Title
Effects of Attention on Multisensory Integration

Author
Barrera, Steven

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Effects of Attention on Multisensory Integration

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Cognitive Science

by

Steven James Barrera

Committee in charge:

Professor Marta Kutas, Co-chair
Professor Jeanne Townsend, Co-chair
Professor Andrea Chiba
Professor Virginia de Sa
Professor Karen Dobkins
Professor Steven Hillyard

2019

The Dissertation of Steven James Barrera is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-chair

Co-chair

University of California San Diego

2019

DEDICATION

To my Dad, who gave me a direction. And to Seana and Callan, who keep me going.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I thank Professor Marta Kutas for her support as co-chair of my committee. Marta has been one of the constants during my time at UCSD. Her guidance, acumen, encouragement, and example have meant a great deal to me.

I also thank Professor Jeanne Townsend for her support as co-chair of my committee. Jeanne has been incredibly patient and helpful to me since I started working in her lab many years ago. She encouraged me to go back to graduate school and has always been a source of optimism and positivity. I am grateful for her advice and mentorship over the years.

I would also like to thank Marissa Westerfield for her support and advice on my project. She has been a great help throughout my graduate school career.

I am thankful for the time and advice offered by Andrea Chiba, Virginia de Sa, Karen Dobkins, and Steven Hillyard over the course of this project.

Chapter 2 is co-authored with Townsend, Jeanne and Westerfield, Marissa. The dissertation author was the primary author of this chapter.

Chapter 5 is co-authored with Townsend, Jeanne and Westerfield, Marissa. The dissertation author was the primary author of this chapter.

VITA

1991	Bachelor of Arts, Psychology, Stanford University
1996	Master of Arts, Philosophy, University of California San Diego
2006-2011	Graduate Student Researcher, University of California San Diego
2007-2010	Teaching Assistant, Cognitive Science, University of California San Diego
2009-2011	Associate Instructor, Cognitive Science, University of California San Diego
2012-2019	Lecturer, Cognitive Science, University of California San Diego
2019	Doctor of Philosophy, Cognitive Science, University of California San Diego

FIELDS OF STUDY

Major Field: Cognitive Science

Multisensory Perception, Embodied Perception, and Cognitive Electrophysiology.
Professor Jeanne Townsend

ABSTRACT OF THE DISSERTATION

Effects of Attention on Multisensory Integration

by

Steven James Barrera

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2019

Professor Marta Kutas, Co-chair
Professor Jeanne Townsend, Co-chair

The world presents information via a variety of sensory channels. To make sense of this information, we must determine what is relevant and ignore unhelpful noise. We then integrate congruent information within and across modalities to build coherent perceptions. Importantly, immediate goals and prevailing environmental factors may interact to affect our perceptual decisions. This dynamic process of multisensory integration is essential to successful perception in the real world, but can also lead to errors. The current project exploits some of

these perceptual errors to explore how endogenous (task-directed) and exogenous (stimulus intensity) factors may influence multisensory integration.

In a series of four experiments, we use the *sound-induced flash illusion* (SFI; Shams *et al*., 2000; 2002) and related audiovisual effects as indices of multisensory integration. Endogenous attention was manipulated using a *focused attention* visual task and a novel *bimodal conditional attention* task. In our first two experiments, we found that participants reported more illusions when attending to both sensory modalities. This effect was larger when the auditory stimuli were presented at near-threshold levels. Perceptual sensitivity (d') was also found to decrease in the bimodal condition. We then manipulated auditory intensity in each of these tasks independently. Reports of the SFI were found to increase with the higher intensity auditory stimuli. However, differences in reporting these illusions within the same task were attributable to both changes in bias (c) and d'.

Event-related potentials recorded in our first experiment revealed that the SFI was associated with smaller P3 potentials than found in valid targets. We also noted differences in the response-locked error positivity (Pe), with illusory stimuli having more positive amplitudes than real targets. However, the earlier occurring error-related negativity (ERN) was indistinguishable in real and illusory targets. This suggests that participants were less confident of the illusion during stimulus evaluation and one stage of response monitoring. We evaluate these results in terms of the *directed attention* and *information reliability* hypotheses (Andersen et al., 2004, 2005) and discuss how these and similar experiments may deepen our understanding of how multisensory perception is impacted at multiple stages of stimulus and response evaluation.

Chapter 1: General Introduction and Background

Now is the time for the burning of the leaves,
They go to the fire; the nostrils prick with smoke
Wandering slowly into the weeping mist.
Brittle and blotched, ragged and rotten sheaves!
A flame seizes the smouldering ruin, and bites
On stubborn stalks that crackle as they resist.
The last hollyhock's fallen tower is dust:
All the spices of June are a bitter reek,
All the extravagant riches spent and mean.
All burns! the reddest rose is a ghost.
Spark whirl up, to expire in the mist: the wild
Fingers of fire are making corruption clean.

From "The Burning of the Leaves," by Laurence Binyon

1.1 Our Multisensory World

What human sense is best suited to capturing the scene described above in Laurence Binyon's "The Burning of the Leaves?" The auditory system could detect the crackling sound as the leaves combust, while the olfactory system might confirm that something is burning and give a hint at the fuel source. Whereas the visual system would see the glow of the flames, much of the urgency of fire comes from the press of the heat on the skin as signaled by the somatosensory system. Importantly, prior knowledge and contextual information aid in the comprehension of these sensory events, helping one choose between an initial reaction of interest rather than alarm. And depending on your focus of attention, you may not notice any of these signals until they reach a magnitude that breaks through to your awareness.

To get a complete understanding of such scenes, it is clear that we use all of our senses. The signals flow from individual events within the environment, along different paths and mediums: photons of light being both projected and reflected; sound waves vibrating and compressing air molecules; scents and heat riding along shifting air currents. Each sensory signal conveys information that is particular to its own modality. Importantly, the paths and sensory transductions of these physical signals remain largely independent. From the perspective of an ideal observer, the elements these signals have most in common are overlapping spatial and temporal profiles. Somehow these disparate signals are transduced, processed, and integrated with near simultaneity by the situated observer. Critically, while the initial transductions may occur independently, the mere presence of co-occurring sensory signals may alter the manner and degree to which each individual signal is eventually processed. Heightened attention to specific elements of the scene may further modify the processing of sensory information in a dynamic fashion. That the brain accomplishes such instances of multisensory integration is an astounding, yet poorly understood characteristic of human perceptual experience.

