# Behavioral and electrophysiological correlates of cross-modal enhancement for unaware visual events

**Márta Szabina Pápai**

DOCTORAL THESIS UPF / 2017

Supervisor:

Dr. Salvador Soto-Faraco

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES

upf. Universitat Pompeu Fabra
*Barcelona*

*A családomnak.*

# Acknowledgements

Köszönettel legfőképp a családomnak tartozom, Anyukámnak, Pityókámnak, Marcinak és Keresztanyunak, a hosszú, töretlen türelemért, a bizalomért és azért a feltétel nélküli szeretetért, amely lehetőséget teremtett arra, hogy az otthontól távol, a megpróbáltatások egész hadával szembe tudjak nézni. E kihívás egyik kézzelfogható eredménye tulajdonképpen ez a disszertáció, amely nélkületek sokkal-sokkal nehezebb lett volna. Szüleimnek külön köszönöm azt a örök odaadást, amely a legelkeseredetebb és legkilátástalanabb pillanatokon is segített továbblendülni.

Ezen kívül köszönettel tartozom azoknak a barátoknak, akik támogattak, akik kitartottak mellettem, ebben a könnyűnek semmiképpen sem mondható időszakban, az otthoniak közül Amyéket és Bobót kiemelve.

Además, me gustaría dar las gracias a Salva, por su ayuda, confiabilidad y sobretodo por su paciencia. Gracias por apoyarme en los momentos difíciles, tanto a nivel profesional como personal, no sólo arriba sino últimamente también abajo del umbral de la consciencia.

Gracias a Manu, por haberme abierto las puertas de su mundo profesional y personal. Gracias por soportarme día a día, y

aguantar esta tesis. También por hacerme crecer como persona, por la lealtad, paciencia y cariño… en fin por la amistad.

Además, me gustaría dar las gracias a Nara por su apoyo y serenidad infinita, a Mire por su compasión tan verdadera como irónica, y también por utilizar repetitivamente la frase mágica: ’¡ooooh esto es fácil, lo podemos hacer!’. Gracias a Dasha por los momentos equilibrados en las conferencias, a Luis por compartir su sabiduría sobre la ley de los números grandes de la populación y hacernos reír en los momentos más desesperados. Gracias al peque Manu, no solo por compartir la misofonía famosa sino también ayudarme a ver que las noticias son dipolos y vienen juntas.

Me gustaría dar las gracias a toda la gente del MRG, pasados y presentes, y a mis compañeros de varios despachos. También gracias a Sandra y Gabriela por estar siempre disponible para ayudarme en los experimentos; a Silvia, Xavi, Florencia y Cristina que me han hecho, con su ayuda, paciencia y amabilidad, estos años más fácil.

También me gustaría dar las gracias a Luca por la comprensión y apoyo, por recibirme siempre con sonrisa y no perder la simpatía ni en esta época última tan frenética.

Y al final, gracias a Joan, Chiara, Alex y al grupo de Maxi por acogerme, por hacerme sentir a gusto aquí en Barcelona tan lejos de mi familia, y por seguir al lado mío. Sin vosotros hubiese sido mucho más difícil.

# Abstract

In everyday life, we are bombarded with information coming simultaneously from different senses. The information is combined in the brain in order to achieve a fast and accurate answer with the least effort possible. Many times, it happens through multisensory integration. However, may multisensory integration occur (automatically) in emergency situations when imperative action is needed, even if the cognitive resources are deployed to other ongoing activities? In order to resolve the query, we capitalized on uninformative, unpredictive, abrupt audiovisual stimulation in binocular rivalry, which let us measure bottom-up multisensory integration in a situation where top-down mechanisms are limited or even absent. The findings of this dissertation challenge previous views about unaware multisensory integration, since a cross-modal behavioral benefit not necessarily derives from bottom-up multisensory integration, instead, first, the individual contribution of the stimuli might offer sufficient explanation for the cross-modal facilitation, second, the putative influence of bottom-up attention cannot be dismissed.

# Rövid összefoglaló

Mindennapjainkban, minden időpillanatban az információk egész garmadával találkozunk. Ezek az információk párhuzamosan és egyszerre több modalitásból érkeznek, majd az észlelés során kombinálódnak, legkisebb energia-befektetessel is gyors és pontos választ eredményezve. Többnyire mindez a multiszenzoros integrációnak tudható be. Ugyanakkor kérdéses, hogy a multiszenzoros integráció létrejöhet-e olyan azonnali reakciót igénylő helyzetekben, amikor a kognitív erőforrásaink máshová, egy éppen esedékes viselkedés megvalósítására összpontosulnak. A kérdés megválaszolása érdekében váratlan, rövid időtartamú audiovizuális ingerek feldolgozását mértük binokuláris rivalitás (binocular rivalry) közben, amely lehetőséget teremtett a multiszenzoros integráció felülről jövő (top-down) kognitív folyamatok nélküli vizsgálatára. Az eredményeink némi interakcióba lépnek a korábbi kutatási adatokkal, miszerint az ilyenkor kapott multimodális facilitáció nem feltétlenül eredeztethető multiszenzoros integrációból, hiszen a külön prezentált, több modalitásból érkező ingerek integráció nélkül is statisztikai serkentést eredményezhetnek, továbbá az ingerek egyenkénti megragadásában az alulról jövő (bottom-up) figyelmi folyamatok szerepe sem kizárható.

# Resumen

En el día a día, recibimos una gran cantidad de información que procede simultáneamente de distintos sentidos. Esta información se combina en el cerebro para conseguir, con el mínimo esfuerzo posible, una respuesta rápida y precisa. Muchas veces, esto ocurre mediante la integración multisensorial. Sin embargo, ¿puede darse integración multisensorial (de forma automática) en situaciones de emergencia donde se necesita una acción imperativa, aunque los recursos cognitivos estén dedicados a otras actividades en curso? Para resolver esta cuestión, utilizamos estímulos audiovisuales de corta duración, presentados de forma repentina, sin contenido informativo ni predictivo en un paradigma de rivalidad binocular. Esto nos permitió medir la integración de abajo a arriba (bottom-up) en una situación en la que los mecanismos de arriba abajo (top-down) están limitados o incluso ausentes. Los hallazgos de esta tesis ponen en duda puntos de vista anteriores acerca la integración multisensorial no consciente, ya que un beneficio intermodal en el comportamiento no resulta necesariamente de la integración de abajo a arriba (bottom-up) sino que, primero, la contribución individual de cada uno de los estímulos (visual y auditivo) puede explicar satisfactoriamente la facilitación intermodal y, segundo, no se puede descartar una aparente influencia de la atención de abajo arriba (bottom-up).

# Prologue

Many of us have had the experience of getting frozen by the unexpected bark of a nearby dog catching us off-guard, whether it can reach us or not. A similar phenomenon occurs in a variety of situations; a horning car stops us immediately when crossing the street or a loud thunder blast startles us for a moment. Seemingly, abrupt, high arousing stimuli can 'capture' our attention regardless of where our cognitive resources are allocated to. This automatic capture of attention is useful in imperative situations, when immediate reaction to a new event in the environment is needed. Sensory information about events often arrives through various senses simultaneously and, many times, our reaction becomes faster and/or precise owing to cross-modal integration of these sensory stimuli, giving rise to more efficient stimulus capture. However, an outstanding question in the current literature is whether integration between modalities can happen automatically (thus pre-attentively), or else, it constrained by top-down processes. The answer to this question bears consequences in terms of theories about cross-modal integration, as well as applicability of cross-modal principles in various fields.

The brain has a natural tendency to treat multisensory information in a bound manner, sometimes resulting in a stronger signal leading to faster and more accurate responses. This cross-modal benefit can be attributed to 'multisensory integration' (Bertelson, Vroomen, De Gelder, & Driver, 2000; Meredith & Stein, 1983, 1996; Stein & Meredith, 1993; Talsma & Woldorff, 2005; Vroomen, Bertelson, & de Gelder, 2001). Yet, research in the last couple of decades has made it increasingly evident that the integration of information across modalities is not a monolithic process, and might happen via a variety of mechanisms (Lippert, Logothetis, & Kayser, 2007; Lovelace, Stein, & Wallace, 2003; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002). One important distinction, at a theoretical and practical level, is between cross-modal integration processes that happen in a purely bottom-up fashion, and those that are supported (or strengthened) by higher-order mechanisms via top-down modulation (Adam & Noppeney, 2014; De Meo et al., 2015; Hartcher-O, Soto-Faraco, & Adam, 2017; Koelewijn, Bronkhorst, & Theeuwes, 2010; Lippert et al., 2007; Moran & Reilly, 2006).

Why is this distinction relevant? The importance of understanding the detailed nature of the cross-modal mechanism has to do with the limitation of resources in the human brain. Bottom-up mechanisms are often automatic and therefore 'cheap' in resources, whereas processes that rely on top-down machinery are often less automatic and more costly in terms of resources consumed. In the complex scenarios of everyday life, our brain must deal with many events simultaneously, and processing resources are precious. Therefore, the question is whether multisensory integration and its associated advantages can occur via bottom-up automatic mechanisms or are guided by slower, more resource limited, top-down processes. Because top-down and bottom-up mechanisms are closely intertwined during the normal act of attention in everyday life, they must be dissociated using experimental paradigms. One potential way to measure bottom-up automatic multisensory integration without top-down selective modulation is based on paradigms that capitalize on reduced levels of visual awareness. We used binocular rivalry to present visual events below awareness and measure their interaction with sounds.

In the thesis, I focus on audiovisual stimulation, taking it as a representative example of multisensory situations. At first, I am going to review multisensory integration, more precisely, the modulation effect of sound on visual stimuli. It will be followed by a brief review of some aspects of visual attention targeting the distinction between top-down and bottom-up attention in time, in space and related to features. Next, I will present studies addressing attention with multisensory stimuli, presented above as well as below the level of awareness. Then, binocular rivalry will be reviewed, followed by binocular rivalry studies with attentional manipulation, and at last, in cross-modal situations where top-down attention was ruled out with a more or less extent. After the literature review, I will present four experiments, the main empirical contribution of this thesis. These experiments aimed to answer the question of whether bottom-up multisensory integration can happen automatically, with a special focus on conditions that prevent selective endogenous attention and expectation. We capitalized on binocular rivalry, a procedure that we will use to measure the modulation effect of audio events on unaware visual stimuli, using first behavioral measures, then a combination of behavior and ERPs. Finally, I am going to discuss our findings in the light of existing evidence, find convergences, address contradictions and raise some further, still unresolved questions.

# Table of contents

# Table of figures

## Figures of the introduction

## Figures of study 1

## Figures of study 2

.

# CHAPTER 1.

# General introduction

## 1.1. *Multisensory integration*

### 1.1.1. *Cross-modal redundancy effect: race model versus co-activation model*

For a long period of time, a dominating view in sensory processing stated that the initial period of perceptual analysis happens in an encapsulated manner. In other words, each sense is processed in isolation and the integration of information across senses only takes place at late stages of processing, in the association or multisensory brain areas ( Driver & Spence, 2000). As a consequence of this view, research was traditionally carried out on each modality separately, allowing fundamental principles of unimodal perception to be established (Evans & Whitfield, 1964; Hubel & Wiesel, 1962; Mountcastle, 1957). Notwithstanding, the technological and theoretical developments of the last decades led researchers to conclude that understanding of perceptual processing in its full breadth needs a conceptual change, that is, a new perspective of cross-modal perceptual processing has been adopted (Calvert & Thesen, 2004; Churchland, Ramachandran, & Sejnowski, 2005). The early studies on cross-modal interaction indeed demonstrated that behavioral response for sensory events is faster when they are accompanied with a concurrent stimulus in another modality, labeled as 'redundancy effect' (Hershenson, 1962; Stein, Meredith, Huneycutt, & McDade, 1989). In the beginning, various explanations were offered to the redundancy effect in alignment with the encapsulated point of view. Raab (1962) proposed a 'race model' suggesting that cross-modal gain can be nothing more than mere statistical facilitation. That is, each sensory signal 'travels' along a separate channel and the one that arrives first elicits the response, like in a race. By simple probability summation, counting with two sensory signals affords a faster reaction time on average than any of the two racers individually. In contrast, a later view of Miller (1982) proposed that the activation from different channels perhaps contributes to a common pool, and builds gradually in order to reach a single criterion for response initiation, labeled as 'co-activation model'. Yet, despite the enhanced behavioral outcome is similar in both cases, the underlying mechanisms are different and, it is not always trivial to tell apart the two causes of the cross-modal benefits.

### *1.1.2. Neuronal correlates in animals and the principles of multisensory integration*

Beyond the early behavioral findings and theoretical frameworks and, in line with co-activation model, foremost Meredith and Stein (1983; 1993) provided physiological proofs of integration of multisensory stimuli, based on single-neuron recordings of cat's Superior Colliculus (SC). Meredith and Stein (1983) demonstrated that deep layers of the SC contain neurons that respond to inputs from different senses, and the receptive fields corresponding to each modality actually belong to single neurons. The receptive fields of different modalities from the same neuron are often aligned in a relatively close spatial register, confirming a multisensory map of space. Meredith and Stein (1986) also observed that when two simultaneous signals are presented from quite disparate locations, then a so-called 'response depression' occurs on multisensory neurons, whose function is to inhibit responses to signals originating from the different external source. SC neural responses, thus, seem to be permissive to a limited spatial extent, while clearly show deterioration beyond a certain spatial separation. Besides the spatial aspect, the responses of SC also tend to accommodate for slight temporal differences deriving from transduction times, neural latencies and speed for stimuli coming from different senses. Altogether, there is some agreement on that integrative mechanisms based on cross-modal information essentially need proximity in space and time, indicating the ‘spatial and temporal principles‘ of multisensory integration (Meredith & Stein, 1986). In single cell literature, the evoked neural response of cross-modal stimuli is characterized by a larger number of action potentials so-called ‘spikes’ than the number of spikes elicited by the stronger individual sensory signal in isolation. In some cases, the activation originating from bimodal signals even exceeds the linear sum of spikes derived from the two unisensory signals, resulting in super-additive neuronal responses that are considered to be clear proof of multisensory integration. Hence, SC neurons instead of merely responding linearly to stimuli coming from different modalities rather seem to embody an integrative process. What is more, this cross-modal enhancement is maximal when the individual sensory stimuli are weakly effective, illustrating the third principle of multisensory integration, the ‘rule of inverse effectiveness’.

### *1.1.3. Principles of multisensory integration in humans*

Multisensory integration has been characterized by principles formed to favor efficient perceptual functioning. The rules discovered in animals have been, to some extent, confirmed in humans. Because our experiments capitalize on audiovisual modulation, therefore I am going to focus mainly on audiovisual protocols (Bolognini, Rasi, & Ládavas, 2005; Frassinetti, Bolognini, Bottari, Bonora, & Làdavas, 2005; Frassinetti, Bolognini, &

Làdavas, 2002; Làdavas, 2008; B. Stein, London, Wilkinson, & Price, 1996). According to this research, a sudden sound has been demonstrated to facilitate the detection of a visual event presented close in space and in time (Frassinetti, et al., 2002). The cross-modal benefit initially was though to happen within temporal asynchronies up to approximately 100 ms, reflecting the window in which neuronal peak activities of cross-modal stimuli still overlap (Bolognini, Frassinetti, Serino, & Làdavas, 2005). The cross-modal benefit furthermore happened to occur up to $16^o$ (Bolognini, et al., 2005) or $45^o$ of spatial disparity in azimuth ( Stein et al., 1996), or $14^o$ in elevation (Spence & Driver, 1997), putatively reflecting the different sizes of the receptive fields of audio and visual neurons. These findings suggest that multisensory integration in humans is constrained, similarly to animal findings, by spatio-temporal proximity. What is more, further confirming results of Bolognini et al. (2005) suggest that sudden (supra-threshold) sound helps to detect the visual stimuli at threshold level, showing a bigger improvement for less effective stimuli, indicating the third rule of multisensory integration, namely the principle of inverse effectiveness. The inverse relationship between the effectiveness of the stimuli and its integration has been repeatedly demonstrated (Hairston, Laurienti, Mishra, Burdette, & Wallace, 2003; Serino, Farnè, Rinaldesi, Haggard, & Làdavas, 2007; Stein et al., 1996), not always resulting in completely homogeneous findings though (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). As a possible explanation to these controversies Holmes (2007) suggested that despite of the rule of inverse effectiveness is grounded on anatomical and physiological mechanisms, the choice of physiological measurement and statistics by itself might also profoundly influence the likelihood of a given dataset to obey the law. Anyhow, in the big picture, one can claim that the rules of multisensory integration based on animal studies have been generally established also in humans, what is more, they seem to be even less restrictive with respect to time and space, as it was initially thought.

Beyond these promising results, some studies have gone even further by targeting possible audiovisual benefits to help patients with affected visual processing. Some clinical studies demonstrated that the localization of visual stimuli happened to be enhanced by sounds on the hemianopic (Frassinetti et al., 2005) and/or neglected visual hemifield of neurological patients (Frassinetti et al., 2005; Frassinetti, Pavani, & Làdavas, 2002; Làdavas, 2008). In parallel, it was also shown that the localization of a sound could be even shifted (ventriloquized) to the direction of a concurrent visual event, even if the visual stimulus was presented in the extinguished visual hemifield of hemineglect patients (Bertelson, Pavani, Ladavas, Vroomen, & De Gelder, 2000). Hence, multisensory integration seems to happen even in the absence of awareness about the visual event or even if sensory processing is not completely intact, highlighting the role of multisensory neurons which continue to respond to stimuli of the impaired sense and, prevent total stimulus deprivation (Làdavas, 2008). The relationship between sensory awareness and multisensory integration will be one of the key aspects of this thesis.

### *1.1.4. To what extent can cross-modal benefit be attributed to specific mechanisms of multisensory integration?*

The above reviewed evidence as well as numerous other findings suggest that an audio stimulus can have beneficial effects on the processing of a simultaneous visual event. Demonstrations of these benefits vary from detection improvement of a dimly flashed light (Frassinetti, Bolognini, et al., 2002; McDonald, Teder-Sälejärvi, & Hillyard, 2000; Teder-Sälejärvi, Russo, Mc Donald, & Hillyard, 2005) through enhanced discriminability of briefly presented visual stimuli (Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008; Vroomen & de Gelder, 2000) up to an increased perceived luminance of a light flash (Andersen & Mamassian, 2008; Stein et al., 1996), just to mention some. While many of these studies suggest that the cross-modal enhancement effect could be attributed to bottom-up multisensory integration processes, other evidences (Stein, 1998) rather highlight an essential role of top-down influences. Thus, it remained unresolved whether automatic multisensory integration can explain the cross-modal benefit, and if so, to what extent multisensory integration happens without the help of top-down processes, such as selective attention. Given the diversity of terminology in the literature, before following with a review of attentional mechanisms, it is important to make a note on the distinction between multisensory integration and multisensory interaction (for further information please find *Box 1*).

### *Box 1. Multisensory integration versus interaction*

In the literature, there is certain confusion about the meaning of some terms, such as for example multisensory integration versus multisensory interaction. Here, we use integration to refer to a combination between cross-modal stimuli that results in a new entity that cannot be derived from linear combinations of the unisensory components. A clear example is hearing the 'da' syllable in the McGurk illusion, when the sound of 'ba' is presented together with the lip movements of 'ga' (McGurk & MacDonald, 1976). Multisensory integration can also facilitate stimulus processing revealing in faster and more accurate detection or discrimination responses. When this benefit surpasses the prediction of the probability summation model, then the integration explanation is often accepted, over and above mere interaction. This behavioral benefit is typically accompanied by brain signals from multisensory neurons that diverge from the summed brain responses from the constituent unisensory neurons, resulting in non-linear effects (Stein & Meredith, 1993), even if not all of the multisensory neurons reveal in super-additive response (Meredith, Allman, Keniston, & Clemo, 2012).

In opposition to the integration, for multisensory interaction despite a stimulus from one modality has effect on a stimulus from another modality, it does not necessarily reveal in a new entity, or a supra-linear effect surpassing the probability summation model. The phenomenon has been labeled as 'interaction' (De Meo et al., 2015) or simply 'cross-modal effect' (Stein et al., 2010). Interactions amongst modalities can be based on more general mechanisms applicable to the unisensory as well as the cross-modal case, such as for example, attention orienting or priming. In principle, integration phenomena could be seen as a form of interaction, but not the other way around.

## ***1.2. Top-down and bottom-up attention***

### *1.2.1. The function of top-down and bottom-up attention*

Oftentimes daily scenarios contain a huge amount of information, which happens to be much more than it is possible or comfortable to process in a single moment. Nevertheless, we have the subjective experience of a seemingly effortless understanding of the environment, thanks to selective attentional mechanisms which pick relevant information out of the irrelevant ones (separate the wheat from the chaff). Attentional processes are modulated by knowledge and assumptions about the world surrounding us, by the sudden appearance of possibly relevant information, as well as by the state of the organism (alert, arousal…). The attentional mechanisms, furthermore, allow us to flexibly direct resources to a specific location (Posner, 1980), a particular moment (Coull & Nobre, 1998) or to a certain feature (Corbetta & Shulman, 2002); hence we can selectively process the information of the environment by prioritizing some goal-related aspects while ignoring others, that are not relevant.

Attention, thus, facilitates sensory processing. However, an important issue in the context of this thesis is that some effects resembling multisensory integration can be partially or totally explained by attention mechanisms, amongst other top-down processes (De Meo et al., 2015; Lippert et al., 2007; Moran & Reilly, 2006). In the cases where this proves to be true, then by the principle of parsimony, one would have to dismiss hypothetical 'multisensory integration' mechanisms to explain cross-modal effects.

Eventually, whilst top-down attentional effects might operate for the majority of 'conscious' sensory processing, under some conditions, stimulus presentation below awareness can allow capitalizing on bottom-up processes measuring multisensory integration without top-down attention, as it has been shown in some recent papers (Lunghi & Alais, 2013; Lunghi, Morrone, & Alais, 2014; Zhou, Jiang, He, & Chen, 2010; Zhou, Zhang, Chen, Wang, & Chen, 2012). This is the approach we used in this thesis. But first, in the following sections I will expand more on specific concerns regarding attention, before elaborating further on the relationship of attentional and multisensory effects.

### *1.2.2. Top-down versus bottom-up visual attention in space*

James (1890) suggested that we can select information in a 'voluntary' as well as an 'involuntary' manner. This functional distinction has been widely accepted, and forms the basis of many of the current ideas about selective attention. The voluntary system corresponds to our ability to willfully monitor information in a goal-directed manner, while the latter involuntary system rather represents an automatic, stimulus-driven orienting response to the location of a sudden, salient stimulus. The voluntary attention system is also called ‘endogenous’ or ‘top-down’, while the involuntary is rather known as ‘exogenous’, ‘bottom-up’ attention (Corbetta & Shulman, 2002; Desimone & Duncan, 1995). Famously, Posner (1980) illustrated these principles, showing that certain cues could shift the focus of spatial attention towards the cued location in a goal-directed or in a stimulus-driven manner. The goal directed way allowed resources to distribute in space based on cue validity, while the stimulus-driven manner was more spontaneous regardless of the informative content. This latter bottom-up attention, as opposed to top-down attention, was independent of the allocation of cognitive resources, therefore seemed to be less susceptible to interference from other mechanisms (Chica & Lupianez, 2009).

### *1.2.3. Top-down versus bottom-up visual attention in time*

Attention orienting can also occur in the domain of time. Like spatial orienting, temporal orienting can occur in a top-down as well as in a bottom-up manner (Coull & Nobre, 1998; Kingstone, 1992). In Coull and Nobre’s study, a symbolic cue indicated the appearance of an upcoming target in one out of two possible, known time intervals. The study revealed in similar effect that was given by spatial cueing paradigm, namely better performance for targets appearing at the cued moment. Please note that endogenous temporal attention can and should be disentangled from expectation, although both can arise from top-down regulation (See *Box 2*).

*Box 2. Differences between endogenous temporal attention and expectation*

Regardless of the intuitive similarity, and their similar behavioral consequences, the concepts of endogenous temporal attention and of expectation can be dissociated. Endogenous temporal attention refers to the state of anticipation of a relevant future event, while the temporal expectation is rather characterized by the extraction of temporal regularities from the environment (Todorovic, Schoffelen, Van Ede, Maris, & De Lange, 2015). In other words, while attention prioritizes stimulus processing on the basis of motivational relevance, expectation rather constrains visual interpretation based on prior likelihood (Summerfield & Egner, 2009). The outcomes of attention and expectation are not always easy to tell apart, specially using behavioral tasks alone (Todorovic et al., 2015). Yet, some experimental demonstration of sensory responses to expectancy violation by ERP recordings (Czigler, Balázs, & Winkler, 2002; Opitz, Mecklinger, Von Cramon, & Kruggel, 1999; Rinne, Degerman, & Alho, 2005).

The temporal attention can be also captured in a bottom-up manner, indicating the moment of potentially important changes in the environment. Bottom-up temporal attention can be captured by irrelevant temporal onsets of a target in task-relevant (Chun, 1997; Maki & Mebane, 2006; Wee, Chua, & Chua, 2004) as well as in task-irrelevant situations (Dalton & Lavie, 2006). The visual system, thus, tends to devote limited processing resources to the most relevant sensory inputs based on location (Posner, 1980) as well as on time (Coull & Nobre, 1998; Kingstone, 1992) in a goal-directed and in a stimulus-driven manner. These attention mechanisms in principle can operate within as well as across modalities and therefore are potentially confusable with cross-modal effects arising from integration (see below).

### *1.2.4. Top-down versus bottom-up visual attention based on features*

Additionally, resources can also be allocated and guided on the basis of parts or features of objects (Folk, Remington, & Johnston, 1992). The brain tends to form, maintain and update representations of relevant objects to parse visual scenes and drive attention according to their features (Boynton, 2009; Haenny, Maunsell, & Schiller, 1988; Martinez-Trujillo & Treue, 2004; Maunsell & Treue, 2006; Treue & Martínez Trujillo, 1999; Yantis, 2000) or potential threat (Corbetta & Shulman, 2002). Feature-based attention, similarly to the other forms discussed above, can occur in a goal-directed top-down manner (Corbetta & Shulman, 2002; Desimone & Duncan, 1995; Egeth, Virzi, & Garbart, 1984; Liu, Larsson, & Carrasco, 2007; Spivey & Spirn, 2000), or in a stimulus-driven bottom-up way, based on salience of features (Corbetta & Shulman, 2002; Duncan & Humphreys, 1989; James, 1890). In line with an evolutionary perspective, this latter form favors the abrupt, unexpected, novel or potentially dangerous attributes of objects that take high priority and are processed immediately even at the expense of other ongoing processes, and their associated behaviors and neural activity (Corbetta & Shulman, 2002).

### *1.2.5. Cross-modal attention*

Even though the attentional mechanisms reviewed above were based on the visual domain, similar top-down as well as bottom-up mechanisms operate between modalities (Corbetta & Shulman, 2002; Egly, Driver, & Rafal, 1994; Macaluso, Eimer, Frith, & Driver, 2003; Macaluso, Frith, & Driver, 2002; Steven Yantis et al., 2002). In this thesis, the distinction of the different attentional mechanisms is an important theme, as we aimed to rule out the top-down attention and recline exclusively on bottom-up processes. Furthermore, once bottom-up sensory processes are singled out, then one must further attribute between bottom-up capture of attention and integration. On the one hand, attention capture is based on exogenous mechanisms and can lead to redundant effects, identified with the 'race model'. On the other hand, bottom-up multisensory integration refers to the mechanism of cross-modal convergence of information, identified with the 'co-activation model' (both models are described in Chapter 1., Section 1.1.1.). Indeed, the attribution between bottom-up attentional capture based on the independent contribution of sensory signals or bottom-up multisensory integration is not always straightforward. Hence, some researchers look for non-linear responses by calculating probability summation on behavioral responses or by electrophysiological recordings (Fort, Delpuech, Pernier, & Giard, 2002a; Molholm et al., 2002a; Murray, Foxe, & Wylie, 2005; Otto & Mamassian, 2012; Pannunzi et al., 2015; Talsma & Woldorff, 2005) to achieve such a distinction. Beyond attentional capture, attention in general and multisensory integration are quite intertwined process, as I am going to discuss in the following chapter.

