neuroscience*, **24**, 9291-9302

Wallace, M. T., & Stein, B. E. (2000). Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences. *Journal of Neurophysiology*, **83**, 3578-3582.

Wallace, M.T., & Stein, B.E. (1994). Cross-modal synthesis in the midbrain depends on input from cortex. *Journal of Neurophysiology*, **71**, 429-432.

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

Wallace, M.T., & Stein, B.E. (2007). Early experience determines how the senses will interact. *Journal of Neurophysiology*, **97**, 921-926.

Wallace, M.T., Carriere, B.N., Perrault, T.J., Vaughan, J.W., & Stein, B.E. (2006). The development of cortical multisensory integration. *The Journal of Neuroscience*, **26**, 11844-11849.

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

Wallace, M.T., Meredith, M.A., & Stein, B.E. (1993). Converging influences from visual, auditory and somatosensory cortices onto output neurons of the superior colliculus. *Journal of Neurophysiology*, **69**, 1797-1809.

Wallace, M.T., Roberson, G.E., Hairston, W.D., Stein, B.E., Vaughan, J.W., & Schirillo, J.A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, **158**, 252-258.

Wang, Y., Celebrini, S., Trotter, Y., & Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *Biomed Central Neuroscience*, **9**,79.

Warnking, J., Dojat, M., Guérin-Dugué, A., Delon-Martin, C., Olympieff, S., Richard, N., Chéhikian, A., & Segebarth C. (2002). fMRI Retinotopic Mapping - Step by Step. *Neuroimage*, **17**, 1665-1683.

Warren, D.H., Welch, R.B., & McCarthy, T.J. (1981). The role of visual-auditory “compellingness” in the ventriloquism effect: implications for transitivity among the spatial senses. *Perception & Psychophysics*, **30**, 557-564.

Welch, R.B., & Warren, D.H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, **88**, 638-667.

Weleber, R.G. (1994). Stargardt’s macular dystrophy. *Archives of Ophthalmology*, **112**, 752-754.

Werner-Reiss, U., Kelly, K.A., Trause, A.S., Underhill, A.M., & Groh, J.M. (2003). Eye position affects activity in primary auditory cortex of primates. *Current Biology*, **13**, 554-562.

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

World Health Organization (2009). Visual impairment and blindness. Fact Sheet N°282 Retrieved March 28, 2011.

URL http://www.who.int/mediacentre/factsheets/fs282/en/index.html.

Wozny, D.R., Seitz, A.R., & Shams, L. (2008). Learning associations between simple visual and auditory features. *Journal of Vision*, **8**, 171.

Zwiers, M.P., Versnel, H., & Van Opstal, A.J. (2004). Involvement of monkey inferior colliculus in spatial hearing. *Journal of Neuroscience*, **24**, 4145-4156.