To successfully navigate our multisensory environment, the incoming streams of continuous information from different modalities must be made perceptually coherent. The fact that integration seemingly occurs rapidly and effortlessly appears to be *prima facie* evidence for the notion that this is a largely automatic process, driven by what is often thought of as 'low-level' mechanisms. However, we are not merely passive observers of events around us; our goal-driven behaviors influence which sensory items we engage and how we perceptually frame our environment. Additionally, repeated exposure to sensory experiences helps shape our future perceptions in the form of associative learning. These effects, too, seem readily evident in everyday activity and argue for a strong role of 'top-down' processing in multisensory perception. In the end, most individuals are able to act in a world experienced as a relatively well-ordered, structured environment. How the human brain quickly intertwines bottom-up and top-down elements of multisensory processing to accomplish this perceptual feat remains an area of vigorous research (see for example Calvert & Thesen, 2004; Driver & Spence, 1998; Stein & Meredith, 1993) and is the subject matter of this thesis.

The general goal of the current project is to explore and quantify the relationships between stimulus-driven (i.e. bottom-up) and goal-driven (i.e. top-down) effects of attention when humans perceive multisensory stimuli. We focus on instances where incongruent auditory and visual stimulus properties are used in differing attention tasks to create variable perceptual ambiguities. Using the ‘sound-induced flash illusion’ as our primary exemplar, we probe the conditions in which the contemporaneous presentation of two auditory ‘beeps’ and one visual ‘flash’ can induce a second ‘illusory flash’ in many individuals (Shams, Kamitani, & Shimojo, 2000; 2002). The question we attempt to answer is: does increased attention to the auditory stimuli, caused by either exogenous stimulus intensity or endogenous task requirements, make perception more accurate or does it induce enhanced integration which leads to more perceived illusions? To address these issues, we collect and evaluate overt behavioral responses and electroencephalographic data as participants perform a number of tasks that elicit the sound-induced flash illusion. These tasks are designed to tease apart the roles that exogenous and endogenous factors may play during sensory integration. Our resulting data are used to create functional, neurophysiologically inspired models of the integrative processes that produce the sound-induced flash illusion. These integrative models can then be generalized to address broader effects of attention on multisensory perception.

1.2 Our Multisensory Perceptions

Shifting from the poetic to the prosaic, it is useful to review some of the empirical findings which describe the manner and degree to which the signals presented in different sensory channels influence one-another during perceptual tasks. This will begin with a treatment of the overt behavioral effects attributed to multisensory integration before turning to a discussion of neurophysiological mechanisms that may underlie those effects. Multisensory integration has been studied in a variety of contexts, ranging from simple, low-level orienting responses to more complex, high-level cognitive tasks. Although it is something of a false dichotomy to divide perceptual processes into 'simple' and 'complex' categories, it remains a useful exploratory device.[1] These categories will be further developed in the context of exogenous and endogenous processes, respectively. The relationship between these two domains is a central element of the current project and is described in more detail below.

1.2.1 Integration in perceptually simple tasks

Beginning with simpler behaviors, it has been noted that multisensory integration imparts a number of practical benefits when reacting to the sudden appearance of objects. When scanning a sparse visual scene, subjects reliably saccade faster to visual targets when they are accompanied by spatially and temporally congruent auditory stimuli (Frens, Van Opstal, & Van der Willigen, 1995; Goldring, Dorris, Corneil, Ballantyne & Munoz, 1996; Hughes, Nelson, & Aronchick, 1998). In these instances, the auditory signal need have no relevance to the task, nor add predictive value to the visual target's location. Similar studies have shown a saccadic reaction time advantage to visual stimuli when they are paired with

[1] Unless otherwise noted, 'simple' and 'low-level' will be used interchangeably to indicate a lesser degree of neural processing. Likewise, 'complex' and 'high-level' will each denote activities thought to require a greater amount of processing.

unrelated somatosensory stimuli (Amlot, Walker, Driver, & Spence, 2003; Diederich, Colonius, Bockhorst, & Tabeling, 2008). The tendency to look more rapidly to locations that produce stimulation in multiple sensory modalities may maximize the utility of incoming sensory information, conferring survival advantages discussed below.

It has been suggested that specialized multisensory cells in deep layers of the superior colliculus (SC) provide the neural basis for this advantage in saccade generation (Colonius & Diederich, 2004; Meredith & Stein, 1986; Stein & Meredith, 1993). These cells are mildly responsive to individually presented unisensory signals (e.g. visual or auditory stimuli), but may become much more active – in a superadditive fashion – in the presence of spatially and temporally aligned stimuli from multiple modalities (e.g. visual and auditory stimuli together). More will be said about this later, but the critical hypothesized role of SC is worth mentioning now as its subcortical locus of operation has been used to suggest that a fast, automatic integration process may precede in-depth cognitive processing (i.e. categorizing the nature of the target after orientation). The specialized cells in SC react to sensorially redundant stimuli and may serve as an automated orienting mechanism to speed responses to a target in the environment. This function makes intuitive evolutionary sense as organisms that can rapidly notice threats might have an advantage over those which cannot. Observers can also benefit from prior knowledge when making saccades. If auditory stimuli have a history of being paired with a visual target, subjects will look faster to paired signals when the prior probability of spatial alignment is higher (Van Wanrooij, Bremen, & Van Opstal, 2010). It seems that even in the simplest cases of multisensory orienting, exogenous and endogenous factors may interact to impact performance.

Given that one functional outcome of multisensory integration is to speed orientation toward targets, one might expect that additional overt behavioral advantages should follow. Early research in behavioral psychology indeed demonstrated manual reaction time (RT) advantages to stimuli presented simultaneously in multiple modalities (Hershenson, 1962; Todd, 1912). Numerous studies have since confirmed that the simultaneous (or near simultaneous) presentation of targets in auditory and visual channels decreases manual RTs (Diederich & Colonius, 1987, 2004; Giray & Ulrich, 1993; Miller 1982, 1986). Reaction times to visual targets are also speeded by the congruent presentation of tactile stimuli (Forster, Cavina-Prates, Aglioti, & Berlucchi, 2002). Additionally, RTs to auditory stimuli and tactile stimuli can be decreased when presented simultaneously (Zampini, Torresan, Spence, & Murry, 2007). Finally, combining visual, auditory, and tactile stimuli all together has also been shown to decrease RTs when compared to bimodal combinations of the same stimuli (Diederich & Colonius, 2004). From this sample of findings, it is clear that responses to environmental stimuli can be facilitated by multisensory integration in a variety of circumstances and combinations.

In addition to increasing the speed of detection and response, target detection accuracy is also enhanced in the presence of multimodal signals (Ngo & Spence, 2010; Vroomen & de Gelder, 2000). Accuracy for visual target detection, as assessed using hit rates and psychophysical sensitivity measures, has been shown to increase when the targets are temporally and spatially concurrent with the presentation of irrelevant auditory stimuli. In one study, the presence of a spatiotemporally congruent auditory signal increased the d' score associated with detection of a below-threshold visual target (Bolognini, Frassinetti, Serino, & Ladavas, 2005). The use of d' as a dependent measure is critical as it demonstrates that perceptual sensitivity is increased in the presence of multisensory stimuli (Frassinetti, Bolognini, & Ladavas, 2002)[2]. Findings using psychophysical sensitivity measures are consistent with the notion that RT effects cannot be merely ascribed to response bias (as measured with *c* or *β*) or statistical facilitation caused by the appearance of several triggering events (Diederich & Colonius, 2004; Miller, 1982).