## *1.3. Multisensory integration and attention*

### *1.3.1. To what extent multisensory integration is independent from top-down attention?*

A salient stimulus calls for bottom-up attention (Corbetta & Shulman, 2002; Coull & Nobre, 1998; Posner, 1980) and, whenever the information comes from various modalities at the same approximate time and/or location, it results in a more salient attentional cue (Aller, Giani, Conrad, Watanabe, & Noppeney, 2015; Matusz & Eimer, 2011; Van der Burg et al., 2008). The superior effectiveness of cross-modal events in capturing attention suggests integration between stimuli, in line with the co-activation model. Yet, whether such a multisensory integration can happen without top-down attention has been the matter of debate (De Meo et al., 2015; Hartcher-O et al., 2017; Talsma et al., 2010). One strategy to disentangle multisensory integration (MSI) from attention is to try and find if MSI can happen in lack of top-down attention and, the outcome of MSI eventually provides the basis for bottom-up attentional capture (a representative example is presented on **Fig. I/1**.**A-B**). If this is true, then one can conclude that MSI happens prior to, and in the absence of, more general attention mechanisms.

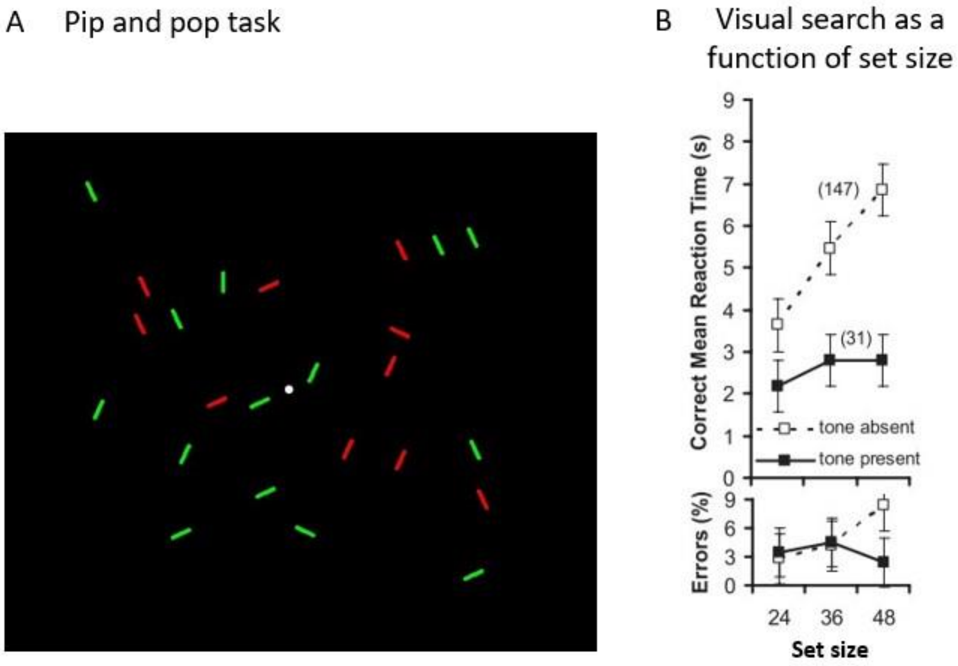


**Figure I/1. A-B** A representative example of multisensory integration, which outcome captures bottom-up attention (Van der Burg et al., 2008). ***A.*** *Pip and pop task.* In a visual search task, participants had to look for a flickering target line (i.e., green vertical line in this case) within the colorful tilted distractor lines. The flickers of the target green line were or were not accompanied by a simultaneous tone. ***B.*** *Visual search as a function of set size.* The serial visual search turned to be parallel (i.e., not sensitive to the set size) when sound was presented, suggesting an automatic integration of the temporally synchronized auditory and visual stimuli (See Van der Burg et al. (2012; 2011) for criticism, explained in detail below in the main text).

Matusz and Eimer (2011) found that audiovisual stimuli tended to elicit a larger spatial cueing effect than a corresponding visual cue and, since tones were task-irrelevant (i.e., independent of top-down task-set), the effect has been attributed to multisensory enhancement of bottom-up attentional cueing. Similarly, Van der Burg et al. (2008) demonstrated a so-called ‘pip and pop’ effect, in which a spatially uninformative audio event drastically decreased the search times when synchronized with a color change of the visual target. Since the pip and pop effect appeared even with statistically uninformative audio stimuli, the influence of top-down attentional cueing was ruled out, giving rise to an interpretation based on bottom-up multisensory integration. In other words, according to van der Burg et al. the sound putatively integrated with the visual target and increased the saliency of the abrupt (color) feature change, which later captured attention in a bottom-up manner (**Figure I/1 A-B**). In line with van der Burg et al., others also favored the explanation of multisensory integration for similar cross-modal benefits, as it seemed to happen regardless of task-relevance, when stimuli were attended but task-irrelevant (Cappe, Thut, Romei, & Murray, 2010) or passively viewed (Vidal, Giard, Roux, Barthélémy, & Bruneau, 2008).

In a different line, according to Lippert, Logothesis and Kayser (2007), De Meo (2015) and, the modeling data of Moran and Reilly (2006) many of the multisensory processes occurring at early stages may still be supported by top-down mechanisms. Van der Burg et al. in a later study (2011) indeed suggested that multisensory integration could not totally explain the pip and pop effect since whenever audio was presented simultaneously with visual distractors, behavioral cost did not occur on target search, something that would have been expected for automatic bottom-up multisensory integration. What is more, Van der Burg (2012) further argued that the pip and pop effect tended to be susceptible to the size of attentional window highlighting the relevance of attentional influence on the phenomenon. In line with this criticism, top-down processes might appear to explain other cases as well. Lovelace, Stein & Wallace (2003) demonstrated that a task-irrelevant light enhanced the detectability of a sound, even if the possibility of response bias was eliminated; what has been attributed to multisensory-mediated gain in stimulus detection. However, the simultaneous presentation of audio and visual signals could also result in a reduction of temporal uncertainty. Therefore, the finding suggests that sometimes multisensory gains cannot be completely disentangled from higher-order influence, like endogenous attention cueing in time or space, expectation, or response bias.

It has been proposed then that attention and multisensory integration might be inevitably intertwined, that is integrated cross-modal stimuli capture bottom-up attention with increased efficiently although such a cross-modal integration is regulated by top-down attention (Talsma et al., 2010). Koelewijn et al. (2010) suggested that the interplay between attention and multisensory integration is determined by the stage of processing at which the interaction happens, as cross-modal interactions can happen from subcortical early areas through to sensory cortical, or up to cortical association areas. Still, the question of to what extent multisensory integration happens in the absence of top-down attention or other top-down processes is still empirically unanswered.

### *1.3.2. ERP correlates of early multisensory integration with attention manipulations*

A plethora of multisensory studies have addressed behaviorally whether bottom-up multisensory integration can happen without the need (or prior to the allocation) of voluntary attention (Bertelson, Pavani, et al., 2000; Bertelson, Vroomen, et al., 2000; Vroomen et al., 2001), or if, instead, attention regulates multisensory integration in a fundamental way. Brain imaging techniques offer promising means to answer this question. Electroencephalography (EEG) and event related potentials (ERPs), amongst other techniques, are widely used tools to measure the expression of multisensory integration under varying conditions of attention.

Some neural correlates of cross-modal processing have been observed at early latencies after the stimulation, as a sign of bottom-up convergence (e.g., Foxe & Schroeder, 2005; Molholm et al., 2002a), although these results must be taken cautiously as many times the influence of top-down processes again could not be totally dismissed (Driver & Noesselt, 2008). Another often accepted neuronal correlate of multisensory processing is the non-linear (i.e. super-additive) neuronal response to cross-modal stimuli, compared to the summed responses of both individual stimuli in isolation (Fort, Delpuech, Pernier, & Giard, 2002b; Giard & Peronnet, 1999; Molholm et al., 2002b; Murray et al., 2005; Talsma & Woldorff, 2005). Accordingly, whenever non-linear responses are observed, multisensory integration can be assumed, while if linear responses are found, then the evidence is not sufficient to support a specific mechanism of multisensory integration. In the later case, a simpler explanation based on the additive effect of unisensory responses is a feasible account for the ERP (and additional behavioral) response.

In literature, the ERP correlates of short-latency non-linear multisensory effects within the first ~40 ms have been repeatedly shown (Fort et al., 2002b; Giard & Peronnet, 1999) and, in some cases (Fort et al., 2002a; Molholm et al., 2002b) have been demonstrated even after ruling out possible anticipatory mechanisms (Teder-Sälejärvi et al., 2002). Precluding anticipation can be essential, since the slow deflection of anticipation, the so-called Contingent Negative Variation, might be considered twice in ERP summation (i.e., for each modality, in case of two senses) as opposed to only once in cross-modal condition, thus the comparison of the sum of unisensory versus cross-modal conditions might incorrectly suggest non-linearity (Murray et al., 2005; Talsma & Woldorff, 2005).

Non-linear responses to multisensory stimuli are mitigated when endogenous attention is not deployed to both of the modalities involved in the cross-modal stimulus (Talsma, Doty, & Woldorff, 2007), suggesting that the early multisensory effect could be already influenced not just by expectation/anticipation but also by endogenous top-down attention. This claim was strengthened by the fact that attention can have an effect on sensory processing even in the first 200 ms (Clark & Hillyard, 1996; Giard, Fort, Mouchetant-Rostaing, & Pernier, 2000; Mangun, 1995; Woldorff et al., 1993). Talsma and Woldorff (2005) specifically measured the role of top-down endogenous spatial attention on ERPs markers of audiovisual integration. Their study showed larger cross-modal responses for attended than unattended stimuli, at latencies of 160ms and beyond. However, please note that in the attended condition of Talsma's study, top-down and bottom-up attentional effects contributed to the audiovisual response. In the last study included in this thesis we capitalized on electrophysiological responses to cross-modal stimuli in order to gain a better understanding of multisensory responses in the absence of top-down attention, as explained in detail below.

### *1.3.3. Multisensory integration for stimuli presented below the level of awareness*

Beyond behavioral and electrophysiological studies in healthy population, numerous studies have measured multisensory integration below the level of awareness, capitalizing on clinical population with altered attention or sensory processing. The audiovisual studies reviewed above (in Chapter 1., Section 1.3.3.), conducted on neglect or/and hemianopic patients provide some examples thereof (op. cit. Frassinetti et al., 2005; Frassinetti, et al., 2002; Làdavas, 2008). Briefly, from these clinical studies we learned that hemianopic or neglect patients are likely to consciously perceive a visual stimulus presented in the blind or neglected hemifield when accompanied with a sound, which would normally be unseen or simply not considered. In the interpretation of these findings, the enhancement of visual awareness by cross-modal simulation has been often attributed to automatic bottom-up multisensory integration, leaving aside the question of whether selective, endogenous attention might have an influential role via top-down mechanisms (for clarification of automaticity, awareness and top-down attention please see *Box 3*).

*Box 3. Automaticity, awareness, and their dissociation from top-down attention*

a. What we mean by automaticity?

The aim of our experiments is to measure automatic, bottom-up multisensory integration, but what does one mean by automaticity? Considering that multisensory interaction is a multifaceted phenomenon, which may happen at different processing stages and via a variety of pathways (e.g., Driver & Spence, 2000; Driver & Noesselt, 2008; Foxe & Schroeder, 2005), it is potentially supported by both bottom-up convergence as well as top-down mechanisms. Yet, many times it is challenging to measure specifically multisensory integration, purely based on bottom-up mechanisms, in the absence of top-down processes. Bottom-up pathways are often assumed to process information in stimulus-driven manner, in a mandatory way without the voluntary control of the observer, often equated to automatic processing.

Notwithstanding, automaticity itself is an elusive term, and its demonstration regarding cognitive processes depends on a variety of criteria, some of which are of a continuous rather than categorical nature (Bargh, 1989; Logan, 1992; Shiffrin & Schneider, 1977). Automaticity in perception is a process that is effortless, not affected by concurrent tasks, and can occur even in the absence of selective attention (Dehaene & Naccache, 2001; Shiffrin & Schneider, 1977). A possibility to address the automaticity of bottom-up multisensory integration is to measure the effects of cross-modal stimulation when it (or one of its components) is not directly accessible to perceptual awareness (Dehaene & Naccache, 2001). If the cross-modal integration effect occurs on task-irrelevant stimuli presented below awareness (where top-down influences are less likely because in principle the multisensory event cannot be the object of selective attention), then one can single out automatic bottom-up multisensory integration processes.

Thus, sensory stimuli can be presented below the level of awareness in some neuropsychological patients (e.g., in hemianopia or neglect), but also in laboratory circumstances (e.g., repetition blindness, attentional blink, masking, binocular rivalry, continuous flash suppression). We used binocular rivalry to present sensory information below awareness.

### *b. Visual awareness*

Awareness is another ill-defined term, not only conceptually, but also in terms of its neuronal correlates. Dehaene, Sergent and Changeux (2003) stated that access to subjective awareness (labeled also as consciousness) is related to the entry of information (e.g., about a sensory stimulus) into a global brain state linking distant areas including the prefrontal cortex, through reciprocal connections, thus making perceptual information reportable by multiple means. Namely, the extension of brain activation to higher association cortices let neural signals be available for a variety of processes, including perceptual categorization, long-term memorization, linguistic processing or intentional action (Mudrik, Faivre, & Koch, 2014).

However, more and more evidences suggest that awareness is a gradual phenomenon, rather than all-or-none. Despite subjective introspection is prone to provide all-or-none information, clinical cases as well as brain imaging techniques reinforced the continuous nature of awareness, characterized by different levels (Bar et al., 2001; Grill-Spector, Kushnir, Hendler, & Malach, 2000; Kouider et al., 2013) activating the neural network to a different extent (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). Importantly, as an agreement between the continuous and all-or-none point of views, some studies also proposed that there is a threshold, which has to be crossed for an experience to become accessible for introspection (Sergent & Dehaene, 2004). However, below this threshold, despite the lack of verbal report, still many perceptual, motor, semantic, emotional and context- dependent processes could occur (Dehaene & Naccache, 2001).

One important aspect for the purposes of the present dissertation, is that encoding of the stimulus remains below the threshold of awareness if bottom-up activation of information is insufficient to trigger a large-scale reverberating of the neural network. A further distinction highly resonates with the gradual concept, stating that awareness can happen at two levels: with perceptual and subjective awareness (Giles, Lau, & Odegaard, 2016). The perceptual awareness constitutes the visual system's ability to process, detect and distinguish amongst stimuli to perform a visual task. While the subjective awareness refers to the visual system's ability to generate a subjective conscious experience. The two phenomena seem conceptually related, although operationally they are distinguishable. Please note that the results of binocular rivalry experiments putatively operate on the perceptual level of awareness (Giles et al., 2016; Zou, He, & Zhang, 2016).

### c. *Dissociation of top-down attention and awareness*

In order to measure automatic processes triggered by unconscious stimuli in the absence of top-down selective attention, the dissociation between attention and awareness might be another useful distinction. On the one hand, an event can benefit from attention without being aware, these types of evidences are provided by binocular rivalry, continuous flash suppression, priming, masking paradigm, repetition blindness or attentional blink (Chen & Yeh, 2008; Faivre, Mudrik, Schwartz, & Koch, 2014; Olivers & Burg, 2008; Vroomen & de Gelder, 2000; Zhou et al., 2010). On the other hand, one could be also aware of an event in the near absence of top-down attention, likewise during zombie behavior (Koch & Tsuchiya, 2006) or gist perception (Li, VanRullen, Koch, & Perona, 2002). Hence, since the two mechanisms dissociate, they can operate independently (Faivre et al., 2014; Koch & Tsuchiya, 2006; Li et al., 2002; Pápai & Soto-Faraco, 2017).

Beyond these clinical studies, several experimental protocols on healthy participants have also targeted multisensory processes for stimuli presented below the level of awareness. These results are important, because when a stimulus is not processes consciously, in principle it cannot trigger voluntary attention orienting. Yet, please note that top-down attention and expectation can still operate, based on other information. Therefore, as we will discuss in this section, one needs to distinguish the different possible contributions to cross-modal responses. In fact, in many studies using unaware stimuli top-down attention may still contribute to the possible expression of multisensory responses, making it difficult to single out bottom-up multisensory convergence.

Vroomen and de Gelder (2000), for example, suggested audio-driven improvements on visual detection in a visual masking paradigm (Bachmann, 1994). The cross-modal enhancement occurred for synchronized sounds but disappeared when the sound was presented slightly before the visual target, leading to the conclusion that the effect was due to multisensory integration rather than to attentional cueing or alerting. A similar finding has been revealed in an attentional blink (Raymond, Shapiro, & Arnell, 1992) study of Olivers and van der Burg (2008), whereby Olivers & van der Burg showed that a visual event escaped from the ‘attentional blink’ more often when it was accompanied by a synchronized auditory event. Additionally, the sound helped to escape the visual event from attentional blink even if it was valid only 18% (i.e., most of the time the tone coincided with visual distractors), again promoting the interpretation of multisensory integration based on automatic, stimulus-driven processes, rather than attention cueing or expectation. What is more, similarly to the result of Vroomen and de Gelder (2000), in Olivers and van der Burg’ study, the effect disappeared if the tone preceded the visual stimulus rather than being simultaneous with it, ruling out alerting and attentional orienting. In line with these findings of masking and attentional blink studies, Chen and Yeh (2008) demonstrated that a simultaneous audio event could also eliminate the visual repetition blindness effect, a phenomenon characterized by failure of perceiving the second occurrence of a repeated item in a rapid serial visual presentation (Kanwisher, 1987). Despite Chen and Yeh's finding was interpreted as multisensory integration, the results were also well within the explanation of attentional capture. This is because the presentation of a sound actually might covertly orient participants' attention to the upcoming visual event (McDonald & Ward, 2000) as well as might simply boost participants' alertness (Robertson, Mattingley, Rorden, & Driver, 1998). While the visual stimulus was rescued from repetition blindness by a sound that appeared before (with 86 ms, thus emphasizing the attentional effect), it did not happen when audio followed the visual event (only with 86 ms falling within the classically considered time window of multisensory integration). Thus, in this case, multisensory integration could not be concluded. From these studies, one can preliminarily conclude that a supraliminal stimulus can help 'rescue' a subliminal stimulus in another modality to awareness in distinct situations, sometimes based on multisensory integration and with the possible contribution of attention.

While these results are promising on their own, a complete subliminal manipulation (i.e. subliminal sound accompanied by subliminal visual, for instance) might be even more powerful by providing a test for automatic multisensory integration in its full breadth. Interestingly, Faivre, Mudrik, Schwarz and Koch (2014) used a congruency priming paradigm to demonstrate that an auditory and a visual stimulus could be integrated even when both of the stimuli were presented below awareness. In this study, pairs of identical or dissimilar audiovisual target letters (i.e., the sound ′b′ with the written letter ′b′ or ′m′, respectively) were preceded by pairs of masked audiovisual prime digits, which could be also identical or dissimilar (i.e., a sound ′6 ′ with a written digit ′6 ′ or ′8 ′, respectively). The relations between the items in the prime and target pairs could either be congruent (items in both pairs identical or dissimilar) or incongruent (items in prime identical while in target dissimilar, or vice versa). Participants had to judge whether the target audiovisual pair was identical or dissimilar. Remarkably, the awareness of audiovisual digit primes was manipulated, such as participants were unaware either of audio or visual digits or both. Congruency priming effect (i.e., congruency related to the similarity of prime and target audiovisual pairs) although revealed even when both of audio and visual prime were presented subliminally, but only if participants underwent prior training of the task using conscious, above threshold stimuli. This suggests that multisensory integration (i.e., based on visual and acoustic features of the audiovisual stimuli) below awareness might be constrained by preceding conscious learning. Yet, there are at least two accounts what cannot be dismissed whenever subliminal multisensory integration is assumed in the study of Faivre et al (2014). First, masking often leads to quite shallow suppression depth (i.e., the suppressed stimulus is not all the time totally unaware); therefore the putative multisensory integration below awareness has to be cautiously interpreted. Second, the constraint of unconscious influence was highly dependent on a learning task a priori, suggesting that previous conscious processing of the cross-modal stimulus is necessary for unconscious cross-modal integration to occur later on. This result in line with the Integrated Information Theory of Consciousness (Mudrik et al., 2014; Tononi, 2004). This theory states that consciousness of some content is necessary for integration below awareness, highlighting the role of consciousness in establishing integrative mechanisms that can later operate unconsciously. However, please note that Noel, Wallace and Blake (2015) highlighted that the cross-modal effect with subliminal stimuli in the study of Faivre et al., (2014) might be also explained by a more parsimonious semantic priming account (i.e., bigger priming effect due to identical audiovisual pairs are based on semantical relations, not necessarily on integration of visual and acoustic features), that is the cross-modal benefit could arise from interactions at amodal, semantic level of analysis (i.e., detection of the auditory and written digit pairs presented at subthreshold intensity and duration), instead of multisensory integration. Thus, unfortunately, the question of whether automatic bottom-up multisensory integration happen remained unanswered.

All in all, an important conclusion arising from the plethora of cross-modal studies suggests that top-down mechanisms many times inevitably affect sensory processing even below or near the level of awareness. Still, paradigms below awareness offer a possibility to rule out or control top-down influences efficiently (i.e., if stimulus presentation not just subliminal but also task-irrelevant) and single out possible automatic bottom-up sensory convergence. This rationale inspired our choice of paradigm and let us decide for binocular rivalry, giving rise to measure bottom-up sensory processes in lack of top-down attention or expectation. In the next and final chapter of the introduction, we briefly describe binocular rivalry and relevant cross-modal studies using this protocol.

## *1.4. Binocular rivalry*

### *1.4.1. Bistable perception, binocular rivalry and, continuous flash suppression*

As it has been mentioned before, in this thesis I have manipulated visual awareness to ensure sensory processing. The presence or content of task-irrelevant sensory stimuli presented below level of awareness cannot ground top-down influences, such as top-down attention or expectation. To achieve sensory stimulation below awareness we capitalized on binocular rivalry. I am going to introduce the technique and its characteristics briefly.

Visual perception is (normally) derived from the two eyes, which receive slightly different views of the scene at any particular fixation (given the different location of the eyes). The neural processes underlying binocular viewing operate to promote single vision by eliminating local interocular conflicts for a goal-directed perception of depth (Levelt, 1965). However, for large interocular conflict, (e.g., for dissimilar images) binocular fusion disrupts (Blake & Boothroyd, 1985), and result in a stochastic alternation between the two distinct percepts entering each eye **(Fig. I/A-B)**. This phenomenon is called 'binocular rivalry' (BR) (Wheatstone, 1838). The image that stays outside consciousness is called the 'suppressed' percept or image, whereas the image that is currently experienced is called 'dominant'. In our experiments, we capitalized on binocular rivalry to render visual images suppressed from awareness for cross-modal stimulation (Blake, 1989; Blake, 2001; Blake & Logothetis, 2002; Levelt, 1965; Wheatstone, 1838).

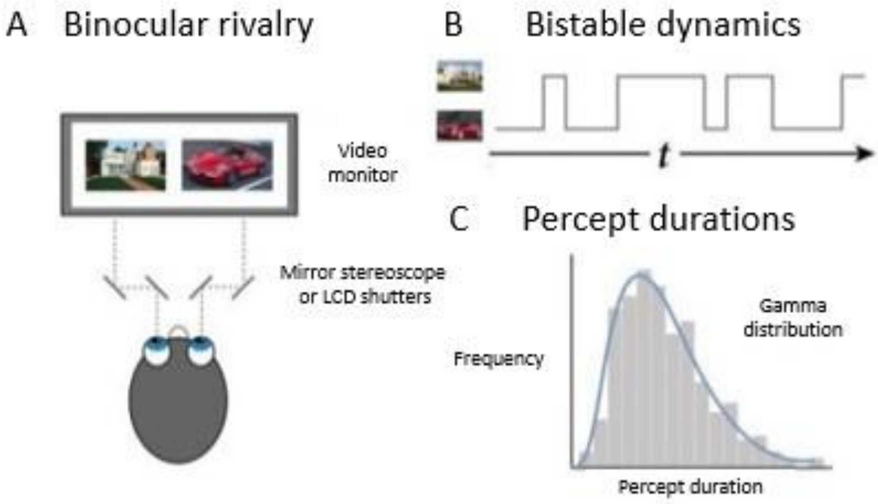


**Figure I/2. A** *Binocular rivalry*. Two dissimilar images are presented into the eyes with mirror stereoscope/LCD shutters. The images do not form a uniform percept rather tend to alternate. **B** *Bistable dynamics*. The time course of the alternations is presented which is stochastic. **C** *Percept durations*. The time of perceptual states form a right-skewed, Gamma distribution. The figure is adapted from Alais and Blake (2013).

BR ensures a good compromise between suppression depth, presentation time, and sensitivity to small modulations (Levelt, 1965) compared to other known techniques like metacontrast masking (Bridgeman, 1971; Francis, 1997), motion induced blindness (Bonneh, Cooperman, & Sagi, 2001) or different forms of flash suppression (McDougall, 1901; Tsuchiya & Koch, 2005; Wilke, Logothetis, Leopold, & Planck, 2003), just to mention some. Suppression depth is characterized by a difference in contrast sensitivity for the suppressed compared to the dominant stimulus, and varies as a function of stimulus strength. Typically, the loss of sensitivity during suppression for balanced rival stimuli is within the range of 0.3 to 0.5 log units relative to their sensitivity under dominance (Nguyen, Freeman, & Wenderoth, 2001). However, please note, that the suppression depth is not a constant, rather a dynamic phenomenon, hence it is weakening over a dominance cycle, reaching a minimum level just prior to a perceptual switch (Alais, Cass, O'Shea, & Blake, 2010).

The latency of the switches between the two competing percepts follows a Gamma-like, right-skewed distribution of dominance durations **(Fig. I/2.C)** (Fox & Herrmann, 1967). When stimuli have equal strength (and importantly there is no big difference in eye-dominance) the otherwise stochastic alternations tend to even out over time. By contrast, when any of the stimuli is stronger than the other (due to differences in contrast, luminance, contour density, spatial frequency, size, motion or velocity), it leads to higher predominance (Levelt, 1965). Thus, the perception of the stronger stimulus remains for an overall increased proportion of the time. An extreme case of this unilateral predominance is the continuous flash suppression (CFS) (Tsuchiya & Koch, 2005), a form of BR that implies a deeper suppression due to the increased predominance of the one of the stimuli.

What kind of neuronal mechanisms stand behind the perceptual fluctuations in BR? Levelt (1965) examined the rivalry dynamics in detail and suggested reciprocal inhibition as the underlying mechanism. The theory of reciprocal inhibition proposes that conflicting visual stimuli receive input that activate their respective, separate neural populations, which send reciprocal inhibition signals to each other. The activation of the winner in this competition, resulting in the dominant percept, is in turn subject to neural adaptation, which gradually attenuates its responses, thus weakening its inhibitory influence over the suppressed population and eventually leading to a turn over (Blake & Logothetis, 2002). These mechanisms offer an explanation to perceptual fluctuation and changes in suppression depth. Researchers have questioned, furthermore, whether reciprocal inhibition and neural adaptation mechanisms are derived from eye-of-origin information. Initially, most of the reciprocal inhibition models assumed that rivalry transpires early in visual processing where inhibitory competition occurs between local features signaled by monocular neurons. This has been confirmed by fMRI study on the blind spot (Tong & Engel, 2001). However, this account did not explain all findings. Kovacs et al. (1996) in their pioneer work broke the homogeneity of conventional stimuli and replaced them by complementary patchworks of rivalry stimuli what resulted in a coherent fluctuation of information originated from different eyes. Reciprocal excitatory connections then seem to promote intra- as well as interocular grouping between neurons with similar preferences, leading to eye- or pattern-based rivalry, favoring a plausible hybrid framework of binocular rivalry (Tong, Meng, & Blake, 2006). However, according to a newer point of view, the influence of external noise (e.g., derived from eye movements or a change in the image) as well as internal noise (e.g., attentional shift, neural noise) on perceptual switches cannot either be dismissed, described by the so-called attractor model (Kim, Grabowecky, & Suzuki, 2006; Lankheet, 2006; Moreno-Bote, Rinzel, & Rubin, 2007). And at last, just to mention few, the most relevant models/theories of binocular rivalry, one of the latest perspective suggests that binocular rivalry does not exclusively depend on bottom-up sensory inputs, rather intertwined with attention. The model of Li et al. (2017) suggest that attention amplifies visual competition by biasing gain toward one of the rival stimuli. After taking into account these basic constraints and mechanisms of BR (without the intention of being exhaustive), next I will review the ERP correlates of stimuli presented under BR.

### *1.4.2. ERP correlates of binocular rivalry*

One approach in ERP studies is to compare physical alternation between stimuli versus BR (Roeber et al., 2008; Veser, O'Shea, Schröger, Trujillo-Barreto, & Roeber, 2008). The moment of a stimulus becomes aware is characterized by a positive ERP deflection around 100 ms (P1) after the onset, followed by a negative deflection around 250ms (N1) and, ending up in a late positive shift around 380ms (P3) (Britz & Pitts, 2011) over ventrolateral occipito-temporal (Roeber et al., 2008; Veser et al., 2008) and prefrontal areas (Veser et al., 2008). In BR studies the P1-N1 components (~100-200ms) have been interpreted to reflect perceptual awareness (Khoe, Mitchell, Reynolds, & Hillyard, 2008; Roeber et al., 2008; Roeber & Schröger, 2004), but they have also been related simply to the saliency of the rivalry visual stimuli (Regan, 1989).

In our experiments, we presented visual contrast increment embedded on the rivalry gratings, which similarly elicits P1-N1-late positivity/P3 components (Mathewson, Gratton, Fabiani, Beck, & Ro, 2009; Metzger et al., 2017), thus we focused on these deflections and timings.

### *1.4.3. Attentional influences in BR*

#### *1.4.3.1. Influence of top-down attention in BR: Unisensory, visual studies*

After a brief review on the basic features of BR, now I will turn the focus on how different attentional mechanisms affect the fluctuation of rivalry visual stimuli. Attentional influence over perception is particularly pronounced when sensory information is ambiguous and, not surprisingly, BR has proved to be quite susceptive to attentional modulation. First of all, BR by itself always requires certain top-down attention as participants are often asked to monitor and report their percepts. But, what happens if top-down attention is directed away from rivalry stimuli? Paffen, Alais and Verstraten (2006) used a peripheral secondary visual task and, Alais et al. (2010) an audio distractor task to direct attention away from rival stimuli and, as a result, rivalry fluctuation slowed down in both cases. The slowing effect was stronger when the secondary visual task was more difficult (Paffen et al., 2006) and, what is more, in the studies of Zhang et al. (2011) and Brascamp and Blake (2012) the alternation even stopped altogether when attention was completely removed from the rival stimuli, line with the modeling data of Li et al. (2017), demonstrating an essential role of endogenous attention in BR dynamics.

Apart from the general role that endogenous attention plays in the monitoring task required in BR protocols, top-down selection to one of the rival visual stimuli prolongs the time the attended image is perceived (dominant) (Chong, Tadin, & Blake, 2005; Hancock & Andrews, 2007; Ooi & He, 1999). The effect of top-down attention is similar to increasing the stimulus strength (i.e., achieved by contrast increment) (Levelt, 1965), suggesting that voluntary attention might enlarge the perceived contrast of the attended stimulus (Carrasco, Fuller, & Ling, 2008; Hancock & Andrews, 2007; Luck, 2004). However, the top-down attention can be deployed to binocular rivalry gratings, without being oriented to a visual flash embedded on any of the binocular rivalry gratings. Chong et al. (2005) as well as Hancock and Andrews (2007) demonstrated that attention deployed to a particular visual feature of any of the binocular rivalry stimulus resulted in lengthened dominance duration of that particular binocular rivalry stimulus while only seeing the binocular rivalry stimulus per se (without focusing on any particular feature) seemed to be insufficient to prolong the duration of a percept. This point is going to be relevant in our studies, as participants were asked to monitor rivalry gratings but no information was given related to visual (either audio) event.

### *1.4.3.2. Influence of bottom-up attention in BR: Unisensory, visual studies*

Perceptual alternation in BR is not only susceptible to top-down selection, but also to bottom-up attentional capture. While, feature-based cues, congruent with the dominant image prolong dominance duration (Dieter, Melnick, & Tadin, 2015), exogenous cues, even if presented under suppression effectively capture bottom-up attention and, often result in a perceptual switch bringing the suppressed image back to awareness (Blake, Westendorf, & Fox, 1990; Dieter et al., 2015; Lack, 1974; Ooi & He, 1999; Paffen & Van der Stigchel, 2010). The presentation of an exogenous cue on the suppressed percept can also produce spatial cuing effect on a subsequent supra-threshold lateralized visual target (Paffen & Van der Stigchel, 2010). What is more, the effect of bottom-up attentional capture, leading to a perceptual switch, is not restricted to stimuli presented within the suppressed image, but it can also occur when a transient event is presented binocularly on the background (e.g., a flash presented behind the rivalry stimuli) (Kanai, Moradi, Shimojo, & Verstraten, 2005).

All in all, evidence regarding to the role of attention in BR are in line with the previously mentioned idea that both top-down and bottom-up attention can increase effective strength of the stimulus, therefore the predominance of the visual stimuli to which the cue is presented (Carrasco et al., 2008; Hancock &

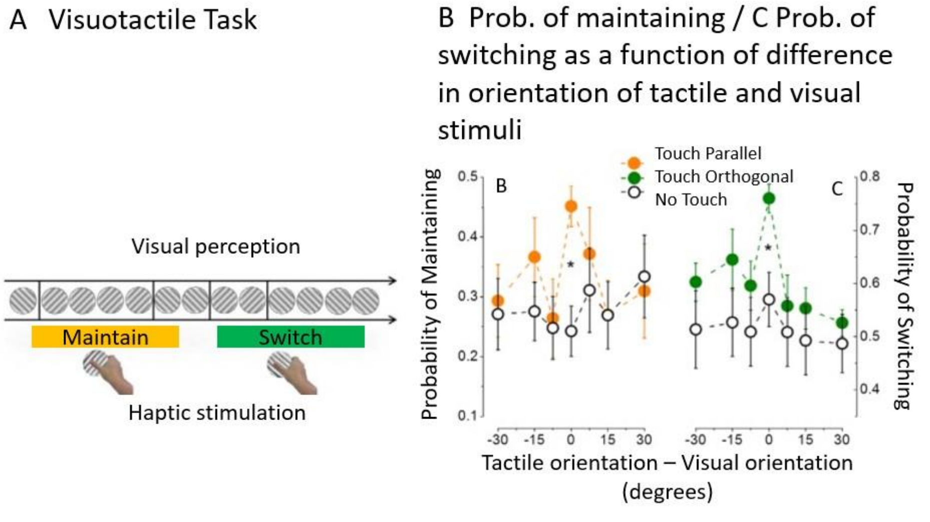
Andrews, 2007; Luck, 2004). The perceptual state tends to prolong as attention is deployed on dominant percept, while a perceptual shift occurs sooner if attention is captured by a salient stimulus either on the suppressed visual stimulus, or just on the background.

### *1.4.4. Cross-modal influences in BR*

In the following sections, I will review BR (and occasionally bistable stimuli) studies with cross-modal manipulations. I have organized the discussion whether top-down mechanisms (endogenous attention, expectation, …) have a clear influence on the cross-modal effect. In general, cross-modal influences are based on some types of feature-congruence between a non-visual stimulus and the visual images under rivalry competition. This congruency across modalities varies from high- to low-level features. Despite both, the high- as well as low-level feature congruency might lead to top-down regulation, confounding possible bottom-up sensory processes (for further clarification please find *Box 4*), still top-down influences are more feasible to rule out or discount if the congruency is based on low-level features alone. We will capitalize on low-level feature congruency between cross-modal stimuli for this reason (explain in detail below).

For example, some studies have reported that visual stimuli can be 'rescued' to awareness from BR suppression based on cross-modal modulation relying on high-level features. This modulatory effects hinge on a variety of semantic or other high-level relationships, such as the meaning of words (Dehaene, Kerszberg, & Changeux, 1998), semantic incongruences in written sentences (Sklar et al., 2012), visual scenes (Mudrik, Breska, Lamy, & Deouell, 2011), arithmetic operations (Bahrami et al., 2010), faces (Barbot & Kouider, 2012) or even emotion (Faivre, Berthet, & Kouider, 2012). This type of congruency can easily trigger expectation, imagery or top-down endogenous selection capitalized on the non-visual stimulus, meaning or other prior-based correlations, that exerts a modulation of the BR dynamics following any these well-known principles (explained above).

Besides the cross-modal effects capitalizing on high-level feature congruency, visual stimuli have been also reported to be 'rescued' from suppression with low-level feature congruency across modalities (a representative example is presented on **Fig. I/3.**). In these cases, the modulation effect could be derived from congruency in the orientation of the stimuli (Lunghi, Binda, & Morrone, 2010), in spatial frequency (Lunghi & Morrone, 2013), in amplitude modulation (Lunghi et al., 2014), or in the combination of the latter two (Guzman-Martinez, Ortega, Grabowecky, & Mossbridge, 2012). Thus, even though top-down regulation can obviously happen for cross-modal stimuli with low-level features, its control is more feasible. For example, in the absence of high-level feature congruency across-modalities, bottom-up multisensory integration might be assumed, when cross-modal effects occur for simple spatio-temporal coincidence despite the two modalities are unpredictable and mutually uninformative about time, space or features (i.e., there can be no cueing effect or expectation built-up). Another case that helps support cross-modal interactions in the absence of top-down regulation is when information in both sensory modalities (the visual event, and the non-visual event) are presented below their respective unimodal thresholds for awareness, but cross-modal effects are seen.


A Visuotactile Task

B Prob. of maintaining / C Prob. of switching as a function of difference in orientation of tactile and visual stimuli

**Figure I/3.** A representative example of a rescue of suppressed stimulus to awareness based on low-level congruency (i.e., in orientation) of tactile and visual features. ***A*** *Visuotactile task.* Participants had to report the orientation of rivalry visual gratings meanwhile congruent or incongruent tactile stimulus (i.e., parallel or orthogonal in orientation, respectively) was monitored.

**B-C.** *Probability of maintaining / probability of switching as a function of difference in orientation of tactile and visual stimuli.* (B) A congruent (parallel) tactile stimulus increased the probability of maintaining the dominance time, while (C) an incongruent (orthogonal) tactile stimulus increased the probability of switch (i.e. decreased the suppression time) of the visual grating within a fine-tuned orientation difference. Adapted by Lunghi and Alais (2013).

*Box 4. Cross-modal congruency based on high- and low-level features*

On the one hand, the cross-modal congruency based on stimuli with high-level features (e.g., speech contains high-level information such as semantics) inevitably relies on not just bottom-up mechanisms deriving from the sensory attributes of the stimuli, but also from top-down processes due to prior knowledge such as statistical regularities, semantic associations and, in language, phonological and syntactic rules. Thus, even in task-irrelevant situations it is possible that the top-down mechanisms become involved in perception, thus information from one modality can perfectly help to anticipate corresponding sensory information from another modality, based on, for example, semantic associations. On the other hand, the cross-modal stimuli can be also congruent in terms of low-level features (e.g., orientation, amplitude modulation, spatial frequency etc.) in the absence of further semantic associations, and hence prone to test bottom-up effects in isolation. However, the top-down regulation cannot be always ruled out even in these cases since stimuli based on low-level features might stay under the influence of cross-modal correspondences deriving from statistical knowledge about world properties (about how features in one modality correlate to certain features in another modality, such as spatial frequency in vision and roughness in touch, visual size and acoustic frequency) (Parise & Spence, 2013) and hence, in the end, might call expectation.

In the following sections, I will summarize some crucial findings illustrating cross-modal congruency effects below awareness, based on high- or on low-level sensory features. In each case, the potential influence of top-down mechanisms will be discussed, giving rise to a clear conclusion on whether a given effect can or cannot be partially attributed to top-down processes. I will discuss cases in which cross-modal effects did not happen for below-awareness stimuli and cases in which they did.

### *1.4.4.1. Failures to find multisensory interactions below awareness using cross-modal congruency between high-level features*

Some studies using BR or related protocols, have reported cross-modal effects that occur only for the dominant visual stimulus, but not when the visual stimulus is under suppression. These cases are referred to as 'failures' to find multisensory interactions below awareness, because in fact, they only occur above awareness. Munhall et al. (2009) capitalized on bistable visual images to study multisensory integration below awareness (even if the suppression depth is quite weak this case) based dynamic version of Rubin vase-face bistable stimulus accompanied by auditory stimuli. An irregular vase was presented in rotating motion, which could be perceived as a moving vase or a face (i.e., moving lips). The bistable stimulus was accompanied by speech sound promoting the McGurk illusion (McGurk & MacDonald, 1976), in combination with the lip movements of the face percept. The cross-modal illusion was limited to face-perception periods of the bi-stable stimulus, thus the multisensory effect worked only above the level of awareness. In another study, Palmer and

Ramsey (2012) found that the McGurk illusion vanished when the sight of lip movements was under flash suppression (using Continuous Flash Suppression, or CFS), arriving to a similar conclusion as that of Munhall et al. (2009). Thus, from these studies we learn that cross-modal integration, at least for the McGurk effect, is very limited or absent when the visual component is not perceived consciously.

Chen, Yeh and Spence (2010) measured BR as a function of cross-modal perceptual congruency between the rival images (of a bird and a car) and sounds (e.g., bird song, car engine, or neutral soundtrack recorded in a restaurant). Congruent sounds, that is, sounds that belonged to objects from the same semantic category as the dominant image, resulted in lengthening of the duration of the corresponding visual percept, compared to any other conditions. Additionally, in a follow-up experiment of the same study the effect of semantic congruency was tested by replacing the bird or car sounds with the corresponding written words. In this case, the cross-modal modulation was not found, thus the cross-modal effect with sounds could not be attributed to congruency between semantic categories (for further information please find *Box 5*). Instead, the effect was attributed to cross-modal perceptual congruency (as authors named it), which was based on visual and acoustic stimulus properties that were associated across modalities. Still, since the cross-modal effect was limited to a lengthening of the dominant percept, but ineffective for images under suppression, we can still include this result in the category of cross-modal effects occurring only above awareness. As such, multisensory effects remained susceptible to top-down mechanisms like attentional selection based on the consciously perceived, sound category.

*Box 5. Perceptual congruency*

The ‘perceptual congruency‘ operates on bottom-up sensory processes of congruent cross-modal features that can be disentangled from the conceptual effect of semantics. We borrowed the expression from (Chen, Yeh, & Spence, 2011), although other authors refer to a similar concept with the terms ‘sensory congruency’ (Zhou et al., 2010) or ‘object-based congruency’ (Iordanescu, Grabowecky, Franconeri, Theeuwes, & Suzuki, 2010). The same concept has been also raised related to a congruency priming effect in the study of Faivre et al. (2014), and reinforced in the criticism of (Noel et al., 2015) (explained in detail below in the main text). The basic notion is that there are certain non-semantic properties that can relate events in different modalities. For example, Pesquita et al. (2013) raised this question in a cross-modal priming effect where the haptic properties of an object primed the recognition of a picture of the same exact object more strongly, than a picture of a different token of the same object. This semantic category was the same in both cases, but the exact object condition provided further cross-modal perceptual congruency.

#### *1.4.4.2. Successful multisensory interactions below awareness using cross-modal congruency between high-level features*

Beyond the above-mentioned findings where cross-modal effects were reported only when both stimuli were aware to the observer, other studies using congruency between high-level features provide evidence of cross-modal effects below awareness. For example, the previously reported study of Palmer and Ramsey (op. cit. 2012) reported a different experiment showing that spatial attention orienting towards the location of suppressed (i.e., hence invisible) lip-movements that matched above threshold speech sounds. Therefore, the conclusion was that unconscious visual speech putatively integrated with the (consciously perceived) speech sounds. Please note that, the difference between the two manipulations (the one mentioned in the previous section and this one) was whether subjects were aware of the influencing modality or not. Based on this difference, Palmer and Ramsey suggested that aware perception of the influencing modality is necessary for cross-modal integration below awareness to happen.

In a different study, Alsius and Munhall (2013) found a similar influence of consciously presented speech sounds on suppressed, congruent moving lips (visual speech), in a CFS paradigm. They used CFS to render visual stimuli of a talking face unaware to participants, whilst supra-threshold audio speech corresponding to the unseen lips was presented. The speaking lips were gradually released from suppression by increasing their contrast over time. In the study of Alsius and Munhall, the image of the talking visual face took less time to overcome suppression when accompanied by congruent speech compared to incongruent speech sounds. The authors interpreted this advantage as a cross-modal benefit to visual events deriving from integration at early stages of sensory processing. However, please note that subjects were again aware of the supra-threshold, influencing auditory stimulus, thus some top-down attentional regulation was highly probable. In a another CFS study (Plass, Guzman-Martinez, Ortega, Grabowecky, & Suzuki, 2014) a talking face below awareness could be congruent or incongruent with a target spoken word in a word categorization task. In the congruent condition, cross-modal facilitation occurred even though subjects were not aware of the articulating face, giving rise to an interpretation of automatic multisensory integration. Yet, please note that in this study subjects had to report the visibility of the face, calling for some sort of top-down selective attention to the (suppressed) visual stimulus.

In the studies discussed above, the nature of the stimuli involved congruency relationships at various levels of complexity (from visual and acoustic features to semantics) and, they were available for prolonged periods of time during the task execution. These characteristics putatively facilitated the influence of top-down processes. What is more, the fact that conscious perception of the influencing stimulus seems to be a requisite for multisensory integration to happen for below awareness visual information, is also consistent with the idea that endogenous selective attention might be a source some of the cross-modal effects reported. In particular knowledge about one stimulus can induce a pattern search for corresponding (congruent) visual features, in very much the same way that endogenous attention to certain features facilitates search, based on instructions. In the cross-modal studies discussed so far, where congruency was based totally or partially in high-level features, the top-down influences could not be dismissed and hence, the conclusion of bottom-up multisensory integration cannot be granted.

However, another study Zhou et al. (2010) found a modulation between based on the sensory congruency (as the authors labelled it, see *Box 5*) between smell and vision presented in a BR and CFS protocols. The categorically congruent smell (e.g., rose) 'rescued' congruent visual images (e.g., picture of a rose) to awareness, compared to an incongruent odor. This congruency effect could not be explained by semantics alone, since verbal information about congruent or incongruent odors alone did not seem to affect visual stimuli like the real fragrance. According to Zhou et al., the effect rather relied on the congruency between stimuli based on bottom-up sensory integration. Similarly, Zhou et al., (2012) in another study demonstrated that smelling an odor from one nostril significantly enhanced the dominance time of the congruent visual image in the contralateral, relative to the ipsilateral visual field. This modulation followed the rules of anatomical organization and, unlikely to be due to top-down influences.

#### *1.4.4.3. Failures to find multisensory interactions below awareness using cross-modal congruency between low-level features*

The majority of cross-modal effects (based on congruency between high-level features) below awareness discussed so far putatively involved some top-down processes (except Zhou et al., 2012). However, cross-modal stimuli sharing only congruent low-level features might offer a more clear-cut opportunity to target bottom-up effects alone in the cases when stimuli are uninformative (i.e., knowledge based on cross-modal correspondence, thus expectation cannot be efficiently activated). In this following section I will discuss studies capitalizing on low-level feature congruency.

Kang and Blake (2005) used an amplitude-modulated sound which could be correlated with flickering gratings in the suppressed or in the dominant eye. Auditory congruency with the dominant visual percept prolonged dominance, although cross-modal congruency did not affect the switch latency of suppressed visual events (i.e., presented below awareness). In another study, Van Ee et al. (2009) found a similar result: again, cross-modal benefits were restricted to the dominant visual percept. Additionally, in that study top-down attention had a special role, the cross-modal effect appeared exclusively when participants actively monitored one of the rival images, but not in passive viewing conditions. Hence, these findings can be framed within known top-down attentional effects in unimodal, visual conditions discussed above (in Section 1.4.3.1., Chong et al., 2005; Hancock & Andrews, 2007), only that the incentive for top-down selection comes from information variable in a non-visual sensory modality.

### *1.4.4.4. Successful multisensory interactions below awareness using cross-modal congruency between low-level features*

Thus, we have been seeing several studies where the cross-modal effects were constrained to awareness (i.e., dominant situation), thus the stimuli again might be susceptible to top-down influences. Hence, without negating the possibility of some bottom-up based findings, these outcomes cannot be conclusive to the effect given the availability of other, more parsimonious accounts. In this section, I will briefly review studies were cross-modal effects, deriving from congruent low-level features, were found below the level of awareness. However, please note that whenever in one modality stimuli are informative about the other modality (i.e. task-relevant or involved expectation based on cross-modal correspondences) top-down influences might still play a role and thus, conclusions about bottom-up multisensory integration uncertain; I will discuss such cases, respectively.

Guzman-Martinez et al. (2012) demonstrated that an amplitude modulated sound shortened the suppression duration (as well as prolonged the dominance duration) of the congruent visual image in a BR protocol. This happened despite participants were asked to ignore audio events; furthermore, none of the participants reported awareness of any relationship between audio and visual stimuli in post-experiment briefings. Thus, one might attribute this effect to multisensory integration taking place below awareness independently of top-down regulation. However, Orchard-Mills et al. (2013) in a subsequent experiment adopted identical stimuli in a visual search paradigm (above awareness, not in a BR protocol) and found that, while the informative audio signals improved search efficiency for the visual stimuli, when audio signal was not informative it failed to affect visual search. The Orchard-Mills' finding raises then the question of whether automatic multisensory integration (i.e., without top-down attentional influence) occurred in the previous study of the research group with completely identical stimuli (Guzman-Martinez et al., 2012).

In another study, Conrad et al. (2010) found that sounds with congruent motion direction prolonged the dominance periods and, importantly, shortened the suppressed periods of random-dot kinematograms presented in a BR protocol. However, as argued by Conrad et al., sounds that were congruent with the suppressed visual stimulus were, by definition, incongruent with the visual motion shown in the dominant percept, and hence could have had an interference effect with the conscious percept, rather than a beneficial effect on the unconscious stimulus. Interestingly, there was a further, non-motion sound condition, which had interference with dominant visual stimulus (with directional motion) but failed to have congruency with the suppressed visual stimulus (characterized by directional motion as well). The perceptual switch of non-motion sound conditions did not differ from no sound condition, thus the lack of cross-modal effect does not favor interference explanation.

Hence, despite the experiments discussed so far in this section provide some initial evidence for cross-modal effects putatively affecting unaware percepts, the cross-modal effects could not surely be attributed to bottom-up integration. Still some other studies, provide more clear-cut findings for bottom-up multisensory integration based on cross-modal congruency. One of them, is the study of Salomon et al. (2013), where participants reported the orientation of a target visual stimulus embedded in a task-irrelevant picture of a hand under suppression (using CFS). The position of the hand was, in principle, irrelevant to the task (orientation discrimination) but when the hand was congruent with the participants' actual hand position, the actual target broke through the CFS more rapidly than in hand position-incongruent trials. Because participants were not aware of the influencing stimulus (i.e., their hand position), which was itself task irrelevant, it is unlikely that multisensory integration might have been influenced by top-down attention or response bias. Hence, visual stimuli were putatively rescued to awareness due to bottom-up multisensory integration. Additionally, in another study Salomon (2015) showed that beyond proprioception, even the vestibular signals integrated with visual stimuli below level of awareness. During yaw rotation interocular suppression was broken more rapidly with congruent compared to incongruent vestibular stimulation. These results therefore suggested multisensory integration below the level of awareness deriving from cross-modal low-level perceptual congruency and, happened putatively independently of top-down attention.

In a different set of studies, Lunghi, Binda and Morrone (2010) reported experiments demonstrating that tactile stimuli can influence suppressed visual events based on congruency in grating orientation. Moreover, the effect showed a narrow selectivity to spatial frequency and to spatial overlap (Lunghi & Morrone, 2013). Even though the influence of higher-order processes were not controlled for and could not be completely dismissed in these particular studies, a subsequent experiment (Lunghi & Alais, 2013) (based on a very similar stimulus set-up) demonstrated that tactile stimuli influenced suppressed visual gratings with orientation selectivity so fine-tuned manner that it was undetectable by participants even when asked explicitly, in a subsequent, supra-threshold visual task. Additionally, Lunghi, Morrone and Alais (2014) found that congruent amplitude modulation of audiotactile stimuli had an effect on unaware visual stimuli, whilst neither the auditory or tactile stimuli alone were effective. Hence, the effect could not be accounted for response bias or top-down attention, and the bottom-up multisensory integration account was favored. Supporting this conclusion, an additional BR study by Lunghi and Alais (2015) demonstrated that the tactile influence on visual processing was indeed, independent of response bias. In the same line, Lunghi, Verde and Alais (2017) furthermore presented a cross-modal effects on suppressed visual stimuli not just by using BR, but also by CFS. Lunghi et al. suggested that the underlying mechanisms of these effects might be that cross-modal influence on tactile or occasionally audiotactile stimulation putatively prevented the congruent visual stimuli from becoming deeply suppressed from awareness (Lunghi & Alais, 2015; Lunghi et al., 2017).

For the scope of the thesis the most important results are then the evidences by Lunghi et al. (2013; 2014), Salomon et al. (2013) and, Zhou et al. (2010, 2012) between cross-modally congruent stimuli. In all these cases, cross-modal stimuli were either mutually uninformative and/or worked despite lack of awareness of the influencing modality. Hence, these results suggest bottom-up multisensory integration. Albeit, even if they used very different stimulation modalities, one remarkable similarity between the studies of Lunghi et al., Salomon et al., and Zhou et al. is that stimulus presentation sustained for a relatively long period of time and, was based on some feature congruency between cross-modal stimulus pairs. The relatively long stimulus presentation might give rise to build up expectation between the stimuli (i.e., attributing some task-relevance to it) thus, facilitating multisensory integration at a certain extent.

Thus, one might fairly raise the question of whether clear evidence of bottom-up multisensory integration could be given by transient uninformative and unpredictive stimulation in a BR task, where low-level cross-modal congruency is confined only to occasional spatio-temporal coincidence of the stimuli. This thesis presents four experiments organized in two studies in order to exploit these possible cross-modal bonds. While the multisensory mechanism near or below the level of awareness stayed mainly at our focus, certain influence of attention was inevitable. However please note that the task-related top-down attention needed to monitor the perceptual dynamics in the BR protocol was not placed on the audiovisual transients of interest, rather to the rivalry stimuli which were irrelevant to the task.

# CHAPTER 2.

## Experimental studies

### 2.1. *Hypotheses*

We have learned that multisensory integration can eventually operate below awareness, however so far it was not clear whether this kind of cross-modal interaction can happen independently from higher-order influences, such as top-down attention or expectation. Although there are several very compelling examples of cross-modal interaction below awareness, the initial issue of whether purely bottom-up pathways can support these kinds of interactions are still under question. Specifically, a case would have to be made about abrupt, quick and unpredictable events.

We based our experiments on a BR protocol, which provides a possibility to monitor perceptual awareness in a controlled way, without necessarily changing the physical stimulation. Our goal was to revel automatic multisensory integration effects, if any. We controlled for task-related top-down modulation as much as possible and aimed to capitalize purely on bottom-up multisensory processes based on low level feature congruency. In particular, in the experiments presented in this thesis we used abrupt, brief, and non-semantic visual and an audio stimulus that were mutually unpredictable, task-irrelevant, and did not share any feature congruency apart of pure temporal coincidence. We focused on modulation effects on the suppressed visual percept, as this was the critical condition to demonstrate cross-modal integration below awareness. The experimental logic is based on the known fact that a sub-threshold visual onset occurring in the suppressed percept speeds up the switch, producing an early emergence of that suppressed percept back to awareness. We build up on this to formulate a multisensory hypothesis. In particular, we hypotized that if multisensory integration happens for unaware visual events, in a bottom-up fashion, then we would have to observe faster perceptual switches from suppression to dominance when an audio stimulus is presented together with a visual event in the suppressed percept, compared to the occurrence of the visual event alone or the auditory event alone.