1.2.2 Integration in perceptually complex tasks

Studies of the response facilitation resulting from multisensory integration have traditionally focused on 'low-level' phenomena using simple stimuli. However, cross-modal effects are also apparent in what are thought of as 'high-level' cognitive tasks. In the domain of spoken language comprehension, it has long been known that the presentation of auditory speech with a congruent display of facial articulation improves speech perception (Helfer, 1997; Sumby & Pollack, 1954). Seeing a speaker as she produces auditory speech has been shown to improve perception at a variety of signal-to-noise ratios and is not limited to overcoming adverse sound environments (Remez, 2005). Even when auditory speech is clearly spoken, the presence of visual facial cues can improve language perception when the type of language used (i.e. structurally or semantically complex) makes it hard to understand (Arnold & Hill, 2001; Reisburg, Mclean, & Goldfield, 1987). Whether the auditory and visual cues act by increasing the signal level for correct categorical perception or confining the problem space by reducing the possible matches that fit the information presented in both sensory modalities, it becomes increasingly clear that complementary auditory and visual signals provide great utility during rapid speech perception.

[2] The use of psychophysical measures in multisensory paradigms will be discussed in chapter 6.

Another ‘high-level’ area where multisensory integration has recently been found to have a significant impact is perceptual learning in structured training environments. The pairing of auditory and visual information during training sessions can increase the rate and total amount of improvement in performance compared to unisensory training alone (Shams & Seitz, 2008). In one case of multisensory training, subjects performed a motion detection and discrimination task. Half of the subjects were trained with audiovisual stimuli, in which visual, auditory, or audiovisual stimuli all contained informative cues regarding motion. The remaining subjects received equivalent training, but with visual stimuli only. When both groups were tested on a visual-only motion detection and discrimination task, it was found that those who had received the multisensory training performed better both within and across training sessions (Seitz, Kim, & Shams, 2006). A similar benefit for auditory voice recognition was found following training with the paired presentation of a speaker’s face with auditory speech, compared to training with the auditory voice alone (von Kreigstein & Giraud, 2006).

Such findings are notable for at least two reasons. First, they demonstrate that attentive multisensory training has quantifiable, long-term advantages over unisensory presentations of the same task. This may be due to deeper initial encoding of the task cues created by the more elaborated multimodal stimuli. Second, they suggest that multisensory training benefits are not confined to the bimodal stimuli, but may extend to tasks in the unisensory domain. This latter point is especially relevant to recent attempts to use multisensory stimuli to overcome unisensory deficits (Ladavas, 2008). For example, in patients with visual field deficits, auditory stimuli can be paired with visual targets to increase their likelihood of conscious detection in the affected region (Frassinetti, Bolognini, Bottari, Bonora, & Ladavas, 2005).

Hemianopia patients who receive orienting training using audiovisual targets show improved visual detection and faster oculomotor responses during visual search tasks compared with controls who received only visual training (Passamonti, Bertini, & Ladavas, 2009).

The successful use of multimodal stimuli to enhance the function of degraded unisensory processes suggests that perception is a dynamic process involving fundamentally multisensory object representations (Shams, Wozny, Kim, & Seitz, 2011). The behavioral advantages of activating these multisensory representations can be seen for tasks both simple (e.g. orienting) and complex (e.g. speech processing). These findings also raise interesting questions about the nature of the neural connections between traditionally unisensory processing areas. However, to fully explain how these multisensory representations can be encoded, manipulated, and put to use, we need to understand the role attention plays in matching task goals to environmental conditions.

1.3 Attention and Multisensory Integration

Over the last two decades, there has been an increasing amount of research examining the impact of attention on multisensory integration (Spence & Driver, 2004). A quick search of pubmed.gov listing for "multisensory integration" for 2000 – 2009 finds 418 entries. This jumped to over 1,650 results from 2010 – 2019. However, the precise role of attention during the integration process remains unclear and has not benefitted as much from systematic study.

Possibly because early biological work in crossmodal effects was based on recordings of multisensory cells (Stein & Meredith, 1993), integration was often conceptualized as an automatic process which would be largely unaffected by attention. For example, some behavioral studies previously found that audiovisual integration can occur in unattended stimuli, suggesting that the process is pre-attentive (Bertelson, Vroomen, De Gelder, &

Driver, 2000; Vroomen, Bertelson, & De Gelder, 2001). However, more recent work has revealed modulatory effects of attention on integration at both ‘low’ and ‘high’ levels of sensory processing (Koelewijn, Bronkhurst & Theeuwes, 2010; Talsma, Senkowski, Soto-Franco & Worldorf, 2010), allowing a more complex picture of the interaction of attention and multisensory integration to emerge. The primary objective of the current project is to begin to quantify how attention can impact the integration of multisensory stimuli. In the section below, we delineate some key operational categories of attention relevant to the present effort, and review noteworthy studies that have begun to make progress in this domain.

Broadly construed, attention can be thought of as comprising a number of dynamic, interactive mechanisms that help in the selection of salient or contextually relevant environmental stimuli for additional processing. Because this general definition can be attached to the analysis of many stimulus elements, a number of attention ‘types’ have been identified which tend to cluster around the specific item under study. For example, mechanisms that are selective on the basis of location would be classified as *spatial attention*. Processing based on association with an object would be classified as *object-based attention*. The same can be said of *feature-based attention*, and so on. While this approach seems to create a multiplicity of attention types, they all refer to a modification in underlying biological responses linked to a specific selection filter (e.g. space, object, and feature). In the present project, we focus on changes in processing associated with attention to specific sensory modalities. While we will call this *modal attention,* it should be construed as simply highlighting the particular processes associated with the selection of specific modalities for enhanced analysis and not exclusive of other selective filters. More will be said about this below, but the critical point here is that we presume each variety of 'attention' refers to a general category of processing modification which can be expressed in different selection activities. That is, we don't choose between a unitary or multiplicative notion of 'attention', but note that the general phenomenon emerges in many instantiations.

The elements which drive selective attention can vary greatly, but tend to be described in two ways: those that impel additional processing by virtue of their stimulus properties and those that receive the imparting of additional resources due to their current relevance or meaning. This basic distinction will be discussed more below. However, this portrayal of attention is particularly relevant in the context of sensory integration as it spans both low-level perceptual and high-level cognitive domains. For example, highly salient stimuli should be more attended whether embedded in a simple saccade task or a complex language task. As it has already been shown that multisensory integration has measurable effects during both simple orienting behaviors and complex cognitive behaviors, it appears that attention can impact integrative processes at multiple levels.