By contrast, if multisensory interactions depend on higher-order mechanisms capitalizing on endogenous selective attention or expectation, then perceptual switch in our experiment will not occur faster in cross-modal condition compared to either unimodal cases, because our paradigm prevents anticipation or the selective effect of attention. Furthermore, in our paradigm, we will be able to test for cross-modal enhancement effects that reflect co-activation, rather than simple redundant effects, both at the level of behavior and of neural correlates. The presence of such co-activation is expected to lead to non-linear responses, and will be taken as conclusive evidence for bottom-up multisensory integration.

## 2.2. *Study 1*

Study 1 has been published in Scientific Reports (2017).

Pápai, M. S., & Soto-Faraco, S. (2017). Sounds can boost the awareness of visual events through attention without cross-modal integration. *Scientific Reports*, 7(January), 41684. http://doi.org/10.1038/srep41684

Pápai MS, Soto-Faraco S. Sounds can boost the awareness of visual events through attention without cross-modal integration. Sci Rep. 2017 Mar 31;7(1):41684. DOI: 10.1038/srep41684

## 2.3. *Study 2*

Study 2 has been submitted for publication in Scientific Reports (2017).

# Behavioral and electrophysiological correlates of the cross-modal enhancement for unaware visual events

*Márta Szabina Pápai[1], Mireia Torralba[1] & Salvador Soto-Faraco[1,2]*

[1]Multisensory Research Group, Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain.
[2]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
Correspondence and requests for materials should be addressed to M.S.P.
(email: marta.papai@upf.edu)

## *2.3.1. Abstract*

According to many reports, cross-modal interactions can lead to enhancement of visual perception, even when visual events appear below awareness. Yet, the mechanism underlying this cross-modal enhancement is still unclear. The present study addressed whether cross-modal integration based on bottom-up processing can break through the threshold of awareness. We used a binocular rivalry protocol, and measured ERP responses and perceptual switches time locked to flashes, sounds or flash-sound co-occurrences. In behavior, perceptual switches happened the earliest when subthreshold flashes co-occurred with sounds. Yet, this cross-modal facilitation never surpassed probability summation, thus suggesting independence rather than

integration of sensory signals. Likewise, the ERPs to audiovisual events did not differ from the summed unimodal ERPs, also suggesting that the cross-modal behavioral benefit for unaware visual events can be explained by the independent contribution of unisensory signals and suggest no need for a multisensory integration mechanism. Hence, even though cross-modal benefits can appear behaviorally, we suggest that this cross-modal facilitation might origin from well-known bottom-up attentional capture processes, contributed by each individual sensory stimulus.

### *2.3.2. Introduction*

Our senses are constantly bombarded by inputs coming from different sensory modalities. Inferring which inputs belong together and which ones not is a non-trivial perceptual problem. It is often claimed [1–3] that this inference is guided by assumptions grounded on spatio-temporal correlations between events in different modalities, on low-level feature-based congruency as well as on higher-level aspects based on meaning or learned associations. This cross-modal binding leads to multisensory integration (MSI) effects, which are claimed to carry beneficial consequences for information processing [3]. However, it is still controversial to what extent MSI can proceed in a purely bottom-up fashion, without any top-down regulation [4,5]. This question is relevant because it bears on automaticity and, hence, on the limitation of processing resources in the brain required for MSI [4,5]. Accordingly, genuine bottom-up MSI would show that multisensory interactions can take place under extremely impoverished sensory conditions. Indeed, such principle has been put forward as a means for rehabilitation in neurological patients suffering from visual hemineglect or hemianopia. Hence it is important to gain an understanding of the underlying mechanisms leading to such visual enhancement.

Supporting a bottom-up view, a good number of recent studies have claimed that MSI can happen for events presented below awareness [6–11]. Many of these studies have used binocular rivalry (BR), whereby a non-visual stimulus exerts an influence on a visual stimulus that is suppressed from awareness. However, the potential role of top-down mechanisms of attention or expectation related to time, space, or to other information provided by the non-visual stimulus (shape, orientation, semantics) are sometimes difficult to rule out [12,13]. Nevertheless, a few studies have revealed MSI effects for stimuli below awareness, supposedly without top-down regulation [6,9,10]. Interestingly, in these studies cross-modal stimuli were typically presented for extended periods of time, promoting the build-up, or recall of existing associations between stimuli pairs that perhaps facilitated the cross-modal interactions below awareness [14]. Therefore, the question remains as to whether MSI below awareness can happen when cross-modal stimuli do not share any feature-based, learned or semantic congruency, above and beyond mere spatio-temporal coincidence.

Recently, Pápai and Soto-Faraco [15] addressed this question capitalizing on a BR protocol where cross-modal events resulted from the spurious co-occurrence of sounds and unaware flashes (on the suppressed eye) presented at unpredictable moments in time, and mutually uninformative. Despite perceptual BR switches happened earlier after cross-modal stimuli compared to unimodal (visual or auditory) events, suggesting an interaction between visual and sound signals [16], this benefit did not exceed statistical facilitation predicted by the probability summation model (PSM) [17,18]. The most parsimonious explanation then, in line with the currently dominant models, was that the cross-modal benefit arose from the independent contribution of events in each modality, putatively involving bottom-up attentional capture (considering that even faint stimuli calls for bottom-up attention to a certain extent [19–21]). Thus, MSI between unaware visual events and sounds could not be concluded from these behavioral results. Yet, failure to violate the benchmark of the PSM does not negate MSI, by principle [22], and therefore the initial question still remains unresolved.

In fact, stimulus-driven effects in attention and MSI are often difficult to separate and deeply intertwined [23–25]. Considering that the distinction between attention and MSI behaviorally is quite challenging (i.e., since multisensory stimuli can integrate over quite wide spatial disparities [26–29] and within a flexible time window [30,31]), electrophysiological techniques might help provide evidence. Here, we gauge the possible neural correlates of bottom-up MSI effects using event-related potentials (ERPs) in a BR task akin to the one in Pápai & Soto-Faraco [15]. To the best of our knowledge, no previous studies have addressed the ERP correlates of audiovisual integration under binocular suppression of the visual event. Nevertheless, it is worth revising previous ERP findings regarding audiovisual integration for supra-threshold stimuli, as well as the correlates of unisensory visual events under suppression.

The neural correlates of MSI for supra-threshold stimuli are well established in the literature, and include expressions in sensory specific [32] as well as in multisensory brain regions (e.g., superior temporal sulcus, right insula, right prefrontal region) [33]. Prior studies have reported that ERP responses to multisensory stimuli (modulation of visual evoked potentials by sounds or *vice versa*) often reveal nonlinearities, beyond the summed responses from the constituent unisensory stimuli [2,32,34,35]. These kinds of non-linearities (i.e., super-additive interactions) are interesting because they reveal MSI processes above and beyond additive effects. In particular, the auditory modulation on visual processing, of interest here, has been found to peak at occipital and parieto-occipital locations [32,34–37]. What is more, some of these cross-modal effects occur very early after stimulus onset, indicating interactions within the first 200 ms [32,34,38], although the contribution of anticipatory processes cannot be ruled out [2,36]. Relevant for the purposes of the present study, Talsma and Woldorff [2] tested for ERP correlates of cross-modal integration for simple audiovisual events presented at, or away from, the focus of spatial attention. Remarkably, the results revealed that the ERP expression of MSI was attenuated for unattended events, compared to audiovisual events at attended locations. Along similar lines, another study from the same group showed that super-additive responses to cross-modal events diminish in absence of modality related top-down attention [39]. In both cases, the attentional effect on MSI appeared at early time windows, in line with attentional effects on unisensory processing, within the first 200 ms [40,41]. Therefore, considering possible top-down effects at early time window (anticipation as well as attention) one might wonder to what extent were previously reported multisensory effects occurred independently of higher-order influences [5,12]. As far as MSI happens to occur above level of awareness it keeps being quite amenable to the modulation by higher-order mechanism. However, because top-down influences can be more controlled (occasionally ruled out) below awareness, BR is a promising technique to single out bottom-up effects.

Previous findings about ERP correlates to visual events below awareness deserve some discussion here. Usually the P1, N1 and 'late positivity' (~ P3) are the visual evoked potentials which are modulated as a function of awareness [42–52]. Generally, the finding about P1 and N1 is that their amplitude tends to diminish without awareness. Despite P1 effects have been suggested to simply reflect subjective visibility of the stimuli [53], there is more consensus about N1 visual potential, although the direction of the modulation by awareness remained controversial (reduced [49–52] or increased [47,48] N1 amplitude under suppression). Similarly, the late positivity component of the visual evoked potential is often modulated by awareness [45,46], but the interpretation of this modulation is less agreed upon [50,54].

In the current study, we used transient visual flash probes at threshold luminance embedded in rival gratings in a BR protocol. We decided for faint flashes to produce moderate stimulus-based perceptual switch, and leave some room for auditory modulation. Otherwise, the flashes occurred at unpredictable moments in time, and where intermixed with an unrelated, and also unpredictable sound. The two events would occasionally coincide. Both sensory events were task-irrelevant, thus mutually uninformative as well as unpredictable (in this way the anticipatory effects on the ERP were mitigated). Importantly, while top-down control is invariably at play in any task, including BR protocols, the present manipulation minimizes stimulus-selective top-down effects of attention or expectancy based on the temporal, spatial, feature-based or semantic congruency of the stimuli, within or across modalities. In this manner, we hoped to single out the effects of bottom-up signal processing, and their integration, if any. Since Pápai and Soto-Faraco [15] using very similar conditions found that behavioural correlates provided evidence for capture from each single modality but did not provide convincing proof of MSI, we expected to exploit the same logic here, and seek whether the neural responses (ERPs) to the stimuli allow concluding on MSI, given that they produce measurable unisensory responses. As mentioned earlier, the violation of the PSM in behaviour can be an indicator of the presence of MSI, although the lack of evidence for PSM violation does not rule MSI out [15,22,37]. Instead, MSI might be present, but express behaviourally below the threshold of PSM. Our approach here is to exploit such expression by using measurements of neural activity.

If MSI occurs for visual events below awareness, then we expect an amplitude modulation of the early visual evoked responses when combined with sounds, compared to the ERP responses to visual events alone, with special focus on parieto-occipital areas. We have chosen parieto-occipital areas to our region of interest, not just capitalized on the location of auditory modulation on visual processing [32,34,36] but also on the location of the first stages of visual consciousness [44,53,55]. In order to produce the correct baseline, the ERPs to the visual alone and auditory alone stimulus conditions will be summed (hereafter referred to as the A+V ERP) and compared to the ERP response to the actual multisensory stimulus (AV ERP). Based on the principle of superposition of electrical fields, if we assume no MSI processes, then the AV ERP would be equivalent to the sum of the individual components, that is the A+V ERP. However, if the ERP to the multisensory event (AV ERP) deviates from the sum of individual ERPs (A+V ERP), then one should infer some integration process, in line with the co-activation model. However, in the contrast of multisensory versus sum of unisensory responses (AV-(A+V)) there would be double amount of baseline activity in the unisensory sum than in the multisensory response. In order to address this issue, activity without event presentation was taken into account, and subtracted from the unisensory sum [2] .

Furthermore, as we will measure behavioural responses (probability of perceptual switch in the BR task), likewise the pervious study of Pápai and Soto-Faraco [15], we will be able to measure the probability of switch time-locked to the sensory events, and calculate whether suppressed audiovisual stimuli reveal the expected cross-modal facilitation in behavior.

## *2.3.3. Results*

### *2.3.3.1. Behavioural results*

#### *2.3.3.1.A. Time-Probability analysis*

The probability of perceptual switches has been measured as a function of time, locked to the events, and the moment of switches (Mean Time to Switch, MTS) has been indicated by the time when probability dropped down to 50%. The ANOVA on the latency data (MTS), returned a significant interaction between perceptual dominance (dominant, suppressed) and event modality (audio, visual and audiovisual) F(2,16)=3.891, p=0.042, which granted for further analysis. Supporting the presence of cross-modal effects, audiovisual events presented under suppression induced faster

MTS (earlier emergence of the suppressed percept to awareness), compared to visual alone and audio alone events, one-tailed (left), paired t(8)=-2.775, p=0.012, and t(8)=-2.531, p=0.017 (alpha level after Bonferroni correction = 0.025). For completeness, we ran an additional two-tailed, paired, t-test (since we did not have a priori hypothesis concerning the direction of the difference) between visual and audio suppressed conditions, which were not different p=0.989. We ran a further two-tailed, paired t-tests between modalities under dominance, what failed to result in statistically significant effects (all p>0.05; **Fig. III/1.A-B**).

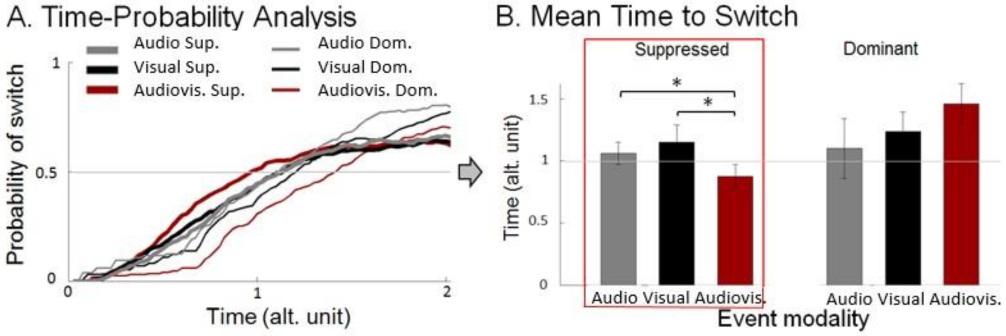


**Figure III/1.A** *Time-Probability Analysis.* The probability of switch is plotted as a function of relative time, expressed in alternation units (sampling points of 0.025 alt. unit). Switch in suppressed condition refers to change in perceptual state from seeing the radial checkerboard to seeing the Gabor patch (as events were presented on/during Gabor patch under suppression; audio, visual and audiovisual suppressed conditions, indicated by grey, black and red colors, respectively). On the other hand, switch in dominant conditions indicates change in perceptual state from seeing the Gabor to seeing the radial checkerboard (as events were presented on Gabor patch under dominance; audio, visual and audiovisual dominant). Zero time point indicates event onset time, which never occurred during piecemeal percept. **B** *Mean Time to Switch (time at which probability crosses 50%).* The bars indicate the mean relative time of switch (in alternation unit with SEM) in each condition separately. Additionally, please note that significant comparisons marked by '*'.

#### *2.3.3.1.B. Probability summation*

Additionally, we tested the empirical behavioral data from the audiovisual suppressed condition against the PSM calculated from the unimodal suppressed conditions and the natural alternation rate (see Methods). The statistical comparison between the PSM and the empirical AV suppressed data throughout time did not reveal violations of the model (one-tailed (right), paired, t-test, all $p>0.05$). Thus, the data clearly illustrates a facilitation of audiovisual events under suppression compared to unisensory conditions, but not a violation beyond the race model (this replicates the behavioral study of Pápai and Soto-Faraco [15]).

### *2.3.3.2. ERP results*

#### *2.3.3.2.A. Hypothesis-driven analysis: Multisensory ERP responses at occipito-parietal electrodes*

Mean ERP values of the ROI electrodes were calculated for audio, visual and audiovisual events. We tested for differences between the summed auditory and visual ERPs and the audiovisual ERPs (summed to the no-stimulus ERPs, see Methods). The paired t-tests (two-tailed) in the time window of 70-540 ms did not reveal any significant deviance (after applying Guthrie and Buchwald's correction for multiple comparisons [56] (**Fig. III/2.**)). Please note this contrast is mostly frequently done using a two-tailed test, since it is more conservative and hence, any significant effects detected can be interpreted more confidently [34,36,37]. We followed the same tradition here. However, given this first negative finding, in order to increase sensitivity to multisensory effects if any, we lowered the significance threshold assuming a one-tailed (directional) paired t-test, based on the assumption of multisensory ERP reveals in higher amplitude than the unisensory sum. This comparison suggested a small window of significant difference after correction, but late in the ERP (between 394-432ms). This effect then happened after early sensory processes. Hence, like the behavioral responses, the ERPs did not reveal a convincing sign of genuine bottom-up MSI between sounds and unaware visual events. Still, in order to understand the underlying mechanisms further, we decided to run some complementary analyses.

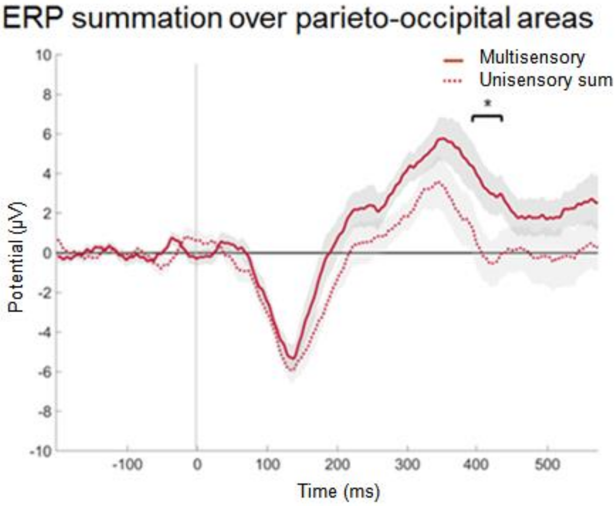


**Figure III/2.** *ERP summation over parieto-occipital areas.* The summed audio and visual suppressed ERPs, called 'Unisensory sum' is contrasted against the sum of audiovisual suppressed and no-stimulus ERPs, called 'Multisensory' over the average of ROI areas (P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, O2) with SEM. The paired t-test (one-tailed (right)) suggested some statistically significant difference, in the time window of 394-432ms after Guthrie and Buchwald's correction. The 0 time point corresponds to stimulus onset.

### *2.3.3.2.B. Visual evoked potentials (VEPs) over ROI area*

Here, we addressed the effects of awareness on VEPs, to check if the visual events under suppression generated a measurable response and also, if our results were in line with previous studies measuring ERPs to visual events presented below awareness. Because of the 3:1 difference in suppressed and dominant number of trials in our design, we randomly picked trials from the suppressed condition to equate the less populated dominant condition (40 +-7 trials/condition after artefact rejection). We then calculated the average VEPs over the ROI area, for suppressed and dominant conditions (**Fig. III/3.A**), and ran paired t-test (two-tailed) between the two waveforms within the time window 70-540 ms. The P1 waves were not very pronounced, what is maybe not so surprising considering the weak stimulus strength of the visual flash [53]. Additionally, both the suppressed and dominant VEPs showed N1 and late positivity (~P3) visual responses. The N1 response to dominant visual events appeared larger than the suppressed one, in line with many previous studies [49–52], and late positivity was reduced in suppressed compared to dominant condition, again, in line with what has been shown in the past

[44,51,55] (despite some discrepancies exist for both of the components [47,48]). Yet, please note that these differences, albeit mostly being in the expected direction, did not reflect statistical significance after Guthrie and Buchwald's correction (all $p>0.05$). Still, the relevant result here is that the VEP appeared for suppressed visual events, whether weaker than the dominant ones or not, but ensuring that these ERPs could be indeed measured in our paradigm.

### *2.3.3.2.C. Multi- and unisensory evoked potentials over ROI area*

The comparison of audiovisual ERPs to the summation of unimodal ERPs presented above did not reveal convincing signs of clearly bottom-up MSI, very much like the behavioral effects did not violate the PSM benchmark. Still, in order to further understand the results, we inspected the single modality and audiovisual ERPs over the ROI area. The paired t-tests (two-tailed) ran in the time window 70-540 ms after the event revealed statistically significant differences between audiovisual and visual ERPs (92-166 ms and 208-432 ms, all $p<0.05$, corrected). Remarkably, the audiovisual and audio ERP waves were not different (all $p>0.05$), in fact they seemed to overlap quite a lot (**Fig. III/3.B**). This pattern of results indicated that the ERP response to the audiovisual events was clearly dominated by the auditory evoked potential. This was true despite there was a visual response when measured alone, and that the electrode cluster we focused on should be visually responsive. Thus, the behavioral advantage in switch latencies for audiovisual events was not particularly reflected in the ERP responses over this ROI and time window. This is perhaps not surprising, because of the nature of the behavioral effects does not lead to conclude on a co-activation, or non-linear interaction at the sensory level.

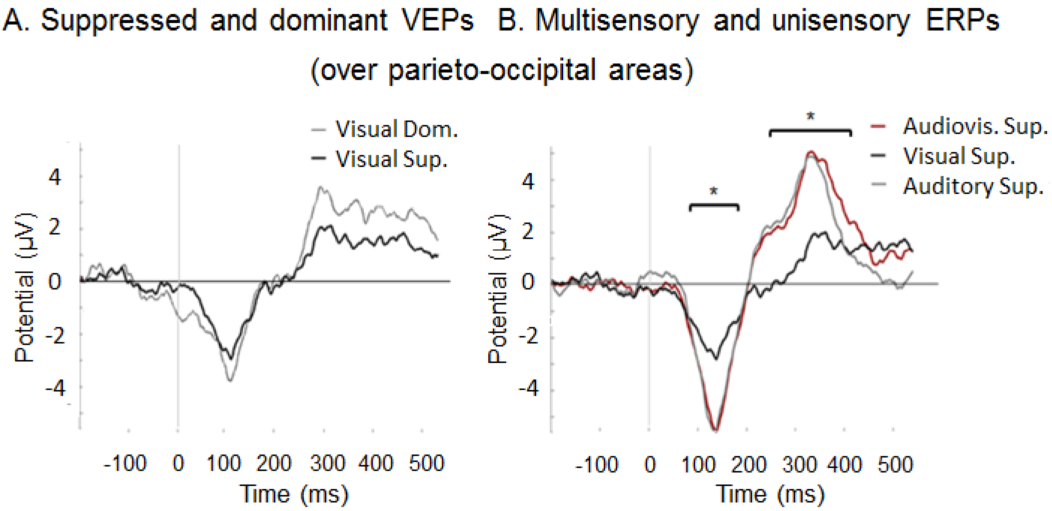
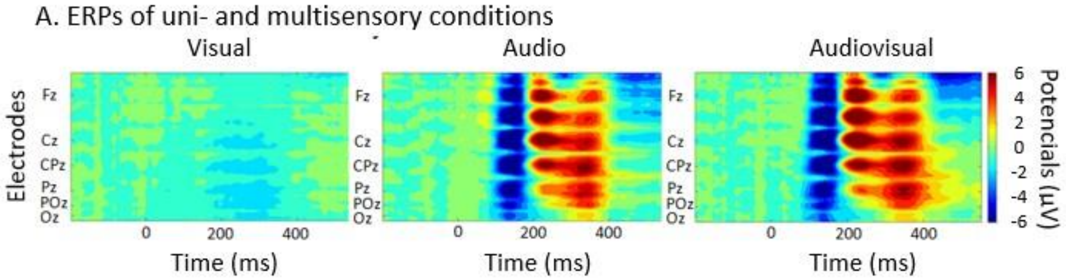


**Figure III/3.A** *Suppressed and dominant VEPs.* Visual N1 and late positivity (~P3) over ROI (P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, O2) areas evoked by dominant or suppressed visual event are plotted respectively. There was no statistically significant difference found between the conditions. Visual dominant condition is indicated by grey, the visual suppressed by black color. **B** *ERPs of multisensory and unisensory conditions.* ERP are presented in audiovisual, visual and auditory suppressed conditions, marked by red, black and grey colors, respectively over ROI areas. Significant differences between visual and audiovisual suppressed conditions are marked by '*'. In both of the graphs 0 time point corresponds to stimulus onset time.

### *2.3.3.2.D. Exploratory analysis: Multisensory ERP responses over all of the scalp*

According to the hypothesis-driven ROI analyses above, MSI cannot be concluded based on the ERP response. One assumption underlying those analyses, based on a good number of previous papers, was that MSI effects might express as a modulation of the VEP, over the parieto-occipital areas, hence the use of an ROI. Still, one might raise the question that putative multisensory effects might express at other scalp locations. Here, we ran exploratory analyses on the ERPs across the scalp. We used the electrode-by-electrode average data over all the scalp locations for each condition (**Fig. III/4.A**) in the time window of 70-540 ms. We first ran two-tailed, followed by a one-tailed (right) paired t-test in order to reveal any weak multisensory effect, if any, by using for the same contrast used in the ROI analysis: comparing the summed unisensory ERPs (audio suppressed plus visual suppressed) versus audiovisual ERPs plus the no-stimulus ERPs (**Fig. III/4.B**). After, cluster-based correction, there were no significant differences (all $p>0.05$). For completeness, similar to the ROI analysis, we also tested for differences between each unisensory and the audiovisual ERPs, at each electrode/time (by paired, two-tailed t-tests). Similarly to the ROI analysis, the statistical test revealed significant differences between audiovisual and visual conditions ERPs: 99-184 ms negative ERP shift over centro-medial and parietal areas all $p<0.05$ (corrected), a positive shift 192-426 ms over cento-medial areas, all $p<0.05$ (corrected), and a later negative shift 442-540 ms over centro-medial areas, all $p<0.05$ (corrected) (**Fig. III/4.**). Yet, the audiovisual ERPs did not differ from auditory ERPs. Thus, all in all, the pattern resulting from the analysis across the scalp was very similar to the one over the ROI area: the summed unimodal ERPs accounted for the audiovisual ERPs.


A. ERPs of uni- and multisensory conditions

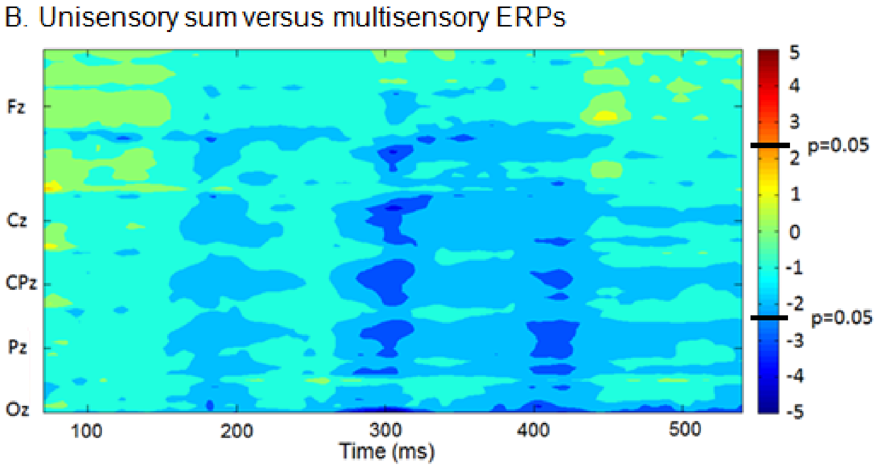

B. Unisensory sum versus multisensory ERPs

**Figure III/4.A** *ERPs of uni- and multisensory conditions.* ERP potentials (µV) are presented in visual, audio and audiovisual conditions as a function of time in all electrodes. 0 time point corresponds to stimulus onset time. **B** *Unisensory sum versus multisensory ERPs.* The t-values of the paired t-test (two-tailed) are plotted as a function of time (70-540ms), across the scalp. The paired t-test did not reveal any statistically significant difference of this contrast after Guthrie and Buchwald's correction. Color map shows t-values, from positive values (red) to negative values (blue). Please note that the t-values of ±2.2622 correspond to the p-value of 0.05.

### *2.3.4. Discussion*

The focus of the current study was to seek for evidence supporting MSI based on bottom-up mechanisms alone. To do so, we used visual stimuli presented below awareness combined with sounds, in a protocol where top-down selection based on expectation or attention would play a minimal role. Thus, the question would be whether a sound can 'rescue' visual events to awareness by means of MSI resulting from bottom-up mechanisms. The behavioral results in the BR task revealed a cross-modal benefit, since the typical switch in perceptual state following a flash on the suppressed eye happened sooner when a sound coincided with the visual event. This would initially be consistent with the hypothesis of bottom-up MSI. However, this cross-modal benefit could not be univocally ascribed to bottom-up MSI (i.e., co-activation), because empirical data from audiovisual events did not deviate from the prediction of probability summation (PSM, e.g., postulated by Raab [17]). This pattern of cross-modal facilitation without solid proof of MSI is very much in line with the previous behavioral study by Pápai and Salvador-Soto [15]. Yet, because bottom-up MSI might have happened without surpassing the limit of probability summation in behavior, we sought for neural correlates of audiovisual integration using ERPs. According to the integration hypothesis, and the logic used in a multitude of previous studies (for above-threshold stimuli), if audiovisual ERPs surpass the threshold set by the summed unimodal ERPs, then MSI can be inferred [32,34–36,38]. Despite a sign of non-linear effect in a late time window (around ~400ms post-stimulus) when lowering the significance threshold, we failed to see evidence for bottom-up integration at sensory stages of processing. Since the ERP analyses could not confirm the bottom-up integration hypothesis, hence by default favored the alternative hypothesis of the independent contribution from each unisensory response to the cross-modal ERP.

Before interpreting this result, it is relevant to discuss the neural correlates of the constituent, single modality, stimuli. Visual evoked potentials to events presented below awareness are often characterized by decreased amplitude of P1 [42–44], that can be even missing for threshold stimuli [53], as was the case in our study. In binocular rivalry paradigms, a diminished N1 and late positivity are often revealed for visual switching from suppression to awareness [44,51], though this pattern seems to be different when strong (very salient) visual probes are presented in the suppressed eye [47,48]. Despite any of the previous BR paradigms measuring VEPs directly comparable to our set-up, the VEPs for faint visual probes embedded in rivalry gratings seem to go in line with previous studies. In our data, both the N1 as well as a late positivity component were evoked in suppressed as well as dominant conditions. Although in both cases these visual components seemed to be attenuated under suppression, in accordance with previous BR studies measuring VEPs to perceptual switches [44,51], these effects were statistically unreliable, probably limited by the few number of trials/condition (since this analysis did not take part of the initial purpose of the study). Yet, what is important for the logic of interpretation in this context is that N1 as well as late positivity were effectively evoked by our visual stimuli even under suppression. Because these components have been also associated to attention capture and orienting respectively, their presence would confirm the possibility of attentional capture by the suppressed stimuli [55] (indeed, the small cross-modal ERP effect by the P300 time window would support that).

The auditory events in our study produced a prototypical Auditory Evoked Potential (AEP) with large response amplitude, given their above threshold strength, over parieto-occipital areas [32,34]. The audio stimulus was supra-threshold regardless of whether the target visual percept was suppressed or dominant, hence its associated AEP with the typical N1-P3 complex reflecting to sensory processing followed by attentional mechanisms did not vary between these two conditions.

Possibly, as a consequence of the weak visual and more robust auditory responses audiovisual ERPs where mostly driven by the auditory response. What is interesting is that regardless the lack of a bottom-up multisensory ERP response, above and beyond the summed unimodal responses in early time window; these audiovisual events did produce a behavioral advantage over each of the unisensory events. This pattern reinforces the idea that such behavioral advantage was not based on bottom-up integration in sensory processing, and probably originated further down the stream of information processing. This interpretation might be strengthened by the small non-linear interaction appearing at late time window, when sensory processing is already quite predisposed to attentional influence. Furthermore, although we initially targeted occipital electrodes in order to pick up visual responses, the exploratory analysis across all scalp electrodes confirmed the ROI-centered findings.

Thus, all in all the results from the cross-modal ERP responses is in line with the behavioral results given by probability summation, favoring the independent contribution of the unisensory stimuli to the cross-modal behavioral benefit, and suggesting that there is no necessity to postulate an additional MSI mechanism to explain this cross-modal advantage. Of course, this does not mean that bottom-up MSI does not occur in other circumstances, but the present result has the implication that purely bottom-up mechanisms may not provide sufficient means for integrative operations below awareness. We tried to single out such mechanisms using audiovisual events where the visual component was presented below awareness, and in the absence of selective top-down attention or expectation about the moment of appearance or the particular feature content of the suppressed event or its association with the sounds. Of course, this does not preclude the possibility that MSI occurs under these circumstances, if it is guided by top-down mechanisms leading to expectation or selective attention.

In the ERP literature, multisensory effects have been already demonstrated by using abrupt audio and visual stimuli [2,32,35,36,38] although since the events were presented above level of awareness, furthermore expectation/anticipation [32,36,38], or top-down attention [2,39] may have had a crucial influence on enabling MSI, evading the present research question. Indeed, when attention has been explicitly manipulated, it seems that multisensory benefits, even those arising from simple temporal coincidence, weaken [2,39]. The present results go one step forward, and suggest that some form of top-down modulation might be needed to enable even the most rudimentary forms of stimulus-driven MSI integration.

Nevertheless, it must be mentioned that behavioral multisensory benefits have been demonstrated in the past for unconscious visual stimuli [6,9,10], although many of these multisensory benefits were never tested against a probability summation baseline, leaving the possibility of more parsimonious explanations open.

Additionally, beyond the lack of probability summation baseline, in many other studies selective attention and/or expectation was simply not controlled for [57,58], resulting in the possibility of a top-down facilitation of the multisensory effect below awareness. Despite these effects are interesting in themselves, they do not speak directly to bottom-up MSI integration.

Hence, amongst the wider context of literature addressing whether MSI can occur for unaware stimuli, the main conclusion to emerge from the present findings is that even though cross-modal benefits can appear behaviorally, and can indeed furnish the observer with an adaptive advantage over unisensory situations, these benefits may not be exclusively grounded on bottom-up mechanisms of sensory integration. Rather, we suggest that the behavioral benefit for cross-modal events is, more likely, based on the combination of bottom-up attentional capture of each unisensory stimuli individually. Despite bottom-up multisensory integration is indeed still a possibility, for now the individual contribution seems to be the most parsimonious explanation when the potential for top-down modulation is minimized.

## *2.3.5. Methods*

### *2.3.5.1. Participants*

Data from 9 naïve observers was used (four female, average age 21.44 ± 2.1 years), and data from one additional participant was excluded (as he failed to run all of the blocks). The participants had normal or corrected-to-normal vision, normal stereo acuity, and presented no strong eye preference (as defined by perceptual predominance during binocular rivalry). The participants received 10 €/hour in return for taking part in the study.

### *2.3.5.2. Ethics statement*

Participants gave written informed consent, and all methods were carried out in accordance with Declaration of Helsinki, under a protocol approved by the local ethics committee of the University of Pompeu Fabra (CEIC - Parc de Salut Mar).

### *2.3.5.3. Apparatus and Stimuli*

Visual stimuli were created in MATLAB using PsychToolbox toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007), displayed on a 19.8 inch CRT monitor (1024 x 768 pixels; 100 Hz refresh rate). The stimuli were displayed on a plain grey background (13.9 cd/m$^2$). The two rival stimuli were contained within circular regions (11.5° ∅) defined by a Gaussian envelope (SD= 0.13°). One rival stimulus was a horizontal Gabor grating with spatial frequency 1.2 cycles/° (mean luminance 17 cd/m$^2$). The other rival stimulus was a radial checkerboard pattern whose mean luminance value was 19 cd/m$^2$. Please note that the luminance values are the group average after the adaptive staircase measurement. This luminance imbalance was set to increase the suppression depth of the Gabor grating, which was always presented on the dominant eye. Both rival visual stimuli (Gabor patch, radial checkerboard) were black-and-white and low-contrast to favour multisensory integration [59,60], and were centred on a black fixation cross (size of 0.25° and luminance of 3 cd/m$^2$) surrounded by a grey circle in the centre (0.5°, luminance of 10 cd/m$^2$). Additionally, each grating was surrounded by a black circle frame (0.2° width) presented simultaneously on the left and right halves of the monitor, with a centre-to-centre separation of 9.7°. The frames were binocularly matched therefore provided dichoptic stimuli for maintaining stable binocular alignment. These stimuli were viewed, one to each eye through a mirror stereoscope, giving a distance from the monitor to the eye of ~57 cm. The observers' head rested in a forehead-chin rest.

When subjects were exposed to the rival stimulus through the stereoscopically, they would experience alternations between the Gabor and the checkerboard. During this alternation, there were three types of event that could occur at random times; a visual flash presented on the lower part of the Gabor patch, a sound, or the flash and sound at the same time. The visual flash consisted of a 30 ms (10 ms fade-in/-out) contrast increment of the lower hemisphere of the Gabor grating. The size of the contrast increment was set individually to be at detection threshold under suppression (inter-participant average luminance =18 cd/m$^2$). The sound was a 500 Hz, 40 dB, 30 ms tone (10 ms ramps in/out) with a 44.1 kHz sampling rate. The sounds where presented from two speakers located on the two sides of the monitor, vertically aligned to the location of lower half of Gabor grating. The timing of the audiovisual stimuli delivery occurred within 1 ms precision, as calibrated on a BlackBox Toolkit (Cambridge Research System). Responses were reported by key presses. The study was run in a dimly lit, sound attenuated test room.

#### *2.3.5.4. Procedure*

We required the participants to covertly monitor their current percept (Gabor patch or the radial checkerboard) meanwhile fixating on the fixation cross, by means of two keys. Participants were instructed to press both keys when a piecemeal mixture of the two patterns was visible. During the experiment, besides the planned pauses, participants were allowed to take a break any time they needed by releasing both of the keys. These trials were repeated in a subsequent run. Each observer participated in two or three 180-min experimental sessions, in two-three consecutive days (varied depending on the alternation rate of the subject).

### *2.3.5.4.A. Pre-experiment calibration*

After few minutes of dark adaptation, each experimental session began with several calibration runs, starting with the calibration of the mirror stereoscope, followed by a training period where participants became acquainted with the BR paradigm. Then, the relative dominance of the radial checkerboard was set between 65% and 75% using an up-down adaptive staircase to adjust contrast of the checkerboard [61]. Next, we set the individual threshold of probe detection for flashes on the Gabor patch under suppression. Please note, that probe threshold measured under suppression result in a relatively strong stimulus if presented under dominance (typical sensitivity lost in suppression is 0.3 to 0.5 log units relative to dominance [6,62]). During threshold measurement, visual flashes were delivered during suppression periods with a delay of 1000 or 1500 ms locked to the initial of the suppression, and the contrast increments were adjusted by an adaptive staircase procedure designed to find the 50% detection threshold [61]. Finally, we tracked the natural alternation dynamics for a 3 minutes session in the very beginning of the experiment.

### *2.3.5.4.B. Experiment*

Observers continuously monitored (and reported by key press) BR alternations between the Gabor patch and the radial checkerboard while their EEG were recorded. The visual, auditory or audiovisual events were presented at pseudo-random moments (see below), and each could be presented when the Gabor was reported dominant or suppressed; hence, the visual dominant (VD) or visual suppressed (VS) events, the auditory dominant (VD) or suppressed events, and the audiovisual dominant (AVD) or suppressed (AVS) events, respectively. Please note that dominant or suppressed, for the audio alone conditions, is a label variable denoting whether the Gabor grating was, respectively, dominant or suppressed, at the moment of sound presentation. This was done to align this condition with the other two in terms of perception.

The intervals between events were composed of a fixed 5-s-delay plus a random 1-3 s jitter which was refreshed if a key press happened. Therefore, event presentation became temporally unpredictable and the events uncorrelated in time, which prevented possible top-down modulation and/or motor preparation [36] (**Fig. III/5.**).

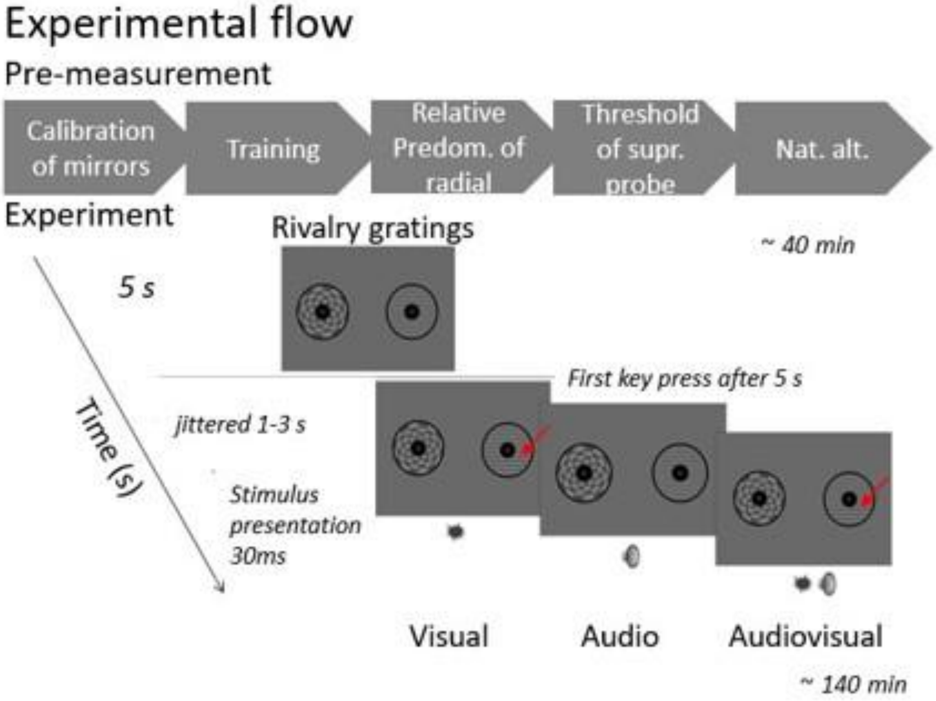


**Figure III/5.** *Experimental flow.* Each session contained of two parts, starting with different steps of pre-measurement and followed by the experiment by itself. During the experiment, each trial started with a 5-s-waiting-period. Furthermore, SOAs were fixed to jittered 1-3s from the first coming key-press occurred after the 5-s-waiting-period. Visual, audio or audiovisual events were presented on the dominant (D) or suppressed (S) Gabor patch. The flash was presented on the lower part of the gratings in visual and audiovisual conditions, marked by red arrows on the graph for better visualization.

Since the events of interest were those under suppression, out of the 280 per modality condition, 3/4 were presented during Gabor suppression and 1/4 under Gabor dominance. Hence, each event modality was equally likely within each dominance condition. No events were ever presented during a piecemeal percept. Participants were asked to report binocular dynamics without any special instruction related to possible visual or audio abrupt stimuli. Considering the demanding nature of monitoring the stochastic fluctuation of rivalry and the unpredictive presentation time of the uninformative (i.e., task-irrelevant) events, we can assume little chance for task-related top-down systematic biases toward the stimuli, other than the introspective monitoring of the rivalry itself.

#### *2.3.5.4.C. EEG recording*

During the experimental sessions, EEG data were acquired using 60 active electrodes (actiCAP, Brain Products GmbH, Munich, Germany) placed after the 10-20 international system, with the tip of the nose as online reference and AFz as ground. Data was referenced offline to the average of the left and right mastoids. Brain Vision Recorder (Brain Products, GmbH, Munich, Germany) was used for signal recording at a sampling rate of 500 Hz. Electrode impedance was kept below 5 kOhm. Horizontal (Heog) and vertical (Veog) electro-oculograms were recorded by two external electrodes, and used for off-line artefact rejection.

#### *2.3.5.5. Analysis*

##### *2.3.5.5.A. Behavioural data*

In order to reduce inter-individual variability in overall alternation rate we normalized the absolute switch time by the natural alternation rate for each individual, measured in periods where no event was presented. We analyzed the time course of the probability of a switch in percept, time locked to event presentation, as a function of event type (time-probability analysis). This analysis was based on the probability of dominant/suppressed percept at each sampling point (0.025 alternation unit), and was informative as to how quickly the dominant percept changed after one of the events (presented under suppression/ dominance) in the design. One useful index in this type of analysis is the Mean Time to Switch (MTS), which indicated the time from event (A, V or AV under suppressed or dominant) presentation, when the probability of switch surpassed the probability of 50% of seeing that particular grating. Conscious report of the radial checkerboard was taken as the suppressed condition (the Gabor patch was not consciously perceived; the condition of interest), and conscious report of Gabor patch was taken as the dominant condition. Those trials where a switch happened before 250 ms were discarded from analysis, as the perceptual change was most likely not evoked by the event. We run a repeated measures ANOVA (with Greenhouse-Geisser correction where appropriate) on the MTS latency data with within participants' factors: percept dominance (*Gabor dominant*, *Gabor suppressed*) and event modality (*audio*, *visual*, *audiovisual*), to compare switch latencies between modality conditions. The a priori hypothesis that audiovisual MTS would be shorter than unimodal ones was tested with one-tailed paired-t tests, whilst other contrasts were tested by two-tailed paired-t tests.

As part of our planned analysis, we also included a contrast of the probability of switch after audiovisual suppressed events against a PSM calculated from the unisensory suppressed events [16,17]. The analysis has been run based on the adapted equation of probability summation for its use with BR switch times [15] (**Equation III/1.**).

$$P(T\leq t \mid R_{AV}) = (P(T\leq t \mid S_{NA}) - P(T\leq t \mid S_A)) + (P(T\leq t \mid S_{NA}) - P(T\leq t \mid S_V)) - (P(T\leq t \mid S_{NA}) - P(T\leq t \mid S_A)) * (P(T\leq t \mid S_{NA}) - P(T\leq t \mid S_V))$$

**Equation III/1.** Let the probability of switch at time T after an audio event be $P(T\leq t \mid S_A)$, and after a visual event be $P(T\leq t \mid S_V)$, and the probability of the empirical distribution of natural alternation (of Gabor suppressed) $P(T\leq t \mid S_{NA})$. Then, one can model the theoretical audiovisual distribution based on the probability summation (redundant audiovisual: $R_{AV}$), as the probability $P(T\leq t \mid R_{AV})$.

The equation considers the probability of switch after any of the unimodal events under suppression ($P(T\leq t \mid S_A)$, $P(T\leq t \mid S_V)$, after audio or visual events respectively), furthermore in order to avoid a double addition of variability unrelated to stimulus presentation, the probability distribution of natural alternation was also taken into account as Gabor suppressed $P(T \leq t \mid S_{NA})$. Then the modelled audiovisual distribution based on probability summation ($P(T \leq t \mid R_{AV})$ could be compared to the empirical probabilities distribution. The PSM provides a benchmark of what would be the quickest possible switch times that one would expect under the assumption of complete independence between visual and auditory sensory processing, according to a simple race model [17].

### *2.3.5.5.B. ERP data*

The EEG data were filtered with a band pass filter between 0.5 and 25 Hz and an additional 50 Hz of notch filter was applied. Data from noisy electrodes were omitted, and the missing data was interpolated from the surrounding electrodes. EEG signal was segmented in epochs time-locked to the onset of suppressed visual, audio and audiovisual events ([-200 600] ms relative to the onset of the event). Additionally, we selected no-stimulus condition trials (used for ERP summation, see below) based on 800 ms pre-stimulus periods (from -1000 to -200 ms relative to the onset of suppressed or dominant events).

We performed automatic artefact rejection on resulting trials: all trials with amplitude exceeding ±100 μV in any of the scalp electrodes were excluded. Furthermore, the automatic artefact rejection was followed by a manual one, with special focus on the parietal and occipital electrodes. Thus, trials with eye-blinks, saccades, head movements, extensive muscle movements such as bite artefacts were removed, resulting in average 119 ±47 trials/condition after artefact rejection (from the 210 collected trials). Those trials where a switch happened before 250 ms were discarded from analysis.

Baseline activity was defined over the -200 ms to 0 ms period of each epoch for the three suppressed conditions. For the no-stimulus condition, we selected the first 200 ms of the trial. Event related potentials (ERPs) were computed for both the tree suppressed (audio, visual and audiovisual, respectively) conditions and the no stimulus condition, for each participant.

*a. Hypothesis-driven analysis*

We focused on visual responses, in order to detect the possible effects of sound on visual processing, as a sign of early multisensory integration, partially in line with other studies [32,35,36,38]. Therefore, we restricted our analysis to a visual region of interest (ROI). The ROI included the electrodes P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, O2. ERP waveforms evoked by audio, visual or audiovisual events were analysed for this ROI area in the time window of 70-540 ms post-stimulus. In order to detect multisensory interactions, ERPs to the audiovisual events were compared to the algebraic sum of ERPs to the unisensory stimuli presented in isolation (i.e. audio and visual events), following earlier studies [34,37,63]. Based on our hypothesis, differences between summed unisensory and the audiovisual ERPs, if any, would suggest nonlinear interaction. Yet, if the summed ERP responses from the unisensory presentations are equivalent to the audiovisual ERPs, then one would have to assume that independent neural responses to each of the unisensory stimulus are simply summed in the audiovisual event presentation [64]. However, during the computation one needs to deal with the methodological problem of adding baseline activity twice, together with the actual ERPs, when calculating the sum of the individual responses [2,36]. Following Talsma & Woldorff [2], to address this problem we calculated baseline neuronal activity what was present in absence of stimuli presentation or key press (no-stimulus condition). We used this no-stimulus activity as a mean to estimate the baseline EEG response and added it to the audiovisual responses before comparing it to the sum of audio and visual neural correlates. The audiovisual and summed audio and visual ERPs were compared using paired two-tailed t-test, and testing directional hypothesis one-tailed test (left), with significant differences being considered when they involved at least 18 consecutive data points (36 ms) at $p<0.05$. This criterion was decided following Guthrie and Buchwald approach [56], in order to correct multiple comparisons.

*b. Exploratory analysis*

Beyond our main focus on the parietal-occipital areas, for completeness, we also looked at possible effects all over the scalp. We used paired t-tests (two-tailed, then one-tailed (right) to directional hypothesis) (significant level p=0.05) on all electrodes in the time window of 70-540 ms, to test the contrast between summed unisensory ERPs and audiovisual ERPs. ERP waveforms across conditions were compared with two-tailed paired t-tests. A cluster-based correction [65] (10.000 randomizations) for electrodes and latencies was applied to correct for multiple comparisons. We used Fieldtrip toolbox [66] and custom-made code for the exploratory statistical analysis.

### *2.3.6. Publication information*

### *2.3.6.1. Data availability*

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

### *2.3.6.2. Acknowledgements*

This study was funded by a Ministerio de Educación, Cultura y Deporte, Ref. FPU/1206471 to MSP, and ERC Starting Grant, Ref. 263145, Spanish Ministry of Science and Innovation (PSI2016-75558-P), and Comissionat per a Universitats i Recerca del DIUE- Generalitat de Catalunya (2014SGR856) to SSF.

### *2.3.6.3. Author contribution statement*

SSF and MSP designed the study, MSP conducted the experiment, SSF, MSP and MT analysed the data and SSF and MSP interpreted the data, and SSF, and MSP wrote the paper.

### *2.3.6.4. Additional information*

Competing financial interest: The authors declare no competing financial interest.

### *2.3.7. References*

1. Driver, J. & Spence, C. Multisensory perception: Beyond modularity and convergence. *Curr. Biol.* **10,** R731–R735 (2000).

2. Talsma, D. & Woldorff, M. G. Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J. Cogn. Neurosci.* **17,** 1098–114 (2005).

3. Stein, B. & Meredith, M. *The Merging Senses*. *Cognitive Neuroscience Series* (1993).

4. Hartcher-O, B. J., Soto-Faraco, S. & Adam, R. Editorial : A Matter of Bottom-Up or Top-Down Processes : The Role of Attention in Multisensory Integration. **11,** 10–12 (2017).

5. ten Oever, S. *et al.* The COGs (context, object, and goals) in multisensory processing. *Exp. Brain Res.* **234,** 1307–1323 (2016).

6. Lunghi, C., Morrone, M. C. & Alais, D. Auditory and Tactile Signals Combine to Influence Vision during Binocular Rivalry. *J. Neurosci.* **34,** 784–792 (2014).

7. Lunghi, C., Binda, P. & Morrone, C. Touch disambiguates rivalrous perception at early stages of visual analysis. *Curr. Biol.* **20,** 143–144 (2010).

8. Lunghi, C. & Morrone, M. C. Early Interaction between Vision and Touch during. *Multisens. Res.* **26,** 291–306 (2013).

9. Lunghi, C. & Alais, D. Touch Interacts with Vision during Binocular Rivalry with a Tight Orientation Tuning. *PLoS One* **8,** 1–8 (2013).

10. Zhou, W., Jiang, Y., He, S. & Chen, D. Report Olfaction

Modulates Visual Perception in Binocular Rivalry. *Curr. Biol.* **20,** 1356–1358 (2010).

11. Faivre, N., Mudrik, L., Schwartz, N. & Koch, C. Multisensory Integration in Complete Unawareness: Evidence From Audiovisual Congruency Priming. *Psychol. Sci.* **25,** 2006–2016 (2014).

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

13. De Meo, R. *et al.* Top-down control and early multisensory processes: Chicken vs. egg. *Front. Integr. Neurosci.* **9,** 1–6 (2015).

14. Mudrik, L., Faivre, N. & Koch, C. Information integration without awareness. *Trends in Cognitive Sciences* **18,** 488–496 (2014).

15. Pápai, M. S. & Soto-Faraco, S. Sounds can boost the awareness of visual events through attention without cross-modal integration. *Sci. Rep.* **7,** 41684 (2017).

16. Miller, J. Divided attention: Evidence for coactivation with redundant signals. *Cogn. Psychol.* **14,** 247–279 (1982).

17. Raab, D. H. Statistical facilitation of simple reaction times. *Trans. N. Y. Acad. Sci.* **24,** 574–590 (1962).

18. Otto, T. U. & Mamassian, P. Noise and correlations in parallel perceptual decision making. *Curr. Biol.* **22,** 1391–

1396 (2012).

19. Ooi, T. L. & He, Z. J. Binocular rivalry and visual awareness: The role of attention. *Perception* **28,** 551–574 (1999).

20. Hancock, S. & Andrews, T. J. The role of voluntary and involuntary attention in selecting perceptual dominance during binocular rivalry. *Perception* **36,** 288–298 (2007).

21. Paffen, C. L. E. & Van der Stigchel, S. Shifting spatial attention makes you flip : Exogenous visual attention triggers perceptual alternations during binocular rivalry. *Attention, Percept. Psychophys.* **72,** 1237–1243 (2010).

22. Pannunzi, M. *et al.* Deconstructing multisensory enhancement in detection. *J. Neurophysiol.* **113,** 1800–18 (2015).

23. Macaluso, E. Modulation of Human Visual Cortex by Crossmodal Spatial Attention. *Sci. (New York, NY)* **289,** 1206–1208 (2000).

24. Mcdonald, J. J., Teder-Sälejärvi, W. A. & Ward, L. M. Multisensory Integration and Crossmodal Attention Effects in the Human Brain. **1791,** (2001).

25. Driver, J. & Spence, C. Attention and the crossmodal construction of space. *Trends in Cognitive Sciences* **2,** 254–262 (1998).

26. Bolognini, N., Rasi, F. & Ládavas, E. Visual localization of

sounds. *Neuropsychologia* **43,** 1655–1661 (2005).

27. Frassinetti, F., Bolognini, N. & Làdavas, E. Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp. Brain Res.* **147,** 332–343 (2002).

28. Teder-Sälejärvi, W. A., Russo, F. Di, Mc Donald, J. J. & Hillyard, S. A. Effects of Spatial Congruity on Audio-Visual Multimodal Integration. 1396–1409 (2005).

29. Spence, C. & Driver, J. Audiovisual links in exogenous covert spatial orienting. *Percept. Psychophys.* **59,** 1–22 (1997).

30. Bolognini, N., Frassinetti, F., Serino, A. & Làdavas, E. 'Acoustical vision' of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Exp. Brain Res.* **160,** 273–282 (2005).