Based on an early distinction by William James (James, 1890), we can further specify the above distinction as dividing attention into two operational categories: (1) passive and involuntary and (2) active and voluntary. The first category, now referred to as *exogenous attention*, is typically characterized as a 'bottom-up' or low-level perceptual process in which stimulus characteristics evoke a series of automated responses. For example, a loud stimulus might engage more auditory processing than a quiet stimulus simply by virtue of the difference in intensity. These automatic or reflexive reactions to the stimuli result in altered neural processing which may yield measureable reaction time advantages as outlined above. Importantly, any modification in neural activity would be contained within the sensory pathways. This can be contrasted with the second category, *endogenous attention*, in which an agent uses ‘top-down’ cognitive resources to willfully select particular stimuli for additional processing, to the exclusion of others (e.g. listening to a specific speaker while consciously ignoring sounds in the background). In this case, the intentional exclusion of some stimuli may decrease processing of competing stimuli or enhance those at the focus of attention, again resulting in behavioral advantages compared to unattended stimuli (for a review see e.g. Carrasco, 2011). Neural activity, in this case, would be required from areas outside of sensory processing to manage this selectivity.

In the previous sections, many of the ‘low-level’ behavioral effects attributed to multisensory integration can be thought of as exogenous and stimulus-driven. For example, the presence of stimuli in both auditory and visual channels may trigger faster saccadic reaction times to a target even if the auditory stimulus is irrelevant and unattended (Goldring *et al*, 1996). Such effects suggest that multisensory integration can be a largely automatic process. The more ‘high-level’ multisensory phenomena (e.g. speech perception and learning) are thought to reflect endogenous processing as they are specific to certain task- and goal-related activities involving agent-directed attention. In the case of speech perception, the individual intentionally focuses on both the movement of a speaker’s lips and the sound of her voice in order to better perceive what is said (Sumby & Pollack, 1954). This endogenous process requires the selection of specific elements from two modalities, which are integrated to form a coherent percept. In such instances, some aspect of the integrative process appears to be enhanced by the demands of the task, even while automatic processes operate in parallel.

In the above examples, it is important to note how attention may be allotted to different sensory modalities during experimental tasks. In the saccade task, only visual targets were selected, while other modalities were ignored. Such designs are sometimes referred to as a *focused attention paradigm* (Diederich & Colonius, 2004). In the speech perception example, valid targets can appear in multiple modalities. This is generally referred to as a *redundant target paradigm* (though sometimes the targets can be redundant within the same sensory modality). Since each paradigm selects information partially based on the sensory channel being attended, we suggest that these paradigms each manipulate modal attention to some degree. Focused attention paradigms attempt to exclude information from an unattended modality, while redundant target paradigms typically require sustained attention to information across two modalities.

Critically, modal attention is rarely an independent variable in either paradigm. Stimulus features (e.g. intensity, spatial location, or onset timing) are more frequently the focus of multisensory studies. Unanalyzed in such studies is the degree to which exogenous factors may interact with modal attention, especially as this may be affected by individual differences. Characterized in this way, we can see how modal attention may implicitly affect the degree to which sensory stimuli from multiple sensory channels may be integrated. With this framework in place, we now examine how attention has been seen to operate in selected multisensory tasks.

1.3.1 The spread of attention across modalities

Most research of low-level and high-level perceptual effects in multisensory integration has tended to use a focused attention paradigm, treating attentional selection as a purely exogenously determined factor. Such studies attempt to ascertain the degree to which integration across the senses is automatic and ‘pre-attentive’. In one such study using a difficult visual search task, subjects were asked to find a vertical or horizontal line amidst a number of similar diagonal distractors (Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). The color of the visual items alternated between red and green and the set size was 24, 36, or 48 items. In some trials, a non-informative auditory 'pip' played when the target changed color. Search times were found to greatly decrease when the auditory pip was presented as the target changed color. When the auditory pip was not present, search time increased as the visual set size increased. This 'pip and pop' effect is consistent with other 'pop-out' effects in which the visual target can be quickly selected by virtue of a highly salient visual cue, such as a unique color (Treisman & Gelade, 1980). The effect was not duplicated by replacing the auditory pip with a non-spatial visual cue, suggesting that the result was not due to simple generalized alerting. Additionally, lowering the reliability of the auditory pairing from 80% to 20% validity did not reduce the effect, arguing against the idea that the RT search advantage was due to a high-level strategic decision or learned association.

The above example, coupled with other studies of audiovisual integration effects (Bertelson *et al*., 2000; Vroomen *et al*., 2001), are consistent with the notion that multisensory integration is a largely automatic process. To the extent that attention is involved, it might appear that exogenous, low-level stimulus characteristics trigger integrative mechanisms. In such an account, endogenous selective attention would be relatively ineffective in modulating these automatic processes. In one test of this hypothesis, Busse and colleagues investigated the role of covert visual spatial attention on the integration of a visual target with an irrelevant auditory sound (Busse, Roberts, Crist, Weissman, & Woldorff, 2005). When an object appeared at an attended location, a simultaneously presented auditory sound received a greater cortical response (as measured by event-related potentials and magnetic resonance imaging) than those sounds accompanying a visual object presented at an unattended location. The authors suggested that the neural processing advantages of visual spatial attention spread to the unattended auditory modality as a function of endogenously directed covert attention. While this spread of attention across modalities itself may be an automatic process, this finding suggests that endogenous, task-directed spatial attention may alter subsequent cross-modal processes.

In a further examination of the role of attention on cross-modal processing, Alsius and colleagues examined how attentional load might influence automatic integration of auditory speech with visual articulatory gestures (Alsius, Navarra, & Soto-Faraco, 2007). Using the well-known *McGurk effect* (McGurk & MacDonald, 1976) as a dependent measure, the study examined how visual mouth movements would affect linguistic categorization of spoken auditory words when subjects were asked to perform simultaneous tactile tasks. When mismatching visual and auditory speech cues are presented, the multimodal sensory information can be combined to form a third categorical percept. This McGurk effect is thought to be the result of automatic cross-modal integration. However, when subjects were required to perform a tactile task while also attending to the audiovisual speech stimuli, it was found that they experienced fewer McGurk illusions. This experiment suggests that attending to the somatosensory modality decreased the amount of cross-modal integration occurring between the auditory and visual sensory channels.

The above examples offer intriguing evidence that endogenous attention may influence multisensory integration. A more extreme argument has been made that some level of attention may be a prerequisite for cross-modal integration. Under the *Feature Integration Theory* (Treisman & Gelade, 1980; Treisman, 1996), attention may act as the ‘glue’ that binds multiple features into unitary perceptual objects. While spatial attention usually plays a prominent role in this account, focused endogenous attention may also play a part in feature integration. Indeed, there is electrophysiological evidence to support the idea that attention to all modalities of a multisensory object may be required to elicit some early neural correlates of cross-modal integration (Talsma, Doty, & Woldorff, 2007). However, one of the problems with existing studies of exogenous and endogenous attention effects on sensory integration is that they are usually studied independently (Talsma *et al*., 2010). In the current project we introduce a single experimental paradigm which is capable of testing both exogenous and endogenous attentional effects on multisensory integration. This will help quantify the relative effects of each and shed light onto possible interactions.

1.4 Measuring and Modeling Multisensory Integration

Approaching the topic of multisensory integration from the perspective of cognitive neuroscience, the present project attempts to address the issue on two analytical levels. First, we would like to provide an operational understanding of multisensory processing within the context of a specific psychological phenomenon. We chose the sound-induced flash illusion (Shams *et al*., 2000; 2002) as our exemplar for reasons to be explained further below. In brief, its startling phenomenological quality of producing distinct perceptual experiences when using identical physical stimuli (i.e. the same experimental participant can sometimes experience a single flash and two beeps accurately or as two flashes and two beeps) makes its observation a powerful dependent measure and proxy for multisensory integration. Second, we would like to explore the degree to which different underlying neural activity is evoked and affected by our multisensory phenomenon, as first delineated at the operational level. These complementary interests can be usefully pursued in tandem. In the present project, we explore the relationship between psychophysical measures of multisensory integration emerging from the sound-induced flash illusion and their electrophysiological correlates.

Understanding the neurological mechanisms by which multisensory integration is accomplished has long been a goal of neuroscience. The measures used in the current project build upon methodological developments in multisensory research developed over the last two decades. To understand contemporary methods and their underlying assumptions, it is helpful to review some recent history in cross-modal research. This is most usefully done by beginning with the seminal work presented by Stein and Meredith in 1993.

1.4.1 The foundational work of Stein and Meredith

While multisensory integration is part and parcel of the larger study of sensory perception, it has benefited from renewed focus in the last thirty years. The increase in multisensory integration research interest and growing consensus in analytical approaches was greatly advanced by the work of Barry Stein and M. Alex Meredith, *The Merging of the Senses* (1993). In this book, Stein and Meredith provide a synthesis of multisensory research spanning decades. And while their examples come largely from experimental work in non-human animals, their underlying premise is that findings in animal models will be directly applicable to research in human perception.

Stein and Meredith begin by establishing that multisensory integration is a fundamental biological process which operates across species. They argue that environmental cues from different modalities are linked together within the nervous system in order to rapidly facilitate judgments and direct consequent behavior. Organisms capable of quickly processing information from multiple sensory sources would enjoy survival benefits over less flexible or computationally impoverished competitors. After providing a brief account of how illusory sensory convergence alters perception and action in humans, Stein and Meredith demonstrate how sensory integration strongly affects behavior in numerous species – even those thought to be evolutionarily primitive by comparison. One example of multisensory integration provided is the lesson of Pavlov's dog (Pavlov & Anrep, 1927). In this iconic case of classical conditioning, repeatedly pairing a sound (a bell ringing) with a food reward conditions a dog to respond to the sound by salivating, even in instances where the food reward is withheld after the sound. Eventually, the integration of auditory and taste stimuli enable the animal to anticipate and react to food more quickly. This behavior is cited as a type of associative learning which requires integration of sensory stimuli at the neural level.

Stein and Meredith observe that the above example can possibly be attributed to a high-order learning mechanism available to species of a certain neurological complexity. However, they also point to similar behavioral processes in snails. Alkon and colleagues (1983) demonstrated that repeatedly pairing multimodal stimuli in marine snails produces long-term changes in behavior. Under normal conditions, a marine snail will move toward a light stimulus. However, when water swirls around them, vibration sensors on the snails trigger a defensive anchoring behavior. Alkon showed that by repeatedly pairing light stimuli with swirling water, one could train the snails to suppress motion towards light and anchor themselves when the light was presented alone. This indicates that a type of cross-modal associative learning can be accomplished by simple organisms. Importantly, Alkon and colleagues also established that the behavioral change was dependent upon a class of cells that were sensitive to both light and vibration stimuli. Stein and Meredith use this example to solidify the notion that sensory integration affects behavior across species and can be usefully examined at a neural level. For the present project, it is interesting to note that this example demonstrates cross-modal effects without appeal to ‘high-level’ or exceptionally sophisticated neural mechanisms. This is instructive when recalling previously mentioned examples of ‘high-level’ multisensory learning effects such as the advantage of using paired audiovisual learning cues in humans (Seitz *et al*., 2006).

After providing a few examples of the types of behaviors which seem to require multisensory integration, Stein and Meredith then begin an in-depth anatomical and functional review of a sub-cortical region known to be reactive to visual, auditory, and somatosensory sensory modalities: the superior colliculus (and its pre-mammalian homologue, the optic tectum). The role of the superior colliculus in mediating visual experience, beyond coordinating saccadic motor movements, was demonstrated in experiments showing that lesions in this area induce visual neglect in cats (Sprague & Meikle, 1965). Stein and Meredith use this finding to support the notion that multiple sensory streams can most effectively modulate behavior when they converge and have access to the same motor output circuits (Stein *et al*., 1976). Stein and Meredith then review a wealth of experiments uncovering the functional significance of the superior colliculus for sensory integration. For our present purposes, however, it will be useful to turn to the methods experimenters employed and the measures which resulted from this research.

Stein and Meredith chronicle how emerging methods in functional neuroanatomy affect our understanding of multisensory integration. Research studies using alert, behaving animals became instrumental in demonstrating how external stimuli induce neural activity, as measured by single-cell recordings (Gordon, 1973; Straschill & Hoffmann, 1970). The use of probes for direct electrical stimulation of multisensory neurons could also be employed to demonstrate the motor outcomes of signal integration. For example, stimulating neurons in the cat superior colliculus demonstrated how eye and ear movements are in spatial register, revealing the map-like organization in this portion of the brain (Stein & Clamann, 1981). These developments were important complements to traditional staining and dissection techniques which provided less functional information than the living, acting organism.

A critical analytical device to come out of the single-cell recording experiments was the means of identifying multisensory responses. Meredith and Stein (1986b) identified a type of cellular firing rate response enhancement in superior colliculus neurons that was selectively sensitive to the combination of both auditory and visual stimuli. When cats were exposed to either auditory or visual stimuli, slow firing rate responses were recorded in these cells. However, when the auditory and visual stimuli were presented simultaneously, a large firing rate enhancement was observed in some neurons. The cellular response to the multisensory stimuli was determined to be larger than the sum of the responses to the unisensory stimuli. This multisensory response is now known as a *multiplicative* or *superadditive* effect. While other measures are also used to indicate multisensory interactions at the cellular level (e.g. cellular depression and sub-threshold activation), testing for superadditivity has become a common means of identifying multisensory operations. Despite the success of this analytical technique, as we note later, there are some reasons to ask whether this standard has been relied on too much (Laurienti *et al.*, 2005; Stanford & Stein, 2007).