31. Meredith, M. A., Nemitz, J. W. & Stein, B. E. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.* **7,** 3215–29 (1987).

32. Giard, M. H. & Peronnet, F. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* **11,** 473–490 (1999).

33. Morís Fernández, L., Visser, M., Ventura-Campos, N., Ávila, C. & Soto-Faraco, S. Top-down attention regulates the neural expression of audiovisual integration. *Neuroimage* **119,** 272–285 (2015).

34. Molholm, S. *et al.* Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cogn. Brain Res.* **14,** 115–128 (2002).

35. Fort, A., Delpuech, C., Pernier, J. & Giard, M.-H. Dynamics of Cortico-subcortical Cross-modal Operations Involved in Audio-visual Object Detection in Humans. *Cereb. Cortex* **12,** 1031–1039 (2002).

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

37. Murray, M. M., Foxe, J. J., Higgins, B. a, Javitt, D. C. & Schroeder, C. E. Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: a high-density electrical mapping study. *Neuropsychologia* **39,** 828–44 (2001).

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

39. Talsma, D., Doty, T. J. & Woldorff, M. G. Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cereb. Cortex* **17,** 679–690 (2007).

40. Mangun, G. R. Neural mechanisms of visual selective attention. *Psychophysiology* **32,** 4–18 (1995).

41. Woldorff, M. G. *et al.* Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc. Natl. Acad. Sci.* **90,** 8722–8726 (1993).

42. Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M. & Ro, T. To See or Not to See: Prestimulus Phase Predicts Visual Awareness. *J. Neurosci.* **29,** 2725–2732 (2009).

43. Kornmeier, J. & Bach, M. The Necker cube - An ambiguous figure disambiguated in early visual processing. *Vision Res.* **45,** 955–960 (2005).

44. Roeber, U. *et al.* Early correlates of visual awareness in the human brain : Time and place from event-related brain potentials. *J. Vis.* **8,** 1–12 (2008).

45. Del Cul, A., Baillet, S. & Dehaene, S. Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* **5,** 2408–2423 (2007).

46. Koivisto, M. *et al.* The earliest electrophysiological correlate of visual awareness? *Brain Cogn.* **66,** 91–103 (2008).

47. Metzger, B. *et al.* Regulating the Access to Awareness: Brain Activity Related to Probe-related and Spentaneous Reversals in Binocular Rivalry. *J. Cogn. Neurosci.* 1–14

(2017). doi:10.1162/jocn

48. Valle-Inclán, F., Hackley, S. a, de Labra, C. & Alvarez, a. Early visual processing during binocular rivalry studied with visual evoked potentials. *Neuroreport* **10,** 21–5 (1999).

49. Ojanen, V., Revonsuo, A. & Sams, M. Visual awareness of low-contrast stimuli is reflected in event-related brain potentials. *Psychophysiology* **40,** 192–197 (2003).

50. Koivisto, M., Revonsuo, A. & Lehtonen, M. Independence of visual awareness from the scope of attention: An electrophysiological study. *Cereb. Cortex* **16,** 415–424 (2006).

51. Kaernbach, C., Schroger, E., Jacobsen, T. & Roeber, U. Vision, Central: Effects of consciousness on human brain waves following binocular rivalry. *Neuroreport* **10,** 713–716 (1999).

52. Marzi, C. A., Girelli, M., Miniussi, C., Smania, N. & Maravita, A. Electrophysiological correlates of conscious vision: evidence from unilateral extinction. *J. Cogn. Neurosci.* **12,** 869–877 (2000).

53. Pins, D. & Ffytche, D. The neural correlates of conscious vision. *Cereb. Cortex* **13,** 461–474 (2003).

54. Eimer, M. & Mazza, V. Electrophysiological correlates of change detection. *Psychophysiology* **42,** 328–342 (2005).

55. Railo, H., Koivisto, M. & Revonsuo, A. Tracking the processes behind conscious perception: A review of event-

related potential correlates of visual consciousness. *Conscious. Cogn.* **20,** 972–983 (2011).

56. Guthrie, D. & Buchwald, J. S. Significance testing of difference potentials. *Psychophysiology* **28,** 240–244 (1991).

57. Aller, M., Giani, A., Conrad, V., Watanabe, M. & Noppeney, U. A spatially collocated sound thrusts a flash into awareness. *Front. Integr. Neurosci.* **9,** 1–8 (2015).

58. Alsius, A. & Munhall, K. G. Detection of Audiovisual Speech Correspondences Without Visual Awareness. *Psychol. Sci.* **24,** 423–431 (2013).

59. Jaekl, P., Pérez-Bellido, A. & Soto-Faraco, S. On the 'visual' in 'audio-visual integration': A hypothesis concerning visual pathways. *Exp. Brain Res.* **232,** 1631–1638 (2014).

60. Perez-Bellido, A., Soto-Faraco, S. & Lopez-Moliner, J. Sound-driven enhancement of vision: disentangling detection-level from decision-level contributions. *J. Neurophysiol.* **109,** 1065–1077 (2013).

61. Kaernbach, C. A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing. *J. Acoust. Soc. Am.* **88,** 2645–2655 (1990).

62. Nguyen, V. a, Freeman, a W. & Wenderoth, P. The depth and selectivity of suppression in binocular rivalry. *Percept. Psychophys.* **63,** 348–360 (2001).

63. Molholm, S., Ritter, W., Javitt, D. C. & Foxe, J. J.

Multisensory Visual-Auditory Object Recognition in Humans: A High-density Electrical Mapping Study. *Cereb. Cortex* **14,** 452–465 (2004).

64. Murray, M. M. *et al.* Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb. Cortex* **15,** 963–974 (2005).

65. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data **164,** 177–190 (2007).

66. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* (2011).

# CHAPTER 3.

# Discussion and conclusions

## *3.1. Bottom-up multisensory integration or top-down regulation?*

In everyday situations, information about the environment arrives to the brain simultaneously from different senses. Oftentimes, this cross-modal redundancy helps us to react faster and more efficiently, compared to processes based on a single modality. A case in point is auditory enhancement of visual processing. While several studies show that sounds presented simultaneously with visual events enhance performance in visual detection (McDonald, Teder-Sälejärvi, & Hillyard, 2000; Frassinetti et al., 2002; Teder-Sälejärvi et al., 2005; Perez-Bellido et al., 2013) or eventually increase perceived luminance (the explanations of these phenomena might be varied, and remain disputed (e.g., Jaekl et al., 2014 for a review). One might assume a fast, bottom-up integration schema, which receives some support from the discovery of direct cross-modal connections between early sensory brain areas (e.g., from audio to visual; Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003), or even Superior-Colliculus-mediated improvements in simple detection tasks (Stein & Meredith, 1993; Stein, 1998). Yet, a wide range of findings suggest variations from this interpretation of feed-forward or bottom-up multisensory integration (Frassinetti et al., 2002; McDonald et al., 2000; Teder-Sälejärvi et al., 2005, Stein et al., 1996, 1998), supporting instead other mechanisms based on recurrent architectures such as response bias (Odgaard, Arieh, & Marks, 2003; Marks, Ben-Artzi, & Lakatos, 2003) anticipation (Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002), imagery (Carter, Konkle, Wang, Hayward, & Moore, 2008), expectation (Lovelace, Stein, & Wallace, 2003) or even attention cueing (Lippert, Logothetis, & Kayser, 2007). While the most recent cross-modal studies controlled for the majority of the above-mentioned mechanisms, the attentional account is more difficult to disentangle, and hence was still not completely ruled out in many occasions (as it has been discussed in Chapter 1., Section 1.4.4.). Therefore, the question whether pre-attentive, automatic multisensory integration can happen has remained disputed.

## *3.2. Evaluation of our findings*

### *3.2.1. Evaluation of the initial hypotheses*

The present thesis addressed whether multisensory integration may occur for events below awareness. Testing cross-modal integration for unconscious events is not the only approach to infer bottom-up integration, but one that has some advantages because the unaware stimulus cannot, in principle, guide strategies that could bias responses. We focused on abrupt, uninformative and unpredictive cross-modal co-occurrence of events that do not share any perceptual congruency, beyond simple spatio-temporal alignment. These conditions are fairly different from the task-relevant, sustained presentation of perceptually congruent stimuli that have led so far to bottom-up multisensory integration below awareness. Yet, we reasoned that the conditions used in the present experiments are best suited to single out the possible effects of bottom-up mechanisms, if any, causing multisensory integration below awareness. This detail is important because the potential influence of task- or stimulus-related top-down attention on putative cross-modal integration effects in our case is strongly diminished or, ideally, absent. Hence, our results bear directly on the question of whether cross-modal integration is possible based solely on bottom-up mechanisms.

To do so, we measured the perceptual switches in a BR paradigm designed to elicit fluctuations between two simple, meaningless images: a Gabor grating, and a checkerboard plaid. The measure of interest was the latency of perceptual switches as a function of occasional visual stimuli (abrupt visual flash below threshold), acoustic stimuli (a hearable, abrupt audio stimulus) or the (unpredictable) combination of the two. Thus, both of the stimuli were abrupt transients, and were uncorrelated, thus gave no chance to building-up cross-modal associations over time. The cross-modal modulation of visual events presented under suppression provided the critical condition of the test. According to our hypothesis, if such cross-modal integration mechanisms are at play, then we can expect a cross-modal behavioral benefit after audiovisual events, resulting in earlier (or more likely) perceptual switches, in comparison to switches after either unimodal stimuli.

Similarly to many previous BR studies using sustained stimulation, we showed cross-modal facilitation effects for abrupt stimulation. Perceptual switches occurred sooner after suppressed visual events were accompanied by sounds, than after any of the two individual unisensory stimuli. Thus, our hypothesis seemingly held. However, after testing our empirical data against the probability summation model, and despite of the similarity of our behavioral findings with prior studies using sustained cross-modal stimulation, we could not conclude on multisensory integration as the necessary cause of this benefit. Hence, the most parsimonious interpretation of our results suggested that each sensory stimulus contributed independently to the cross-modal effect, as opposed to multisensory integration. In the following sections I will discuss the results of our four experiments and discuss them in relation to the literature.

### *3.2.2. The effect of abrupt audiovisual stimulation in BR, without top-down selective attention*

The results from Experiment 1 help illustrate one of the main findings of this thesis, and the logic behind the interpretation of results. As stated above, we found earlier perceptual switches after audiovisual events in comparison with audio and visual events by themselves, but this advantage could be putatively based on the individual contribution of unisensory stimuli. Similarly to Experiment 1, a recent study by Aller et al., (2015) also used abrupt spatial audiovisual stimulation in dynamic CFS to investigate whether a sound facilitates the detection of a visual suppressed flash, depending on audiovisual spatial congruency. They found that sounds spatially collocated with visual stimuli boosted the suppressed percept into perceptual awareness. The authors suggested that the effect happened due to automatic multisensory integration. However, like in other studies using BR or other methods of measuring cross-modal enhancement effects, there were some features of their paradigm that could allow top-down orienting. In their particular study, the audio events always coincided with the visual events in time, therefore acquiring a predictive value by reducing the temporal uncertainty about the onset of the visual target (additionally, please note that auditory events in this study had some task-relevance, due to the presence of auditory catch trials). One could therefore explain Aller et al.'s results with the temporal attention cueing mechanisms already proposed by Lippert et al., (2007) to account for cross-modal enhancement in the detection of near threshold visual events. So, Aller et al.'s study was a good example for cross-modal interactions using short abrupt stimuli, when visual events are maintained below the threshold of awareness, but it was not conclusive regarding bottom-up multisensory integration. In our studies, we tried to avoid the predictive value of sounds (and hence, their role in reducing temporal uncertainty) by making the sound unpredictive to visual events (just 25% of sounds coincided with a visual event in the suppressed percept), so that any potential effects would be likely based on bottom-up mechanisms.

### *3.2.3. Dissociating bottom-up attentional capture from multisensory integration*

If bottom-up multisensory integration cannot be concluded, then, where does the cross-modal benefit observed in Experiment 1 come from? We entertained the hypothesis that such cross-modal benefit would have to come from the contribution of each individual components of the cross-modal pair, independently. Hence, in Experiment 2 of this dissertation, we meant to disentangle multisensory integration from attentional effects. To do so we used a sound offset (gap of silence embedded in background noise) instead of a sound onset, an approach that has been used before to test attentional accounts. For example, Ngo & Spence (2012) adapted a previous experiment from Vroomen & de

Gelder (2000), reporting the ‘freezing’ phenomenon, but substituting the audio component of the audiovisual stimulus with a gap of silence. Because they still replicated the ‘freezing’ effect with gaps (absence) in the audio stimulus, the phenomenon has been attributed to attentional capture rather than to multisensory integration. In line with this logic, we found that the cross-modal facilitation in our BR protocol stood the same when using a sound offset (gap of silence). However, it is important to note, that the nature of Ngo and Spence’ manipulation and ours was very different in terms of predictability. The sound gaps in our study were statistically unpredictable as opposed to the ones used by Ngo and Spence (2012), which took place within an otherwise regular audio stream and thus, called for expectation and prediction. Nonetheless, there is a possible criticism to this logic, since the cessation of noise does not mean lack of input (absolute silence), as background room noise was rather inevitable (in our case, estimated to be 35 dB [A] by sound meter). But if we at least consider that audio gaps were a weaker stimulus than tones, then by means of inverse effectiveness, a stronger multisensory response would have been expected, something that did not happen. Therefore, again, the most straightforward interpretation of the results would be bottom-up attention capture deriving from the individual contribution of the two sensory stimuli, rather than bottom-up sensory integration. One could relate this distinction with the distinction between the race model

(Raab, 1962) and the co-activation model (Miller, 1982), explained in the introduction (Chapter 1., Section 1.1.1.).

Following this ‘attention capture’ logic, should we consider our auditory gaps just as any other transient, albeit consisting of a strength decrement instead of increment? There are precedents in the literature addressing the capacity of sound intensity increment/decrements to integrate with visual increment/decrements (Andersen & Mamassian, 2008). In Andersen and Mamassian’s study, there was a general improvement in audiovisual detection with respect to single modality stimuli, regardless the direction of the change (increment/decrement) in each modality, or its cross-modal congruency (e.g., visual increment paired with sound decrement). Despite the authors attributed this audiovisual improvement to multisensory integration, in the literature there is consensus about that salient stimulus can draw exogenous, involuntary attention to its location as well as to the time its occurrence. Therefore, whilst irrelevant sounds could perhaps increase visual sensitivity by integration (Stein, London, Wilkinson, & Price, 1996; Bolognini, Frassinetti, Serino, & Làdavas, 2005) they might also simply provide reduction of uncertainty (Lippert, Logothetis, & Kayser, 2007; Perez-Bellido et al., 2013). Thus, we believe that Andersen and Mamassian’s results, similarly to ours, are well within the attentional hypothesis, which was also considered in their paper.

### *3.2.4. The relevance of spatial alignment between auditory and visual events*

In the two first experiments, the auditory stimuli were placed relatively close but not perfectly aligned to the visual events. One might then wonder whether bottom-up multisensory integration failed just because of the lack of perfect spatial alignment. As mentioned in the introduction, the spatial rule of MSI, initially proposed to describe cross-modal spatial register in the pattern of neural responses in the Superior Colliculus of animals, has been shown to apply to psychophysical detection in humans in visual-auditory protocols (Frassinetti et al., 2002, 2005; Bolognini et al., 2005; Làdavas, 2008), albeit it is not as general as it was initially assumed (e.g., Murray et al., 2005; Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005). Frassinetti et al., (2002) and Bolognini et al., (2005), for example, demonstrated that detection of masked visual stimuli improved by task-irrelevant audio stimuli presented in the same rather than different spatial location. In those papers, the spatial as well as the temporal selectivity of the effect was exceptionally precise (within 100 ms in time and within 16° in visual angle). Therefore, in Experiment 3 we aligned the visual and audio stimuli in space in order to increase the likelihood of multisensory integration. However, the spatially congruent and incongruent audiovisual trials were statistically not different from each other, both showing a cross-modal benefit, not beyond the probability summation threshold.

Please note that despite the lack of perfect spatial alignment is a potential alternative explanation that is worth testing, the literature is full of cases where multisensory interactions occur despite spatial misalignment (a notable example is Murray et al., 2005). Indeed, cross-modal benefits have been reported for unaware visual stimuli in the study of Lunghi (2010) notwithstanding a difference of 15 cm between the location of tactile and visual stimuli. All in all, it seems that the absence of clear evidence for multisensory converge, above and beyond attention grabbing of each sensory modality, could not be explained by the lack of proper spatial alignment in the first two experiments.

### *3.2.5. Neural correlates of cross-modal interactions with unaware visual events*

The conclusion from Experiments 1-3 is that, despite there is an advantage of cross-modal signals in terms of producing faster perceptual switches in BR, this advantage cannot be tagged to multisensory integration. Yet, an important problem when relying on the lack of effects is the so-called type II error. In our particular case, one cannot be totally sure about whether multisensory integration happened or not, just because the cross-modal benefit did not surpass the probability summation benchmark. As we have stated in the introduction, whereas presence of non-linear interactions may be taken as a sign of integration, their absence leaves us without a definite proof. The only resort is then to call the principle of parsimony, whereby known attention capture effects could simple explain the results. This query motivated the fourth experiment in this thesis, reported in the second manuscript, addressing possible multisensory effects that, despite falling short of a super-additive advantage in behavior, might express in electrophysiological responses. Despite we still found a behavioral cross-modal benefit, the ERP summation test, which is used in literature for testing non-linear cross-modal effects above awareness, see (Giard & Peronnet, 1999; Murray et al., 2005; Talsma & Woldorff, 2005; Teder-Sälejärvi et al., 2002), failed to provide conclusive grounds for multisensory integration. Rather, the neural correlates happened to be in line with the explanation of the individual contribution of the stimuli, showing a prominent influence of audio stimulation in the cross-modal responses. Interestingly, regardless of the lack of a bottom-up multisensory ERP response, above and beyond the summed unimodal responses in the early time window; these audiovisual events did produce a behavioral advantage over each of the unisensory events. This pattern reinforced the idea that such behavioral advantage was not based on bottom-up multisensory integration and, probably originated further down the stream of information processing. This interpretation might be strengthened by the small non-linear interaction appearing at a late time window in the ERPs, when sensory processing was already quite predisposed to attentional influence. Please note that, although we initially targeted occipital electrodes in order to pick-up visual responses with increased statistical power, a follow-up, exploratory analysis across all scalp electrodes actually confirmed the ROI-centered findings.

### *3.2.6. Individual contribution of stimuli calls for bottom-up attentional capture*

Considering the transient nature of the stimuli, the contribution of bottom-up, exogenous, attention probably played a role (still in lack of top-down attention) in the cross-modal facilitation consistently observed in our experiments. Yet, as mentioned before, the safest interpretation of this facilitation is based on the independent contribution of each sensory event, in line with the literature concerning attention capture above awareness (Prinzmetal, Park, & Garrett, 2005; Remington, Johnston, & Yantis, 1992) and from suppressed stimuli, below awareness (Blake, Westendorf, & Fox, 1990; Dieter, Melnick, & Tadin, 2015; Lack, 1974; Ooi & He, 1999; Paffen & Van der Stigchel, 2010). Indeed, bottom-up attention capture is not a new phenomenon in BR literature. As described in the introduction (Chapter 1., Section 1.4.3.2.), these capture effects have been tested many times, however always on unisensory contexts within vision (Blake et al., 1990; Kanai, Moradi, Shimojo, & Verstraten, 2005; Ooi & He, 1999; Paffen & Van der Stigchel, 2010). The bottom-up attentional capture tend to elicit perceptual switches, although not every salient stimulus necessarily captures attention (Klimesch, Sauseng, & Hanslmayr, 2007; Van Rullen & Koch, 2003), as well as not all of the perceptual switches originate from attentional capture, of course (Pastukhov & Braun, 2007).

As we aimed to rule out the top-down attention deployed to stimuli, we capitalized on non-informative and task-irrelevant low-level stimulus without cross-modal feature congruency. At the same time we also managed to prevent a strong bottom-up attentional capture from visual events. The latter point was important, so that there was room for the possible additional effects of sounds, and the interaction between sound and flash Despite the weak visual stimuli used, their effect in single modality conditions was very similar to previously reported findings in the literature, for flashes on the rivalry stimuli (Paffen & Van der Stigchel, 2010) or on the background (Kanai et al., 2005), producing earlier switches. Additionally, we have shown that an abrupt audio stimulus also evoked perceptual switches. To the best of our knowledge, this is the first study to demonstrate that the presentation of a transient auditory stimulus results in faster perceptual switch in BR. Remarkably, the audio and visual stimuli elicited perceptual switches with very similar probabilities, regardless of that auditory stimuli were presented clearly above threshold, while visual stimuli below threshold level (beyond the suppression of BR for visual stimulus). Since the currently emerging views of BR states that the sensory processes are distributed at multiple levels: low- as well as high-level cortical areas (Leopold & Logothetis, 1996; Tong, Meng, & Blake, 2006), thus, an unseen visual information can be maintained and propagated from occipital visual regions to parietal and frontal areas (King, Pescetelli, & Dehaene, 2016) where visual information is highly susceptible to higher-order influences, such as attention. In addition to the switches with equal probabilities after unisensory events, perceptual switches happened also earlier after unisensory stimulation than in what would be expected without stimulation, a latency provided by fluctuations during the baseline situation when there were no events presented (Chapter 2., Supl. Mat., Section 2.2.S1.2.A.). These switches after unisensory events thus, are also proof of the attention capture of abrupt stimuli was effective for both auditory as well as visual events.

As discussed earlier, the conclusions from the present findings derive from several failed attempts to find signs of multisensory integration under conditions where it should putatively happen (Experiments 1-4), or finding successful cross-modal benefits under conditions where no multisensory integration should happen (Experiment 2). Despite this, one could still argue that cross-modal benefits might arise from the activation of multisensory neurons (or other integration mechanisms) that do not result in a super-additive effect (Meredith, Allman, Keniston,

& Clemo, 2012). The conclusion of the putative lack of bottom-up multisensory integration is based on the parsimony principle, whereby a simpler, known explanation (in this case, bottom-up attention grabbing of each modality signal independently) can already account for the results without the need to postulate a further mechanism (in this case, multisensory integration). This conclusion needs to be taken at face value, and perhaps further clarified by the two following disclaimers. First, the conclusion of this thesis does not necessarily mean that bottom-up multisensory integration does not occur in other circumstances, for example, when stimulation is above perceptual awareness. In fact, this is probably the case under some circumstances (e.g., Jaekl et al., 2014; ten Oever et al., 2016 for reviews). The present result has the implication that purely bottom-up mechanisms (achieved by unpredictive and uninformative, abrupt stimulus presentation) may not provide sufficient means for integrative operations below awareness. Second, the conclusion of this lack of integration does not bear on the possibility that multisensory integration occurs for unaware stimuli, if it is guided by top-down mechanisms such as expectation, imagery or selective attention. We did not test these cases, mostly because the focus was to address the potential for multisensory integration grounded on bottom-up integrative mechanisms alone. As it has been illustrated in the introduction, there are quite some examples of cross-modal interactions below awareness supported by top-down mechanisms (e.g. Aller et al., 2015; Guzman-Martinez, Ortega, Grabowecky, & Mossbridge, 2012).

### *3.2.7. Connecting the behavioral and neural responses to cross-modal events*

In this thesis, the account of the independent contribution of each sensory stimulus, has been strengthened by the neural correlates. Interestingly, in Experiment 4, while behaviorally the switches after the cross-modal events occurred earlier than after unimodal stimuli, the ERP data showed a slightly different pattern. Eventually, the auditory evoked potentials (AEPs) were statistically not different from ERPs elicited by multisensory events. While the auditory events (always above awareness) elicited strong AEPs, the threshold-level visual flash presented under suppression resulted in weak but measurable visual evoked potentials (VEPs). Hence, the auditory stimulus provided stronger ERPs than the weak visual stimulus, thus its dominance on the cross-modal ERP response may be not so surprising. This suggests a tentative explanation: on the one side, threshold visual stimuli on the suppressed eye might provide some weak basis for perceptual switch toward that percept, leaving the system at an ambiguity point. On the other hand, auditory events provide the system with a stronger signal, that is not specific to one or the other eye. It has been shown that external noise tends to elicit perceptual switches (Kanai et al., 2005; Lankheet, 2006; Moreno-

Bote, Rinzel, & Rubin, 2007), therefore, it is possible that sounds might have injected noise in the system that summed up to the weak but biasing signal coming from vision contributed to tilt the balance towards a perceptual switch.

### *3.2.8. The importance of testing for non-linear responses to multisensory stimuli*

As it has been discussed at length in the discussion, the multitude of studies measuring cross-modal responses under BR protocols can receive a variety of interpretations. Within these studies, some are of interest for us because they have provided conclusive evidence for bottom-up multisensory integration. Therefore, it is important to discuss this discrepancy between these studies and our conclusions. The main difference between the findings demonstrating bottom-up multisensory integration in past literature and our results derives from three factors: from the inclusion of a test of probability summation in the analysis of the behavioral data, from the inclusion of neuroimaging, ERP methods and, from the use of abrupt, unpredictive and uninformative stimuli whose cross-modal congruency is limited to spatio-temporal coincidence.

Despite the several studies that have demonstrated bottom-up multisensory integration putatively in lack of top-down modulations (Lunghi & Alais, 2013; Lunghi, Morrone, & Alais, 2014; Salomon, Lim, Herbelin, Hesselmann, & Blanke, 2013; Zhou, Jiang, He, & Chen, 2010; Zhou, Zhang, Chen, Wang, & Chen, 2012) were quite convincing, none provided a test against probability summation. Even if in some cases such a measurement was not necessary, since findings followed anatomical organizations (i.e., the tactile stimulation on vision was fine-tuned to one octave in orientation (Lunghi & Alais, 2013) or the olfactory stimulus to one nostril had effect precisely on the visual stimulus presented to the corresponding lateralized visual hemifield (Zhou et al., 2012) leading to conclude on bottom-up multisensory integration, in the other cases though, it was not necessarily so evident.

To the best of our knowledge, our studies are the first to address the evaluation of cross-modal responses below awareness using the probability summation model and, to look at the ERP correlates. On the one hand, for calculating probability or ERP summation, the component stimuli of the cross-modal pair need to be measured independently, as well as paired (e.g., to calculate PSM the probability of switch after each sensory stimulus needs to be measured independently (Otto & Mamassian, 2012). For instance, in Lunghi et al's study (2010) the probability of perceptual switch after the binocular presentation of the gratings can be measured alone, however, the probability of switch after touch always involves the probability of switch related to the visual gratings, therefore the two accounts are not measured independently. In fact, in many of the studies where cross-modal facilitation was found below awareness, this constrain (i.e., independently measurable single modalities) was not met, impeding a test for probability summation (Salomon, Kaliuzhna, Herbelin, & Blanke, 2015; Salomon et al., 2013; Zhou et al., 2010). Thus, the alternative explanation based on the independent contribution of each sensory stimulus was left unresolved, or at least remained untested. In the following sections, I will discuss the putative constraints of bottom-up multisensory integration.

## *3.3. The multisensory effect below the level of awareness*

### *3.3.1. Is cross-modal spatio-temporal alignment insufficient to produce bottom-up multisensory integration?*

So far, the picture is that multisensory integration fails to occur upon the co-occurrence of abrupt simple events in two different modalities, but it can nevertheless happen in cases of cross-modal congruency between more complex features (body posture, spatial frequency, …). One might ask, then, what stands behind these past findings conclusively showing multisensory integration below awareness? The main concept of our studies was to rule out top-down selective attention between stimuli, thus we used abrupt, uninformative and unpredictive, low-level stimuli pairs what were confined to spatio-temporal congruency. Instead, in other studies (Lunghi & Alais, 2015; Lunghi et al., 2010, 2014; Lunghi, Verde, & Alais, 2017; Lunghi & Morrone, 2013;

Salomon et al., 2015, 2013, Zhou et al., 2010, 2012) the stimuli presentation was long-lasting (sustaining for several seconds in some cases) and cross-modal stimuli could bear some form of congruency between cross-modal features beyond pure spatio-temporal co-occurrence.