Following a review of experiments exploring the effects of integration within the superior colliculus, Stein and Meredith arrived at a set of rules which characterize the impact of multisensory stimuli on the receptive fields of multisensory cells. First, multisensory response enhancement typically requires the multimodal stimuli to be spatially coincident. The exact degree of coincidence is determined by size of the receptive fields of the involved neurons. If the stimuli are spatially removed from one-another, you will see either no interaction or possibly inhibition at the cellular level. Secondly, although exact temporal coincidence is not required of the stimuli, the neural temporal patterns resulting from the activity of the multimodal stimuli must overlap. That is, cellular responses to two modalities must remain co-active for a multisensory effect to be observed. Based on Stein and Meredith's review, they put the optimal temporal window for multimodal interaction at approximately 100 ms.

The third rule to emerge from their study is the observation that neurons do not change their receptive fields based on multisensory input. The unimodal receptive fields of these neurons remain constant, even though multisensory stimuli elicit multiplicative behavior. Additionally, Stein and Meredith point out that the neurons of the superior colliculus do not respond to multiple inputs from the same modality in the same way as they respond to stimuli from multiple sensory streams. This simply highlights the fact that stimuli within modalities do not create the same superadditive responses in the superior colliculus as produced from multisensory stimuli, as might occur if enhancement was due to summed stimulation energy. In their final rule, Stein and Meredith observed that the largest enhancement in multisensory responses is brought about by the weakest, minimally effective unimodal stimuli. Since multisensory enhancement identification is a comparative response rate, the largest and most robust percentile change in activity will be observed using the smallest effective stimuli.

While Stein and Meredith's review of animal work focused on the multisensory receptive properties of neurons in the superior colliculus, similar sensory integration has been found in the anterior ectosylvian sulcus (AES) of cats. In tests of the spatial and temporal properties of AES neurons, Wallace and colleagues (1992) found that overlapping receptive fields for auditory, visual or somatosensory stimuli displayed multisensory enhancement similar to that seen in superior colliculus neurons. This is significant for at least two reasons: (1) it demonstrates that the rules for multisensory integration proposed by Stein and Meredith are applicable to areas outside of the superior colliculus and may, therefore, be representative of multisensory interactions in general; and (2) the above rules seem operative in both cortical and sub-cortical structures. Interestingly, the multisensory response profile of AES neurons has been found to develop postnatally (Wallace *et al.*, 2006) and plays an important part in shaping multisensory functions of target superior colliculus neurons (Jiang *et al.*, 2006; Wallace & Stein, 2000).

Stein and Meredith's rules for sensory integration have provided the theoretical framework for a wide variety of multimodal perceptual studies. These rules, situated in studies spanning numerous species, now provide a solid foundation for exploring both the psychophysical and neural underpinnings of a host of multisensory processes. However, perhaps an even greater achievement of their 1993 publication was to make the broad body of electrophysiological work available to a wider audience. Due to their lucid and detailed description of complex neurobiological experiments, new generations of psychologists, cognitive scientists, and computer scientists have gained access to important insights that may otherwise have remained hidden from view. The cross-disciplinary pollination enabled by their exposition is difficult to overemphasize as a source of innovation in modern multisensory research.

While the above methods have proven very successful in animal studies, a different approach is generally required in human studies. Aside from occasional clinical applications, invasive techniques such as single-cell recordings and direct neural stimulation are not permissible.[3] For studies in human populations, non-invasive techniques such as electroencephalography and event-related potentials (EEG/ERP), functional magnetic resonance imaging (fMRI), and positron emission tomography (PET) prove more practical and feasible. Noteworthy here is the continued belief that Stein and Meredith's rules of multisensory integration, and specifically the measure of superadditivity, will continue to apply at the level of whole brain imagery and measurement. We will employ EEG/ERP recordings for uncovering neural indications of multisensory integration in chapter 5. We now turn to a discussion of how conflicting multisensory stimuli and resulting perceptual illusions have been used to examine multisensory processes. We then focus on the specific multisensory phenomenon targeted in this project and our specific operational aims.

1.4.2 Multisensory illusions

Stein and Meredith began their 1993 analysis of multisensory interaction by reviewing a number of perceptual illusions experienced by humans. Illusions provide an interesting avenue by which to explore the nature of perception. An observational illusion is an extreme instance in which an environmental stimulus appears differently than it would if judged by the same observer in ideal circumstances. An illusion-generating stimulus or scene necessarily contains some element of perceptual ambiguity. This ambiguity may be due to atypical, exogenous cues. For example, the angle of view may prevent depth cues from being apparent, providing a misleading notion of the size of an object. Just as important, endogenous states, such as an observer's expectations, may induce nonstandard interpretations or biases which conflict with objective conditions.

[3] An interesting alternative approach not explored here is the possible use of transcranial magnetic stimulation for exploring the multisensory role of brain areas in healthy participants (Tunik *et al.*, 2007).

The careful manipulation of ambiguity during perceptual tasks can therefore offer a means of testing how and when baseline perceptual processes both (1) operate under typical circumstances and (2) compensate when standard observational conditions are compromised. More precisely, recording how the perceptual system operates under conditions of uncertainty lays bare both the organism's reliance on external factors and the perceptual tendencies of the developed system. In the case of multisensory illusions, the manipulation of exogenous or endogenous factors is even more important as this can serve as an independent variable which may change the nature of the resulting sensory integration. The reported incidence of multisensory illusions and their physiological correlations becomes a useful dependent measure of the amount of sensory integration which occurs in the system.

The phenomenon of sensory integration in cross-modal illusions has been an alluring topic of research for decades. For example, the *ventriloquism effect* (Howard & Templeton, 1966) is an illusion in which an auditory stream is perceptually linked to visual spatial cues to provide a seamless, unified perception of otherwise spatially separated stimuli. In the prototypical case, when an observer views an animated ventriloquist's dummy and simultaneously hears a spatially separated voice, the observer tends to attribute the sound to the dummy. A more common example of the phenomenon is the experience most people have when watching television or a movie. Voices and sound-effects are automatically attached or attributed to the synchronous movements seen on the screen. In this way, the visual cues are said to 'capture' the auditory stimuli. Such integration is thought to be pre-attentive and not affected by the deliberate allocation of spatial attention (Bertelson *et al.*, 2000). This phenomenon continues to be an active subject of research and has been employed to study audio-visual integration (e.g. Bischoff *et al.*, 2007).

Another cross-modal illusion is the *McGurk effect* (McGurk & MacDonald, 1976). In this much-studied phenomenon, an individual's perception of speech sounds is heavily influenced by visual cues provided by the speaker. When viewing a speaker's face, the visual experience of seeing the syllable /ga/ mouthed while a different syllable /ba/ is spoken can change the observer's auditory experience to that of a third syllable /da/. The influence of visual cues on auditory sensory information is so profound in this case that it changes the identity of the perceived phoneme in real-time. Magnetoencephalogram (MEG) studies (Mottonen *et al.*, 2002; Sams *et al.*, 1991) have demonstrated that changes in visual speech stimulate activity in the auditory cortex, even in the absence of auditory speech. This suggests that the visual signals may prime certain syllables for activation and recognition in the auditory channel. Additionally, in a functional magnetic resonance imaging (fMRI) study of auditory and visual speech that varied in temporal congruity, Jones and Callan (2003) found that auditory cues can activate areas near the occipital-temporal junction thought to be associated with visual motion processing. These findings together suggest that information is passed between auditory and visual areas during speech perception in a dynamic fashion, with each channel affecting the processing in the other.

While the above examples of cross-modal interaction generally focus on visual influences on auditory stimuli, there are also well-known cases of auditory cues impacting visual perception. Among these is *auditory driving* (Gebhard & Mowbray, 1959; Shipley, 1964), in which the rate of a fluttering sound influences the perceived flickering rate of a light. Similar temporal judgment effects of sound over visual input have also been reported in single visual flash and auditory beep presentations (Fendrich & Corballis, 2001). In the Fendrich and Corballis study, when a single flash is preceded by a single beep, subjects reported that the flash occurred earlier in time. If the beep follows the flash, however, the flash is reported later. The tendency for the two stimuli to be drawn together was termed *intersensory temporal locking* (ITL). One possible operational explanation for these phenomena is based on the greater inherent precision of the auditory sense for detecting temporal change. Given the superiority of the auditory system for accurately tracking temporal information, visual sensory data would be weighed less when auditory and visual cues come into conflict. For audiovisual stimuli, perceptual ambiguities would tend to be resolved in favor of the auditory channel.

In the *sound-induced flash illusion* (hereafter referred to as the SFI), when a single visual flash is accompanied by two or more auditory beeps in close temporal proximity, individuals often perceive not one, but two flashes (Shams *et al.*, 2000; 2002). As in the previous cases, the auditory stimuli exert powerful effects on how very simple visual events are perceived. Shams *et al*. (2002) argued that the SFI is qualitatively different than the ITL phenomena. One of the reasons for this claim is that while the single flash could be made to appear twice when accompanied by two beeps (fission), two flashes could not be made to appear as a single flash when paired with a single beep (fusion). However, a later experiment using the flash-illusion phenomena did give rise to this fusion effect (Andersen *et al.*, 2004). Such evidence re-introduces the possibility that the ITL and sound-flash illusion are based on the same underlying multisensory processes. It may be that the second beep in the sound-flash illusion causes the observer to misperceive the single flash onset and offset as two separate events, causing the report of a second illusory flash. Given the simple nature of the stimuli used and the varying effects which can be produced by their temporal and spatial integration, a great deal more remains to be learned from this paradigm.

1.5 Current Project Goals

Through the examination of illusions such as those described here, it is believed that we can learn something important about the psychological phenomena and underlying neural mechanisms involved in more typical multisensory perception. By testing perceptual responses that are known to be at odds with the sensory information provided by the environment, the study of illusions can help us gain new insights as to how our sensory systems normally operate. In that spirit, the current project uses the sound-induced flash illusion to explore the effects of modal attention and exogenous attention on cross-modal sensory integration. Specifically, the project focuses on the following questions:

1) To what degree can multisensory integration be affected by endogenous, task-directed modal attention? (Chapter 2)

2) How does exogenous auditory intensity affect the SFI and other multisensory effects? (Chapters 2, 3, and 4)

3) How do task-directed modal attention and stimulus intensity interact? (Chapter 2, 3, and 4)

4) What electrophysiological indices can serve as markers for late effects of multisensory processing in the SFI? (Chapter 5)

5) What psychophysical methods are best suited to helping us gain a deeper understanding of multisensory processes? (Chapter 6)

The SFI paradigm provides a powerful means of addressing these questions. Illusions are especially useful experimental phenomena as they can be manipulated to produce varying reportable perceptions from a single sensory input state. The equivocal nature of the sensory input tells us something of how the perceptual and cognitive systems are performing when they yield differing results. Illusions also provide exciting data points to inform our models of sensory processing. Such models require more than overt behavioral measures for validation. Along with speeded reaction times (RTs) and accuracy measures, Signal Detection Theory analysis is employed to help determine changes in sensitivity to multisensory stimuli. EEG and ERP recordings and analyses also offer additional measures of the covert processes necessarily involved in illusory perception.

Although the opportunities for empirical exploration afforded by the multisensory illusion paradigm are appealing, our results must be compared to findings from more typical instances of multisensory integration for them to have applicability to a generalized understanding of cross-modal interaction. Comparison of illusory multisensory phenomena with non-illusory multisensory facilitation effects is also addressed in our series of experiments. Additionally, we critically examine theoretical and analytical tools to help interpret our behavioral and EEG/ERP multisensory recordings. More will be said in Chapter 6 regarding the methodological challenges that remain to be solved for measuring multisensory phenomena.

1.6 Conclusions

As with any project, it is hard to know when to stop considering additional conditions, experiments, or ideas to pursue. There is no doubt that many relevant and important elements are missing from the above treatment of multisensory integration. However, by pursuing the objectives within the SFI paradigm outlined here, we hope to make progress in helping to explain how exogenous and endogenous attentional systems may dynamically interact to affect multisensory integration.

The sound-induced flash illusion provides an intriguing window into these issues. However, one of the challenges of this and similar projects is to make our findings relevant beyond this narrow phenomenon. One way we are attempting to do this, as noted above, is by comparing illusory integration with more 'typical' instances of cross-modal pairings within our experiments. An additional opportunity to apply some of our findings is inspired by the setting in which our experiments have been developed. Our laboratory has previously investigated the hypothesis that irregularities in sensory processing and multisensory binding are responsible for some of the features present in autism (Brock *et al.*, 2002; Rippon *et al.*, 2007). EEG recordings have been identified as a means of exploring this specific issue (Brown *et al.*, 2005), and our lab has examined the differences between congruent versus incongruent pairings of pictures and auditory stimuli in autism. One of the challenges in autism research is to find simple perceptual tasks that are appropriate for subjects across the autism spectrum. This requires that the tasks do not presume detailed background knowledge or complex motor ability in participants. The present SFI paradigm was specifically pursued as it is simple and yields clear perceptual differences in most experimental participants (also see van der Smagt, van Engeland, and Kemner, 2007). If some symptoms of autism represent a failure mode of multisensory integration, it will be instructive to consider how the findings from the present experiments fit in with those from other autism studies. While direct testing of autistic populations falls outside the scope of what we could hope to accomplish in the current project, it is hoped that the analytical and experimental methods developed herein may find broader use in such studies.