Then, is just spatio-temporal alignment insufficient for such an integrative effect? Conrad et al. (2010) found cross-modal effect (not completely clear whether based on an interference under dominance or congruency under suppression, though, see Chapter 1, Section 1.4.4.4.) between auditory motion and unaware visual motion stimuli presented for a relatively long period of time in their BR study. Interestingly, abrupt spatio-temporal alignment did not provide enough bases for cross-modal effect, whilst long-lasting presentation of directionally congruent motion did. From this finding, one can conclude that either longer presentation or congruency between cross-modal features might be relevant for the cross-modal modulation in BR, I will discuss this consideration in the following sections, respectively.

### *3.3.2. The role of the time of presentation*

From the previously mentioned Conrad et al.' study we learned that the cross-modal effect was limited to long-lasting, sustained presentation of congruent cross-modal stimuli, as it was the case in the other above-mentioned studies with successful reports of cross-modal integration below awareness (Lunghi & Alais, 2015; Lunghi et al., 2010, 2014, 2017; Lunghi & Morrone, 2013; Salomon et al., 2015, 2013, Zhou et al., 2010, 2012). Indeed, a longer presentation of the cross-modal stimuli with congruent features might give rise to build-up on, or at least reinforce, some cross-modal perceptual associations between stimuli. This build-up could well happen thanks to the opportunity for the information in one sensory modality to induce the activation of perceptual properties of the unaware visual stimulus which, over time, might happen to coincide. It might happen due to predictable time, location and sensory features (considering Zhou et al., 2010), not missing out the possibility of mental imagery (Pearson, Clifford, & Tong, 2008) thus, achieving priority in the suppressed percepts and, producing quicker emergence to awareness (Conrad et al., 2013). These kinds of processes are more likely for the longer presentation of cross-modal stimulation. Is it possible that the cross-modal congruency related to the long-lasting presentation of the stimuli thus, might affect or even constrains the cross-modal effect below awareness?

### *3.3.3. The nature of cross-modal congruency, is it semantics?*

Can the multisensory integration effect presented in literature be attributed to high-level processes such as semantics? By semantics we refer to relationships between meaning thus, a conceptual relation between words, signs, symbols and what they stand for.

The study by Zhou et al. (2010), already discussed before, showed that cross-modal modulation below awareness can happen between olfaction and visual stimuli based on perceptual congruency (Chapter 1., Section 1.4.4.1. *Box 5*), but not based only on semantic relations. Similarly, in the study of Chen, Yeh and Spence (2010) the semantics by itself did not reveal in cross-modal modulation (i.e., the spoken word of animals did not have the effect like animal sounds), albeit in this case the cross-modal effect was restricted to the dominant percept. Hence, semantics may play a role, but it does not seem to be enough to explain cross-modal modulation below the level of awareness.

### *3.3.4. The nature of cross-modal congruency, is it cross-modal association?*

If cross-modal modulations below the level of awareness cannot be solely explained by semantics, are they based on cross-modal associations? By the cross-modal associations we refer to cases when a feature in one stimulus modality is associated to a feature of another modality, caused by the learning over correlations, incidental or explicit. Hence, this associations are not strictly semantic (e.g., association of bright colors and a high-pitch tone), and are not necessarily natural, since they can be arbitrarily acquired by exposure. Einhauser, Methfessel and Bendixen (2017) actually tested such newly acquired audiovisual associations (i.e., during a 20-min-session before the experiment) in a BR protocol.

The cross-modal effect from sounds to their associated suppressed visual images did not appear below awareness, although the newly acquired audiovisual associations did bias perception under dominance. This result is in line with top-down attentional effects (Chong, Tadin, & Blake, 2005; van Ee, van Boxtel, Parker, & Alais, 2009). Thus, seemingly the cross-modal integration effects below awareness cannot be attributed to semantics or quickly learned cross-modal associations.

### *3.3.5. The nature of cross-modal congruency, is it perceptual congruency?*

One further candidate to explain bottom-up multisensory integration below the level of awareness is 'perceptual congruency', a concept introduced by Chen et al. (2011), and used by many others labelling it in another name (discussed in the introduction of this thesis, see Chapter 1., Section 1.4.4.1., *Box 5*). Perceptual congruency derives from congruent sensory features (e.g., visual and acoustic, tactile or olfactory attributes, …), which can be dissociate from cross-modal associations, where the connection is arbitrary learned. For example, the sound and the image of a bird is perceptually congruent, while the sound of a bird can be associated with red circle due to a learning of the arbitrary connection. In fact, the results of Lunghi et al., (2010, 2014, 2017), Lunghi and Morrone (2013), Zhou et al. (2010) as well as Salomon et al. (2015, 2013) may well be categorized in this class, giving cross-modal effect below the level of awareness based on orientation and/or spatial frequency for touch and vision (Lunghi et al., 2010, 2017; Lunghi & Morrone, 2013), amplitude modulation depth for audition, touch and vision (Lunghi et al., 2014), position of a participants' hand (Salomon et al., 2013), vestibular information (Salomon et al., 2015), or bottom-up processing of olfactory and visual features (Zhou et al., 2010). According to a possible suggestion, the perceptual congruency between stimuli might gain relevance if any or both of the stimuli are relatively ambiguous, therefore congruent information might help to unify different sensory inputs to compensate for ambiguity (Yang & Yeh, 2014). However, please see the study of Moors et al. (2015) for contradictory results. In Moors et al.'s study despite spatio-temporal alignment and cross-modal perceptually congruency (i.e., looming audio and visual stimuli), there was no modulation below awareness. Still, one negative result may not suffice for the falsification of an assumption when many positive results are available in the same line.

### *3.3.6. Is previous conscious exposure of the cross-modal stimuli pairs a prerequisite for multisensory integration below awareness?*

The majority of findings suggest that the perceptual congruency seems to be promising candidate requisite for bottom-up multisensory integration, above and beyond spatio-temporal alignment. Albeit, clear constraints for unaware multisensory integration are still not established. However, talking about sensory processes below the level of awareness, need a brief enumeration of some consciousness theories. As I mentioned already in the introduction section about awareness (Chapter 1., Section 1.3.3., *Box 3*), according to the global workspace model of awareness, a certain information becomes conscious when the representation of this information becomes globally available in the brain via the fronto-parietal system (Baars, 1988; Dehaene & Naccache, 2001), which facilitates access between otherwise independent brain functions. In addition, other theories state that visual consciousness emerges via recurrent interactions that enable information exchange across multiple levels of the cortical hierarchy (perhaps already on visual areas) (Lamme, 2006; Lamme & Roelfsema, 2000). Beyond the difficulty of localization and conceptualization of consciousness, a new theory particularly links the concept of consciousness with the integration of information. Tononi (2012, 2004, 2008) proposed that consciousness actually corresponds to the capacity of a system to integrate information. Aligning with Tononi's proposal, Mudrik et al. (2014) postulated the integration information theory in multisensory contexts. Mudrik et al. suggested that cross-modal stimuli must have been consciously experienced (perceived) beforehand in order to integrate below awareness. The theory also proposes that the more complex the stimuli the bigger the necessity such previous conscious exposure. However, as we learned from the study of Einhauser et al. (2017), conscious experience of arbitrary cross-modal associations did not provide enough bases for unaware multisensory processing later on, suggesting that perhaps cross-modal stimuli must be not just previously perceived consciously but also perceptually congruent (beyond mere spatio-temporal alignment). According to the theory of Mudrik (2014), once the association between the congruent cross-modal stimuli is established in awareness, this association can operate even when stimuli are unaware or with less degree of awareness. Hence, one might speculate that whenever the stimuli are task-relevant and, presented for a relatively long period of time, the build-up of cross-modal associations previously established in awareness might be promoted, giving rise to unaware integration. When stimuli are task-irrelevant, mutually uninformative and, presented at unexpected moments and for brief periods of time, the build-up of such cross-modal associations would be ineffective under unaware conditions, even if they had been well established under prior conscious experience. Clearly, the cross-modal events tested in this thesis fall in the latter class.

Nevertheless, I believe that our results, even if they indeed bear a clear relationship to consciousness, are not finally conclusive as per consciousness theories. It is worth noting that in the thesis, perceptual consciousness was used as a manner to understand the depth of processing afforded by integration across sensory modalities, but it was not the goal of our scientific enquiry. Additionally, as it was mentioned in the introduction (Chapter 1., Section 1.3.3., *Box 3*), one might want to be cautious about the interpretations related to consciousness theories, as our findings putatively relate only to perceptual awareness, not reaching the level of subjective awareness (Giles, Lau, & Odegaard, 2016). A representative example of this distinction is the evidence by Zou, He and Zhang (2016), who demonstrated that invisible stimulus features could induce binocular rivalry indicating the phenomenon might be caused by differences in perceptual signal strength rather than conscious selection processes (i.e., the phenomenon putatively operated on the level of perceptual awareness without reaching the level of subjective awareness). According to this divergence, the perceptual awareness constitutes the visual system's ability to process stimuli in order to perform a visual task without necessarily reaching the level of subjective awareness where the visual system is able to generate a subjective conscious experience. Keeping in mind this distinction, and the limitations of our manipulations (that operate at the level of perceptual awareness), one cannot be conclusive about whether our findings can claim anything pro or contra any of the current consciousness theories.

## *3.4. Implication of the results*

### *3.4.1. Practical relevance of bottom-up multisensory integration*

Referring to the core question of the thesis, what is the relevance of bottom-up multisensory integration? Thinking of practical consequences, in emergency situations for example, we would greatly benefit from such an automatic and pre-attentive integration of sensory information, when instant reaction is needed independently of other cognitive mechanisms (likewise attention, expectation, imagination, memory retrieval, etc.), which might be eventually summoned to goal-directed behavior (Santangelo & Spence, 2007). Such principles have been suggested to provide good basis for alerting systems for drivers, for example (Ho, Gray, & Spence, 2014). What is more, bottom-up multisensory integration could provide basis for rehabilitation in states of reduced awareness, for not completely intact sensory processes (i.e., unilaterally or bilaterally reduced responsiveness) (Bolognini, Rasi, Coccia, & Ladavas, 2005; Làdavas, 2008; Lewald, Tegenthoff, Peters, & Hausmann, 2012), or even to include plasticity (Pascual-Leone, Amedi, Fregni, & Merabet, 2005). So, the query of to what extent cross-modal stimuli integrate in an automatic bottom-up manner (i.e., without the influence of attention, expectation, imagery or other cognitive mechanisms) might be a relevant question.

### *3.4.2. Are the results of this thesis applicable?*

Even if our results do not suggest multisensory integration, are they still applicable? In clinical studies, a logic and stimuli similar to our studies (i.e., abrupt audiovisual stimulus presentation) has been used to propose rehabilitation techniques based on multisensory integration for patients with visual hemineglect and or with hemianopia (Bolognini, Rasi, Coccia, & Ládavas, 2005; Frassinetti et al., 2005; Bolognini, Rasi, & Ládavas, 2005; Làdavas 2008). On the one hand, in hemianopia the visual sensory processes are defective that is resulting from the unilateral postchiasmatic damage, which determines the loss of vision in the hemifield that corresponds retinotopically to the damaged area (Zihl & Kennard, 2003), although beyond the persistent cortical blindness in some cases the patients may exhibit residual conscious vision or reflective responses to not consciously perceived stimuli (Stoerig & Cowey, 1997). On the other hand, in the hemispatial neglect, the patients suffer from an attentional deficit thus, usually fail to report, respond or orient to visual stimuli presented contralaterally to the lesioned hemisphere (Vallar, Halligan, Fink, Marshall, & Vallar, 2003). The above-mentioned studies of Bolognini et al. and Frassinetti et al. showed that visual information flashed in the affected hemifield of the patient could nevertheless significantly improve their auditory localization performance, despite the patients remained unaware of the presence of the visual stimulus. The beneficial effect of cross-modal stimulation thus, has been attributed to bottom-up multisensory integration of an unaware visual event and an aware sound. According to our results, one might ask though, to what extent Bolognini et al. and Frassinetti et al. explanation (i.e. assuming bottom-up multisensory integration) is the most parsimonious as opposed to individual contribution of stimuli calling for bottom-up attention. The bottom-up attentional capture entirely or at least partially might take place in the cross-modal effect in both of the clinical populations. In hemianopia, a therapy called 'visual restoration therapy' (Kasten, Wüst, Behrens-Baumann, & Sabel, 1998) already capitalized on the residual visual sensory processes and uses repeatedly presented visual flash on the affected hemifield. Besides, related to the visual neglect, the bottom-up attentional capture might be also possible in visual search task (Mangano et al., 2014; Pinto, Leij, Sligte, Lamme, & Scholte, 2013), hence regardless of the top-down attentional deficit, bottom-up attention might not be necessarily affected. Thus, in the above mentioned clinical studies (even if the ventriloquist effect is quite different from BR) as well as in related therapeutic approaches, the bottom-up attentional account might be also considered as a feasible principle, beyond and besides bottom-up multisensory integration.

## *3.5. Conclusions*

While it is widely assumed that we are not consciously aware of most of the cross-modal processing going on during everyday life perception, to date it was not known whether cross-modal integrative processes, in the sense of convergence of information, may occur based on purely bottom-up sensory processes. One can attempt to produce cross-modal effects when one of the sensory modalities is presented below awareness (as in binocular rivalry protocols). Yet, in order to single out these putative bottom-up integration processes across modalities one would have to control or rule out top-down processing, such as that of selective endogenous attention, expectation and semantically guided anticipation across modalities. If these are properly controlled, one can try and reveal whether cross-modal integrative phenomena still stand. We presented brief unpredictable and uninformative task-irrelevant audio stimuli, to measure their influence on unaware visual stimuli. Based on the results of our studies we concluded that spatio-temporal coincidence of abrupt simple stimuli putatively did not provide basis for bottom-up multisensory integration. This result might suggest such the phenomenon of unconscious cross-modal integration in the absence of top-down selective attention might be constrained to cross-modal perceptual congruency amongst complex perceptual features when the stimuli are available for an extended period of time, or else, operate only when top-down influences are available. Another conclusion of this thesis is that, even in the absence of integration, the independent contribution of stimuli in each sensory modality might be also sufficient to achieve improved (fast and precise) reactions, based on bottom-up attentional capture. Even in clinical cases, where putative bottom-up integration was proposed as a rehabilitation strategy, such an interpretation might be also give the basis for an alternative interpretation of the cross-modal rehabilitation principles.

## Bibliography

Adam, R., & Noppeney, U. (2014). A phonologically congruent sound boosts a visual target into perceptual awareness. *Frontiers in Integrative Neuroscience*, *8*(September), 1–13. https://doi.org/10.3389/fnint.2014.00070

Alais, D., & Blake, R. (2013). Binocular rivalry and perceptual ambiguity. *Oxford Handbook and Perceptual Organization*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Alais, D., Cass, J., O’Shea, R. P., & Blake, R. (2010). Visual sensitivity underlying changes in visual consciousness. *Current Biology*, *20*(15), 1362–1367. https://doi.org/10.1016/j.cub.2010.06.015

Alais, D., van Boxtel, J. J., Parker, A., & van Ee, R. (2010). Attending to auditory signals slows visual alternations in binocular rivalry. *Vision Research*, *50*(10), 929–35. https://doi.org/10.1016/j.visres.2010.03.010

Aller, M., Giani, A., Conrad, V., Watanabe, M., & Noppeney, U. (2015). A spatially collocated sound thrusts a flash into awareness. *Frontiers in Integrative Neuroscience*, *9*(2), 1–8. https://doi.org/10.3389/fnint.2015.00016

Alsius, A., & Munhall, K. G. (2013). Detection of Audiovisual Speech Correspondences Without Visual Awareness. *Psychological Science*, *24*(4), 423–431. https://doi.org/10.1177/0956797612457378

Andersen, T. S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision Research*, *48*(25), 2537–2544.

https://doi.org/10.1016/j.visres.2008.08.018

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. *A Cognitive Theory of Consciousness*. Retrieved from http://www.loc.gov/catdir/description/cam032/87020923.html

Bachmann, T. (1994). *Psychophysiology of visual masking : the fine structure of conscious experience*. Nova Science Publishers.

Bahrami, B., Vetter, P., Spolaore, E., Pagano, S., Butterworth, B., & Rees, G. (2010). Unconscious Numerical Priming Despite Interocular Suppression. *Psychological Science*, *21*(2), 224–233. https://doi.org/10.1177/0956797609360664

Bar, M., Tootell, R. B. H., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., … Dale, a M. (2001). Cortical mechanisms specific to explicit object recognition. *Neuron*, *29*(2), 529–535.

Barbot, A., & Kouider, S. (2012). Longer is not better: nonconscious overstimulation reverses priming influences under interocular suppression. *Attention, Perception, & Psychophysics*, *74*(1), 174–184. https://doi.org/10.3758/s13414-011-0226-3

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. *Unintended Thought*.

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & De Gelder, B. (2000). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia*, *38*(12), 1634–1642. https://doi.org/10.1016/S0028-3932(00)00067-1

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*(2), 321–332. https://doi.org/10.3758/BF03205552

Blake, R. (1989). A neural theory of binocular rivalry. *Psychological Review*, *96*(1), 145–67. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2648445

Blake, R. (2001). A Primer on Binocular Rivalry, Including Current Controversies. *Brain and Mind*, *2*(1), 5–38. https://doi.org/10.1023/A:1017925416289

Blake, R., & Boothroyd, K. (1985). The precedence of binocular fusion over binocular rivalry. *Perception & Psychophysics*, *37*(2), 114–124. https://doi.org/10.3758/BF03202845

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews. Neuroscience*, *3*(1), 13–21. https://doi.org/10.1038/nrn701

Blake, R., Westendorf, D., & Fox, R. (1990). Temporal perturbations of binocular rivalry. *Perception & Psychophysics*, *48*(6), 593–602. https://doi.org/10.3758/BF03211605

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). “Acoustical vision” of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*(3), 273–282. https://doi.org/10.1007/s00221-004-2005-z

Bolognini, N., Rasi, F., Coccia, M., & Ladavas, E. (2005). Visual search improvement in hemianopic patients after audio-visual stimulation. *Brain*, *128*(12), 2830–2842. https://doi.org/10.1093/brain/awh656

Bolognini, N., Rasi, F., & Ládavas, E. (2005a). Visual localization of sounds. *Neuropsychologia*, *43*(11), 1655–1661. https://doi.org/10.1016/j.neuropsychologia.2005.01.015

Bolognini, N., Rasi, F., & Ládavas, E. (2005b). Visual localization of sounds. *Neuropsychologia*, *43*(11), 1655–1661. https://doi.org/10.1016/j.neuropsychologia.2005.01.015

Bonneh, Y. S., Cooperman, A., & Sagi, D. (2001). Motion-induced blindness in normal observers. *Nature*, *411*(6839), 798–801. https://doi.org/10.1038/35081073

Boynton, G. M. (2009). A framework for describing the effects of attention on visual responses. *Vision Research*, *49*(10), 1129–1143. https://doi.org/10.1016/j.visres.2008.11.001

Brascamp, J. W., & Blake, R. (2012). Inattention Abolishes Binocular Rivalry: Perceptual Evidence. *Psychological Science*, *23*(10), 1159–1167. https://doi.org/10.1177/0956797612440100

Bridgeman, B. (1971). Metacontrast and lateral inhibition. *Psychological Review*, *78*(6), 528–39. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5122073

Britz, J., & Pitts, M. A. (2011). Perceptual reversals during binocular rivalry: ERP components and their concomitant source differences. *Psychophysiology*, *48*(11), 1490–1499. https://doi.org/10.1111/j.1469-8986.2011.01222.x

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology*, *98*(1–3), 191–205. https://doi.org/10.1016/j.jphysparis.2004.03.018

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-Visual Multisensory Interactions in Humans: Timing, Topography, Directionality, and Sources. *Journal of Neuroscience*, *30*(38), 12572–12580. https://doi.org/10.1523/JNEUROSCI.1099-10.2010

Carrasco, M., Fuller, S., & Ling, S. (2008). Transient attention does increase perceived contrast of suprathreshold stimuli: a reply to Prinzmetal, Long, and Leonhardt (2008). *Perception &*

*Psychophysics*, *70*(7), 1151–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18979688

Carter, O., Konkle, T., Wang, Q., Hayward, V., & Moore, C. (2008). Tactile rivalry demonstrated with an ambiguous apparent-motion quartet. *Current Biology*, *18*(14), 1050–4. https://doi.org/10.1016/j.cub.2008.06.027

Chen, Y.-C., Yeh, S. L., & Spence, C. (2010). Crossmodal constraints on human visual awareness: Auditory semantic context modulates binocular rivalry. *Journal of Vision*, *10*(7), 885–885. https://doi.org/10.1167/10.7.885

Chen, Y. C., & Yeh, S. L. (2008). Visual events modulated by sound in repetition blindness. *Psychonomic Bulletin & Review*, *15*(2), 404–408. https://doi.org/10.3758/PBR.15.2.404

Chen, Y. C., Yeh, S. L., & Spence, C. (2011). Crossmodal constraints on human perceptual awareness: Auditory semantic modulation of binocular rivalry. *Frontiers in Psychology*, *2*(SEP), 1–13. https://doi.org/10.3389/fpsyg.2011.00212

Chica, A. B., & Lupianez, J. (2009). Effects of endogenous and exogenous attention on visual processing: An Inhibition of Return study . *Brain Res*, *1278*, 75–85. https://doi.org/10.1016/j.brainres.2009.04.011

Chong, S. C., Tadin, D., & Blake, R. (2005). Endogenous attention prolongs dominance durations in binocular rivalry. *Journal of Vision*, *5*(11), 1004–1012. https://doi.org/10.1167/5.11.6

Chun, M. M. (1997). Temporal binding errors are redistributed by the attentional blink. *Perception & Psychophysics*, *59*(8), 1191–1199. https://doi.org/10.3758/BF03214207

Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (2005). A Critique of Pure Vision. *In: C. Koch and J. Davis (Eds.), Large-Scale Neuronal Theories of the Brain*, 1–25. https://doi.org/10.1207/S15326969ECO1502_5

Clark, V. P., & Hillyard, S. a. (1996). Spatial Selective Attention Affects Early Extrastriate But Not Striate Components of the Visual Evoked Potential. *Journal of Cognitive Neuroscience*, *8*(5), 387–402. https://doi.org/10.1162/jocn.1996.8.5.387

Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision*, *10*(10), 27. https://doi.org/10.1167/10.10.27

Conrad, V., Kleiner, M., Bartels, A., Hartcher O'Brien, J., Bülthoff, H. H., & Noppeney, U. (2013). Naturalistic Stimulus Structure Determines the Integration of Audiovisual Looming Signals in Binocular Rivalry. *PLoS ONE*, *8*(8).

https://doi.org/10.1371/journal.pone.0070710

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 215–229. https://doi.org/10.1038/nrn755

Coull, J. T., & Nobre,  a C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *J Neurosci*, *18*(18), 7426–35. https://doi.org/0270-6474/98/187426-10$05.00/0

Czigler, I., Balázs, L., & Winkler, I. (2002). Memory-based detection of task-irrelevant visual changes. *Psychophysiology*, *39*(6), 869–73. https://doi.org/10.1017/S0048577202020218

Dalton, P., & Lavie, N. (2006). Temporal attentional capture: effects of irrelevant singletons on rapid serial visual search. *Psychonomic Bulletin & Review*, *13*(5), 881–885. https://doi.org/10.3758/BF03194013

De Meo, R., Murray, M. M., Clarke, S., Matusz, P. J., Soto-Faraco, S., & Wallace, M. T. (2015). Top-down control and early multisensory processes: Chicken vs. egg. *Frontiers in Integrative Neuroscience*, *9*(3), 1–6. https://doi.org/10.3389/fnint.2015.00017

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211. https://doi.org/10.1016/j.tics.2006.03.007

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(24), 14529–14534. https://doi.org/10.1073/pnas.95.24.14529

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness:basic evidence. *Cognition*, *79*(79), 1–37.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc Natl Acad Sci U S A*, *100*(14), 8520–8525. https://doi.org/10.1073/pnas.1332574100

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dieter, K. C., Melnick, M. D., & Tadin, D. (2015). When can attention influence binocular rivalry? *Attention, Perception & Psychophysics*, *77*(6), 1908–18. https://doi.org/10.3758/s13414-015-0905-6

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on “sensory-specific” brain regions, neural responses, and judgments. *Neuron*, *57*(1), 11–23. https://doi.org/10.1016/j.neuron.2007.12.013

Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, *10*(20), R731–R735. https://doi.org/10.1016/S0960-9822(00)00740-5

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Egeth, H. E., Virzi, R. a, & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(1), 32–39. https://doi.org/10.1037/0096-1523.10.1.32

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology.*, *123*(2), 161–77. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8014611

Einhäuser, W., Methfessel, P., & Bendixen, A. (2017). Newly acquired audio-visual associations bias perception in binocular rivalry. *Vision Research*, *133*, 121–129. https://doi.org/10.1016/j.visres.2017.02.001

Evans, E. F., & Whitfield, I. C. (1964). Classifcation of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat. *The Journal of Physiology*, *171*, 476–93. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1368845&tool=pmcentrez&rendertype=abstract

Faivre, N., Berthet, V., & Kouider, S. (2012). Nonconscious influences from emotional faces: A comparison of visual crowding, masking, and continuous flash suppression. *Frontiers in Psychology*, *3*(MAY), 1–13. https://doi.org/10.3389/fpsyg.2012.00129

Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory Integration in Complete Unawareness : Evidence From Audiovisual Congruency Priming. *Psychological Science*, *25*(11), 2006–2016. https://doi.org/10.1177/0956797614547916

Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory Integration in Complete Unawareness: Evidence From Audiovisual Congruency Priming. *Psychological Science*, *25*(11), 2006–2016. https://doi.org/10.1177/0956797614547916

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *22*(13), 5749–59. https://doi.org/20026562

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology. Human Perception and Performance*, *18*(4), 1030–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1431742

Fort, A., Delpuech, C., Pernier, J., & Giard, M.-H. (2002a). Dynamics of Cortico-subcortical Cross-modal Operations Involved in Audio-visual Object Detection in Humans. *Cerebral Cortex*, *12*(10), 1031–1039. https://doi.org/10.1093/cercor/12.10.1031

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

Fox, R., & Herrmann, J. (1967). Stochastic properties of binocular rivalry alternations. *Perception & Psychophysics*, *2*(9), 432–436. https://doi.org/10.3758/BF03208783

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, *16*(0959–4965 (Print)), 419–423. https://doi.org/10.1097/00001756-200504040-00001

Francis, G. (1997). Cortical dynamics of lateral inhibition: metacontrast masking. *Psychological Review*, *104*(3), 572–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9243965

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., & Làdavas, E. (2005). Audiovisual integration in patients with visual deficit. *Journal of Cognitive Neuroscience*, *17*(9), 1442–1452. https://doi.org/10.1162/0898929054985446

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332–343. https://doi.org/10.1007/s00221-002-1262-y

Frassinetti, F., Pavani, F., & Làdavas, E. (2002). Acoustical vision of neglected stimuli: interaction among spatially converging audiovisual inputs in neglect patients. *Journal of Cognitive Neuroscience*, *14*(1), 62–9. https://doi.org/10.1162/089892902317205320

Giard, M. H., Fort, A., Mouchetant-Rostaing, Y., & Pernier, J. (2000). Neurophysiological mechanisms of auditory selective attention in humans. *Frontiers in Bioscience : A Journal and Virtual Library*, *5*, D84-94. https://doi.org/10.2741/Giard

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci*, *11*(5), 473–490. https://doi.org/10.1162/089892999563544

Giles, N., Lau, H., & Odegaard, B. (2016a). What Type of Awareness Does Binocular Rivalry Assess? *Trends in Cognitive Sciences*, *20*(10), 719–720. https://doi.org/10.1016/j.tics.2016.08.010

Giles, N., Lau, H., & Odegaard, B. (2016b). What Type of Awareness Does Binocular Rivalry Assess? *Trends in Cognitive Sciences*, *20*(10), 719–720. https://doi.org/10.1016/j.tics.2016.08.010

Grill-spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*(8), 837–843. https://doi.org/10.1038/77754

Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Mossbridge, J. (2012). Report Interactive Coding of Visual Spatial Frequency and Auditory Amplitude-Modulation Rate. *Current Biology*, *22*(5), 383–388. https://doi.org/10.1016/j.cub.2012.01.004

Haenny, P. E., Maunsell, J. H., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Experimental Brain Research*, *69*(2), 245–259. https://doi.org/10.1007/BF00247570

Hairston, W. D., Laurienti, P. J., Mishra, G., Burdette, J. H., & Wallace, M. T. (2003). Multisensory enhancement of localization under conditions of induced myopia. *Experimental Brain Research*, *152*(3), 404–8. https://doi.org/10.1007/s00221-003-1646-7

Hancock, S., & Andrews, T. J. (2007). The role of voluntary and involuntary attention in selecting perceptual dominance during binocular rivalry. *Perception*, *36*(2), 288–298. https://doi.org/10.1068/p5494

Hartcher-O, B. J., Soto-Faraco, S., & Adam, R. (2017). Editorial : A Matter of Bottom-Up or Top-Down Processes : The Role of Attention in Multisensory Integration, *11*(February), 10–12. https://doi.org/10.3389/fnint.2017.00005

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *The Journal of Experimental Psychology*, *63*(3), 289–293. https://doi.org/10.1037/h0055703

Ho, C., Gray, R., & Spence, C. (2014). Reorienting Driver Attention with Dynamic Tactile Cues. *IEEE Transactions on Haptics*, *7*(1), 86–94. https://doi.org/10.1109/TOH.2013.62

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Neuropsychologia*, *45*(14), 3340–3345. https://doi.org/10.1016/j.neuropsychologia.2007.05.025

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*, 106–54. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1359523&tool=pmcentrez&rendertype=abstract

Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception & Psychophysics*, *72*(7), 1736–1741. https://doi.org/10.3758/APP.72.7.1736

Jaekl, P., Pérez-Bellido, A., & Soto-Faraco, S. (2014). On the "visual" in "audio-visual integration": A hypothesis concerning visual pathways. *Experimental Brain Research*, *232*(6), 1631–1638. https://doi.org/10.1007/s00221-014-3927-8

James, W. (1890). The Principles Of Psychology. In *London: MacMillan*. Retrieved from https://books.google.com/books?hl=hu&lr=&id=sah8CgAAQBAJ&pgis=1

Kanai, R., Moradi, F., Shimojo, S., & Verstraten, F. A. J. (2005). Perceptual alternation induced by visual transients. *Perception*, *34*(7), 803–822. https://doi.org/10.1068/p5245

Kang, M.-S., & Blake, R. (2005). Perceptual synergy between seeing and hearing revealed during binocular rivalry. *Psichologija*, *32*, 7–15.

Kanwisher, N. G. (1987). Repetition blindness: type recognition without token individuation. *Cognition*, *27*(2), 117–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3691023

Kasten, E., Wüst, S., Behrens-Baumann, W., & Sabel, B. a. (1998). Computer-based training for the treatment of partial blindness. *Nature Medicine*, *4*(9), 1083–1087. https://doi.org/10.1038/2079

Khoe, W., Mitchell, J. F., Reynolds, J. H., & Hillyard, S. a. (2008). ERP evidence that surface-based attention biases interocular competition during rivalry. *Journal of Vision*, *8*(3), 18.1-11. https://doi.org/10.1167/8.3.18

Kim, Y.-J., Grabowecky, M., & Suzuki, S. (2006). Stochastic resonance in binocular rivalry. *Vision Research*, *46*(3), 392–406. https://doi.org/10.1016/j.visres.2005.08.009

King, J.-R., Pescetelli, N., & Dehaene, S. (2016). Selective maintenance mechanisms of seen and unseen sensory features in the human brain. *BioRxiv.org*, 1–30.

https://doi.org/http://dx.doi.org/10.1101/040030

Kingstone, A. (1992). Combining Expectancies. *The Quarterly Journal of Experimental Psychology*, *44*(September), 69–104. https://doi.org/10.1080/14640749208401284

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews*. https://doi.org/10.1016/j.brainresrev.2006.06.003

Koch, C., & Tsuchiya, N. (2006). Attention and consciousness : two distinct brain processes, *11*(1). https://doi.org/10.1016/j.tics.2006.10.012

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372–84. https://doi.org/10.1016/j.actpsy.2010.03.010

Kouider, S., Stahlhut, C., Gelskov, S. V, Barbosa, L. S., Dutat, M., de Gardelle, V., … Dehaene-Lambertz, G. (2013). A neural marker of perceptual consciousness in infants. *Science*, *340*, 376–80. https://doi.org/10.1126/science.1232509

Kovacs, I., Papathomas, T. V., Yang, M., & Feher, A. (1996). When the brain changes its mind: Interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences*, *93*(26), 15508–15511. https://doi.org/10.1073/pnas.93.26.15508

Lack, L. C. (1974). Selective attention and the control of binocular rivalry. *Perception & Psychophysics*. https://doi.org/10.3758/BF03205846

Làdavas, E. (2008). Multisensory-based approach to the recovery of unisensory deficit. *Annals of the New York Academy of Sciences*, *1124*, 98–110. https://doi.org/10.1196/annals.1440.008

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501. https://doi.org/10.1016/j.tics.2006.09.001

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*. https://doi.org/10.1016/S0166-2236(00)01657-X

Lankheet, M. J. M. (2006). Unraveling adaptation and mutual inhibition in perceptual rivalry. *Journal of Vision*, *6*(4), 304–10. https://doi.org/10.1167/6.4.1

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, *379*(6565), 549–53. https://doi.org/10.1038/379549a0

Levelt, W. J. M. (1965). On Binocular Rivalry. https://doi.org/0.4249/scholarpedia.1578

Lewald, J., Tegenthoff, M., Peters, S., & Hausmann, M. (2012). Passive auditory stimulation improves vision in hemianopia. *PLoS ONE, 7*(5). https://doi.org/10.1371/journal.pone.0031603

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–601. https://doi.org/10.1073/pnas.092277599

Li, H.-H., Rankin, J., Rinzel, J., Carrasco, M., & Heeger, D. J. (2017). Attention model of binocular rivalry. *Proceedings of the National Academy of Sciences*, (July), 201620475. https://doi.org/10.1073/pnas.1620475114

Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research, 1173*(1), 102–109. https://doi.org/10.1016/j.brainres.2007.07.050

Liu, T., Larsson, J., & Carrasco, M. (2007). Feature-based attention modulates orientation-selective responses in human visual cortex. *Neuron*, *55*(2), 313–323. https://doi.org/10.1016/j.neuron.2007.06.030

Logan, G. D. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology*, *105*(2), 317–339.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research*, *17*(2), 447–453. https://doi.org/10.1016/S0926-6410(03)00160-5

Luck, S. J. (2004). Understanding awareness: one step closer. *Nature Neuroscience*, *7*(3), 208–9. https://doi.org/10.1038/nn0304-208

Lunghi, C., & Alais, D. (2013). Touch Interacts with Vision during Binocular Rivalry with a Tight Orientation Tuning. *PLoS ONE, 8*(3), 1–8. https://doi.org/10.1371/journal.pone.0058754

Lunghi, C., & Alais, D. (2015). Congruent tactile stimulation reduces the strength of visual suppression during binocular rivalry. *Scientific Reports*, *5*(1), 9413. https://doi.org/10.1038/srep09413

Lunghi, C., Binda, P., & Morrone, C. (2010). Touch disambiguates rivalrous perception at early stages of visual analysis. *Current Biology*, *20*(4), 143–144.

Lunghi, C., & Morrone, M. C. (2013). Early Interaction between Vision and Touch during. *Multisensory Research*, *26*(3), 291–306. https://doi.org/10.1163/22134808-00002411

Lunghi, C., Morrone, M. C., & Alais, D. (2014). Auditory and Tactile Signals Combine to Influence Vision during Binocular Rivalry.

*The Journal of Neuroscience*, *34*(3), 784–792. https://doi.org/10.1523/JNEUROSCI.2732-13.2014

Lunghi, C., Verde, L. Lo, & Alais, D. (2017). Touch accelerates visual awareness. *I-Perception*, *8*(1). https://doi.org/10.1177/2041669516686986

Macaluso, E., Eimer, M., Frith, C. D., & Driver, J. (2003). Preparatory states in crossmodal spatial attention: spatial specificity and possible control mechanisms. *Experimental Brain Research*, *149*(1), 62–74. https://doi.org/10.1007/s00221-002-1335-y

Macaluso, E., Frith, C. D., & Driver, J. (2002). Supramodal Effects of Covert Spatial Orienting Triggered by Visual or Tactile Events. *Journal of Cognitive Neuroscience*, *14*(3), 389–401. https://doi.org/10.1162/089892902317361912

Maki, W., & Mebane, M. (2006). Attentional capture triggers an attentional blink. *Psychonomic Bulletin & Review*, *13*(1), 125–131. https://doi.org/10.3758/bf03193823

Mangano, G. R., Oliveri, M., Turriziani, P., Smirni, D., Zhaoping, L., & Cipolotti, L. (2014). Impairments in top down attentional processes in right parietal patients: Paradoxical functional facilitation in visual search. *Vision Research*, *97*, 74–82. https://doi.org/10.1016/j.visres.2014.02.002

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 4–18. https://doi.org/10.1111/j.1469-8986.1995.tb03400.x

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. In *International Journal of Psychophysiology* (Vol. 50, pp. 125–145). https://doi.org/10.1016/S0167-8760(03)00129-6

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, *14*(9), 744–751. https://doi.org/10.1016/j.cub.2004.04.028

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To See or Not to See: Prestimulus Phase Predicts Visual Awareness. *Journal of Neuroscience*, *29*(9), 2725–2732. https://doi.org/10.1523/JNEUROSCI.3963-08.2009

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychon Bull Rev*, *18*(5), 904–909. https://doi.org/10.3758/s13423-011-0131-8

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2006.04.001

McDonald, J. J., Teder-Sälejärvi, W. a, & Hillyard, S. a. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*(6806), 906–908. https://doi.org/10.1038/35038085

McDonald, J. J., & Ward, L. M. (2000). Involuntary listening aids seeing: evidence from human electrophysiology. *Psychological Science*, *11*(2), 167–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11273425

McDougall, W. (1901). On the seat of the psychophysical processes. *Brain*, *24*(4), 579–630. https://doi.org/10.1093/brain/24.4.579

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Meredith, M. A., Allman, B. L., Keniston, L. P., & Clemo, H. R. (2012). Are Bimodal Neurons the Same throughout the Brain? In *The Neural Bases of Multisensory Processes*. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK92874/

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389–391. https://doi.org/10.1126/science.6867718

Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of muitisensory neurons in cat superior colliculus. *Brain Research*, *5*, 350–354.

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, *75*(5), 1843–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8734584

Metzger, B., Mathewson, K. E., Tapia, E., Fabiani, M., Gratton, G., & Beck, D. M. (2017). Regulating the Access to Awareness: Brain Activity Related to Probe-related and Spentaneous Reversals in Binocular Rivalry. *Journal of Cognitive Neuroscience*, 1–14. https://doi.org/10.1162/jocn

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002a). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. In *Cognitive Brain Research* (Vol. 14, pp. 115–128). https://doi.org/10.1016/S0926-6410(02)00066-6

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002b). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research, 14*(1), 115–

128. https://doi.org/10.1016/S0926-6410(02)00066-6

Moors, P., Hanne, H., Wagemans, J., De-Wit, L., & van Ee, R. (2015). Suppressed visual looming stimuli are not integrated with auditory looming signals: evidence from continuous flash suppression, *32*(0), 1–27.

Moran, R. J., & Reilly, R. B. (2006). Neural mass model of human multisensory integration. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 1,* 5559–62. https://doi.org/10.1109/IEMBS.2006.259588

Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology*, *98*(3), 1125–39. https://doi.org/10.1152/jn.00116.2007

Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat’s somatic sensory cortex. *Journal of Neurophysiology*, *20*(4), 408–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13439410

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: expanding the limits of unconscious processing. *Psychological Science, 22*(6), 764–770. https://doi.org/10.1177/0956797611408736

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2014.04.009

Munhall, K. G., Hove, M. . ten, Brammer, M., & Paré, M. (2009). Audiovisual integration of speech in a bistable illusion, *19*(9), 735–739. https://doi.org/10.1126/scisignal.2001449.Engineering

Murray, M. M., Foxe, J. J., & Wylie, G. R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *NeuroImage, 27*(2), 473–478. https://doi.org/10.1016/j.neuroimage.2005.04.016

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., … 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*(7), 963–974. https://doi.org/10.1093/cercor/bhh197

Ngo, M. K., & Spence, C. (2012). Facilitating masked visual target identification with auditory oddball stimuli. *Experimental Brain Research, 221*(2), 129–136. https://doi.org/10.1007/s00221-012-

3153-1

Nguyen, V. a, Freeman, a W., & Wenderoth, P. (2001). The depth and selectivity of suppression in binocular rivalry. *Perception & Psychophysics*, *63*(2), 348–360. https://doi.org/10.3758/BF03194475

Noel, J. P., Wallace, M., & Blake, R. (2015). Cognitive neuroscience: Integration of sight and sound outside of awareness? *Current Biology*, *25*(4), R157–R159. https://doi.org/10.1016/j.cub.2015.01.007

Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. *Perception & Psychophysics*, *65*(1), 123–132. https://doi.org/10.3758/BF03194789

Olivers, C. N. L., & Burg, E. Van der. (2008). Bleeping you out of the blink : Sound saves vision from oblivion. *Brain Research*, *1242*(2000), 191–199. https://doi.org/10.1016/j.brainres.2008.01.070

Ooi, T. L., & He, Z. J. (1999). Binocular rivalry and visual awareness: The role of attention. *Perception*, *28*(5), 551–574. https://doi.org/10.1068/p2923

Opitz, B., Mecklinger, A., Von Cramon, D. Y., & Kruggel, F. (1999). Combining electrophysiological and hemodynamic measures of the auditory oddball. *Psychophysiology*, *36*, 142–147. https://doi.org/10.1017/S0048577299980848

Orchard-Mills, E., Van der Burg, E., & Alais, D. (2013). Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies. *Journal of Vision*, *13*(3), 1–17. https://doi.org/10.1167/13.3.6

Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology*, *22*(15), 1391–1396. https://doi.org/10.1016/j.cub.2012.05.031

Paffen, C. L. E., Alais, D., & Verstraten, F. A. J. (2006). Attention speeds binocular rivalry. *Psychological Science*, *17*(9), 752–756. https://doi.org/10.1111/j.1467-9280.2006.01777.x

Paffen, C. L. E., & Van der Stigchel, S. (2010). Shifting spatial attention makes you flip : Exogenous visual attention triggers perceptual alternations during binocular rivalry. *Attention, Perception and Psychophysics*, *72*(5), 1237–1243. https://doi.org/10.3758/APP

Palmer, T. D., & Ramsey, A. K. (2012). The function of consciousness in multisensory integration. *Cognition*, *125*(3), 353–364. https://doi.org/10.1016/j.cognition.2012.08.003

Pannunzi, M., Pérez-Bellido, A., Pereda-Baños, A., López-Moliner, J., Deco, G., & Soto-Faraco, S. (2015). Deconstructing multisensory enhancement in detection. *Journal of Neurophysiology*, *113*(6), 1800–18. https://doi.org/10.1152/jn.00341.2014

Pápai, M. S., & Soto-Faraco, S. (2017). Sounds can boost the awareness of visual events through attention without cross-modal integration. *Scientific Reports*, *7*(January), 41684. https://doi.org/10.1038/srep41684

Parise, C. V, & Spence, C. (2013). Audiovisual Cross-Modal Correspondences in the General Population. *Oxford Handbook of Synesthesia*, (November), 790–815. https://doi.org/10.1093/oxfordhb/9780199603329.013.0039

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review of Neuroscience*, *28*, 377–401. https://doi.org/10.1146/annurev.neuro.27.070203.144216

Pastukhov, A., & Braun, J. (2007). Perceptual reversals need no prompting by attention. *Journal of Vision*, *7*(10), 1–17. https://doi.org/10.1167/7.10.5.Introduction

Pearson, J., Clifford, C. W. G., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology*, *18*(13), 982–6. https://doi.org/10.1016/j.cub.2008.05.048

Perez-Bellido, A., Soto-Faraco, S., & Lopez-Moliner, J. (2013). Sound-driven enhancement of vision: disentangling detection-level from decision-level contributions. *Journal of Neurophysiology*, *109*(4), 1065–1077. https://doi.org/10.1152/jn.00226.2012

Pesquita, A., Brennan, A. A., Enns, J. T., & Soto-Faraco, S. (2013). Isolating shape from semantics in haptic-visual priming. *Experimental Brain Research*, *227*(3), 311–322. https://doi.org/10.1007/s00221-013-3489-1

Pinto, Y., Leij, A. R. Van Der, Sligte, I. G., Lamme, V. A. F., & Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *Journal of Vision*, *13*(3), 1–14. https://doi.org/10.1167/13.3.16.doi

Plass, J., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2014). Lip reading without awareness. *Psychological Science*, *25*(9), 1835–1837. https://doi.org/10.1177/0956797614542132

Posner, M. I. (1980). Orienting of attention. *Q J Exp Psychol*, *32*(September), 3–25. https://doi.org/10.1080/00335558008248231

Prinzmetal, W., Park, S., & Garrett, R. (2005). Involuntary attention and identification accuracy. *Perception & Psychophysics*, *67*(8), 1344–1353. https://doi.org/10.3758/BF03193639

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037/0096-1523.18.3.849

Regan, D. (1989). Human brain electrophysiology. Evoked potentials and evoked magnetic fields in science and medicine. *Electroencephalography and Clinical Neurophysiology*, *73*(1), 84. https://doi.org/10.1016/0013-4694(89)90022-9

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics*, *51*(3), 279–290. https://doi.org/10.3758/BF03212254

Rinne, T., Degerman, A., & Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRI study. *NeuroImage*, *26*(1), 66–72. https://doi.org/10.1016/j.neuroimage.2005.01.017

Robertson, I. H., Mattingley, J. B., Rorden, C., & Driver, J. (1998). Phasic alerting of neglect patients overcomes their spatial deficit in visual awareness. *Nature*, *395*(6698), 169–172. https://doi.org/10.1038/25993

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. In *International Journal of Psychophysiology* (Vol. 50, pp. 19–26). https://doi.org/10.1016/S0167-8760(03)00121-1

Roeber, U., & Schröger, E. (2004). Binocular rivalry is partly resolved at early processing stages with steady and with flickering presentation: A human event-related brain potential study. *Neuroscience Letters*, *371*(1), 51–55. https://doi.org/10.1016/j.neulet.2004.08.038

Roeber, U., Widmann, A., Trujillo-Barreto, N. J., Herrmann, C. S., Shea, R. P. O., & Schröger, E. (2008). Early correlates of visual awareness in the human brain : Time and place from event-related brain potentials. *Journal of Vision*, *8*(21), 1–12. https://doi.org/10.1167/8.3.21.Introduction

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147–53.

https://doi.org/10.1093/cercor/bhl024

Salomon, R., Kaliuzhna, M., Herbelin, B., & Blanke, O. (2015). Balancing awareness: Vestibular signals modulate visual consciousness in the absence of awareness. *Consciousness and Cognition*, *36*, 289–297. https://doi.org/10.1016/j.concog.2015.07.009

Salomon, R., Lim, M., Herbelin, B., Hesselmann, G., & Blanke, O. (2013). Posing for awareness : Proprioception modulates access to visual consciousness in a continuous flash suppression task. *Journal of Vision*, *13*(7), 1–8. https://doi.org/10.1167/13.7.2.doi

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology. Human Perception and Performance*, *33*(6), 1311–1321. https://doi.org/10.1037/0096-1523.33.6.1311

Sergent, C., & Dehaene, S. (2004). Is Consciousness a Gradual Phenomenon ? *Psychological Science*, *15*(11), 720–728. https://doi.org/10.1111/j.0956-7976.2004.00748.x

Serino, A., Farnè, A., Rinaldesi, M. L., Haggard, P., & Làdavas, E. (2007). Can vision of the body ameliorate impaired somatosensory function? *Neuropsychologia*, *45*(5), 1101–7. https://doi.org/10.1016/j.neuropsychologia.2006.09.013

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), 19614–19619. https://doi.org/10.1073/pnas.1211645109

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*(1), 1–22. https://doi.org/10.3758/BF03206843

Spivey, M. J., & Spirn, M. J. (2000). Selective visual attention modulates the direct tilt aftereffect. *Perception & Psychophysics*, *62*(8), 1525–1533. https://doi.org/10.3758/BF03212153

Stein, B. E. (1998). Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. In *Experimental Brain Research* (Vol. 123, pp. 124–135). https://doi.org/10.1007/s002210050553

Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Alex Meredith, M., Perrault, T. J., … Lewkowicz, D. J. (2010).

Semantic confusion regarding the development of multisensory integration: A practical solution. *European Journal of Neuroscience*. https://doi.org/10.1111/j.1460-9568.2010.07206.x

Stein, B., London, N., Wilkinson, L., & Price, D. (1996). Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis. *Journal of Cognitive Neuroscience, 8*(6), 497–506. https://doi.org/10.1162/jocn.1996.8.6.497

Stein, B., & Meredith, M. (1993). *The Merging Senses*. *Cognitive Neuroscience Series*.

Stein, London, N., Wilkinson, L., & Price, D. (1996). Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis. *Journal of Cognitive Neuroscience*, *8*(6), 497–506. https://doi.org/10.1162/jocn.1996.8.6.497

Stein, Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli. *Journal of Cognitive Neuroscience, 1*(1), 12–24. https://doi.org/10.1162/jocn.1989.1.1.12

Stoerig, P., & Cowey, A. (1997). Blindsight in man and monkey. *Brain, 120*(3), 535–559. https://doi.org/10.1093/brain/120.3.535

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*(3), 679–690. https://doi.org/10.1093/cercor/bhk016

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098–114. https://doi.org/10.1162/0898929054475172

Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*, *14*(1), 106–114. https://doi.org/10.1016/S0926-6410(02)00065-4

Teder-Sälejärvi, W. A., Russo, F. Di, Mc Donald, J. J., & Hillyard, S. A. (2005). Effects of Spatial Congruity on Audio-Visual

Multimodal Integration, (V), 1396–1409.

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*(9), 1396–1409. https://doi.org/10.1162/0898929054985383

ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M. M., & Matusz, P. J. (2016). The COGs (Context-Object-Goals) in multisensory processing. *Experimental Brain Research*. https://doi.org/10.1007/s00221-016-4590-z

Todorovic, A., Schoffelen, J. M., Van Ede, F., Maris, E., & De Lange, F. P. (2015). Temporal expectation and attention jointly modulate auditory oscillatory activity in the beta band. *PLoS ONE*, *10*(3), 1–16. https://doi.org/10.1371/journal.pone.0120288

Tong, F., & Engel, S. A. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature*, *411*(6834), 195–9. https://doi.org/10.1038/35075583

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, *10*(11), 502–511. https://doi.org/10.1016/j.tics.2006.09.003

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5*, 42–64. https://doi.org/10.1186/1471-2202-5-42

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, *215*(3), 216–242. https://doi.org/215/3/216 [pii]

Tononi, G. (2012). Integrated information theory of consciousness : an updated account, 290–326.

Treue, S., & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–101. https://doi.org/10.1038/nn1500

Vallar, G., Halligan, P. W., Fink, G. R., Marshall, J. C., & Vallar, G. (2003). Spatial cognition : Evidence from visual neglect Spatial cognition : evidence from visual neglect, *6613*(August 2017). https://doi.org/10.1016/S1364-6613(03)00032-9

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(5), 1053–1065. https://doi.org/10.1037/0096-1523.34.5.1053

Van der Burg, E., Olivers, C. N. L., & Theeuwes, J. (2012). The attentional window modulates capture by audiovisual events. *PLoS ONE*, *7*(7). https://doi.org/10.1371/journal.pone.0039137

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *NeuroImage*, *55*(3), 1208–1218. https://doi.org/10.1016/j.neuroimage.2010.12.068

van Ee, R., van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(37), 11641–11649. https://doi.org/10.1523/JNEUROSCI.0873-09.2009

Van Rullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(03)00095-0

Veser, S., O’Shea, R. P., Schröger, E., Trujillo-Barreto, N. J., & Roeber, U. (2008). Early correlates of visual awareness following orientation and colour rivalry. *Vision Research*, *48*(22), 2359–2369. https://doi.org/10.1016/j.visres.2008.07.024

Vidal, J., Giard, M. H., Roux, S., Barthélémy, C., & Bruneau, N. (2008). Cross-modal processing of auditory-visual stimuli in a no-task paradigm: A topographic event-related potential study. *Clinical Neurophysiology*, *119*(4), 763–771. https://doi.org/10.1016/j.clinph.2007.11.178

Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*(4), 651–659. https://doi.org/10.3758/BF03194427

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(5), 1583–1590. https://doi.org/10.1037/0096-1523.26.5.1583

Wee, S., Chua, F. K., & Chua, F. K. (2004). Capturing attention when attention blinks. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 598–612. https://doi.org/10.1037/0096-1523.30.3.598

Wheatstone, C. (1838). Contributions to the Physiology of Vision.Part II. On Some Remarkable, and Hitherto Unobserved, Phaenomena of Binocular Vision,(Continued). *Proceedings of the Royal Society of London*, *6*(0), 138–141. https://doi.org/10.1098/rspl.1850.0044

Wilke, M., Logothetis, N. K., Leopold, D. A., & Planck, M. (2003).

Generalized Flash Suppression of Salient Visual Targets, *39*, 1043–1052.

Woldorff, M. G., Gallen, C. C., Hampson, S. a, Hillyard, S. a, Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, *90*(18), 8722–8726. https://doi.org/10.1073/pnas.90.18.8722

Yang, Y. H., & Yeh, S. L. (2014). Unmasking the dichoptic mask by sound: Spatial congruency matters. *Experimental Brain Research*, *232*(4), 1109–1116. https://doi.org/10.1007/s00221-014-3820-5

Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. *Control of Cognitive Processes: Attention and Performance Xviii*, 73–103. https://doi.org/10.2337/db11-0571

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., & Courtney, S. M. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nature Neuroscience*, *5*(10), 995–1002. https://doi.org/10.1038/nn921

Zhang, P., Jamison, K., Engel, S., He, B., & He, S. (2011). Binocular rivalry requires visual attention. *Neuron*, *71*(2), 362–9. https://doi.org/10.1016/j.neuron.2011.05.035

Zhou, W., Jiang, Y., He, S., & Chen, D. (2010). Report Olfaction Modulates Visual Perception in Binocular Rivalry. *Current Biology*, *20*(15), 1356–1358. https://doi.org/10.1016/j.cub.2010.05.059

Zhou, W., Zhang, X., Chen, J., Wang, L., & Chen, D. (2012). Nostril-Specific Olfactory Modulation of Visual Perception in Binocular Rivalry. *The Journal of Neuroscience*, *32*(48), 17225–17229. https://doi.org/10.1523/JNEUROSCI.2649-12.2012

Zihl, J., & Kennard, C. (2003). Disorders of Higher Visual Function. In *Neurological Disorders: Course and Treatment: Second Edition* (pp. 255–263). https://doi.org/10.1016/B978-012125831-3/50219-7

Zou, J., He, S., & Zhang, P. (2016a). Binocular rivalry from invisible patterns. *Proceedings of the National Academy of Sciences USA*, *113*(30), 8408–8413. https://doi.org/10.1073/pnas.1604816113

Zou, J., He, S., & Zhang, P. (2016b). Binocular rivalry from invisible patterns, *113*(30). https://doi.org/10.1073/pnas.1604816113