Outside of possible application in autism spectrum disorder, there are numerous opportunities to apply the SFI paradigm to other developmental and clinical groups.

Multisensory integration abilities are known to change over the course of child and adult development (Brandwein *et al.*,2011; Barutchu *et al.*, 2010), with some enhancements seen in healthy aging populations (Laurienti *et al.*, 2005; Peiffer *et al.*, 2006). In some clinical populations, changes in multisensory integration may be viewed as indicative of a *disconnection syndrome*, in which aberrant cortical connectivity interferes with typical functioning. Disconnection accounts have been implicated in schizophrenia (Friston, 1999) and Alzheimer's dementia (Delbeuck, Van der Linder, & Collette, 2003). The simplicity of the SFI paradigm allows easy adaption to these very different populations. Gaining a better understanding of the potential interplay between exogenous stimulus factors and endogenous attentional systems in multisensory integration should help gain traction in a variety of populations of interest and provide insights into how we all perceive our multisensory world.

Chapter 2: Endogenous attention manipulations of multisensory integration

1. Introduction

When I was about 8 years old, my father took me to my first professional baseball game. It was an exciting scene – filled with new sights, sounds, and smells – and it was difficult to take in all the action. The game began and we were sitting in the outfield seats, when I noticed something odd. The batter hit the ball, but I didn't hear the 'crack' of the bat until about a second later. It was both fascinating and disorienting. The motion appeared somewhat 'broken' as what I heard didn't match what I was seeing. The experience wasn't what I had come to expect from watching on television. When I asked my dad what was happening, he explained that the sound waves traveled slower than the light from the bat, so I would see movements at our distance before I heard therm. This was also true of the satisfying 'slap' of the ball as it hit the catcher's mitt after each pitch and the roar of the crowd after they stood up across the stadium. It was a new experience, but I soon grew used to it. However, if my attention wandered during the game, the 'crack' of the bat would draw me back to see the batter already starting to run to first base as I struggled to find where the ball had been hit.

Of course, our environment is constantly filled with a variety of physical signals that stimulate our sensory systems. These signals travel through different physical mediums and activate the appropriate sensory receptors in a largely independent manner. Some of those signals are caused by the same objects and are perceived as simultaneous events, while others are generated stochastically and remain perceptually distinct. Successfully navigating new situations depends on the ability to quickly and accurately process those signals, determining which are immediately relevant and belong together, and which can be ignored. All creatures in the animal kingdom seem able to integrate information across sensory channels in a way that helps them quickly adapt to their own evolutionary niche (Stein & Meredith, 1993). Multisensory integration thus provides a vital means of binding together disparate environmental sensory signals into whole, regular, perceptible objects that can be acted upon. However, just as integrating signals is critical, so, too, is the ability to selectively attend to a sensory stimulus in a single channel to the exclusion of other, competing modalities (Carrasco, 2011). While the ubiquitous abilities of multisensory integration and selective attention may seem at odds, their cooperative function is of great importance. Through experience, we learn to expect sensory regularities in everyday life that can be vital for survival or simply for enjoying a baseball game.

Multisensory integration, also known as crossmodal integration, is an enduring topic of scientific study. The facilitatory effects of receiving information from multiple senses have been reported for over a century (Todd, 1912). For example, the presence of redundant signals in different modalities can increase the speed at which targets are detected (Diederich & Colonius, 1987; Miller, 1982). Auditory stimuli have been found to improve detection accuracy of visual stimuli (Vroomen & de Gelder, 2000), and also increase sensitivity to sub-threshold visual stimuli (Bolognini *et al.*, 2005). But while many effects of multisensory integration have been well documented, the underlying mechanisms and situational constraints which modulate their emergence remains a fertile area of research (Calvert *et al.*, 2004).

One means of examining multisensory integration has been to exploit changes in perception that occur when sensory information is somehow incongruent across modalities. A common example of this is through the use of the *ventriloquism effect* (Howard & Templeton, 1966). In this well-known illusion, visual motion cues (e.g. articulatory lip movements) can be perceptually linked to a spatially separated stream of audible speech. Here, the close timing of visual cues and sounds cause a shift in spatial perception such that the speech appears to come from the moving lips. While this illusion typically involves visual capture of an auditory stimulus, it has also been demonstrated that an auditory stream can capture visual cues under conditions where the auditory information is more reliable (Alais & Burr, 2004). What this type of sensory binding has in common with speed and accuracy measures is how it implies a *quantitative* shift in perception. That is, the changes caused in these phenomena are a matter of degree in speed, accuracy, or spatial orientation. This is different from *qualitative* effects, in which the binding of sensory information across modalities categorically changes the nature of the thing that is perceived.

Two prime examples of much rarer qualitative effects are the *McGurk effect* and the *sound-induced flash illusion*. In the McGurk effect (McGurk & MacDonald, 1976), speech sounds are combined with visual articulatory cues to create a different perceived sound. In the classic case, a spoken /ba/ sound and mouthed /ga/ motion are combined to produce the perception of /da/. This is a much studied phenomenon that has been found across languages, stimulus sets, and task instructions (reviewed in Alsius *et al.*, 2018). A similarly powerful qualitative multisensory effect is the sound-induced flash illusion (Shams *et al.*, 2000; 2002). In this illusion, when a beep and flash are presented together, a second illusory flash can be perceived if a second beep occurs within 100 ms of the original pair of stimuli. This illusory perception has proven resistant to feedback (Rosenthal *et al*., 2009) and has been extended to show that illusory beeps can also be produced when the original pair is followed by a flash (Andersen *et al*., 2004).

One critical area of multisensory integration that still eludes understanding is how selective attention may alter both quantitative and qualitative effects. Although progress has been made in this domain in recent years (Koelewijn et al., 2010; Talsma et al, 2010; Tang et al, 2016), questions remain regarding the degree to which sensory integration is either automatic in function or susceptible to modification by attention. In favor of the automaticity of integration, Bertelson and colleagues (2000) found that reports of the ventriloquist effect are largely unaffected by the direction of spatial attention. Additionally, Soto-Faraco et al., (2004) used a speeded classification paradigm to suggest that the McGurk effect caused syllabic interference in an automatic, pre-attentive fashion. In a visual search task, Van der Burg *et al.* (2008); found that an uninformative, but synchronously presented auditory 'pip' could cause an automatic 'pop-out' effect for a visual target. These findings suggest that the binding of simultaneously presented stimuli across the senses is an automatic process and largely independent of directed attention.

In contrast to the automaticity hypothesis, others have suggested that some amount of attention may be necessary for multisensory binding to occur (Talsma *et al.*, 2007; Treisman & Gelade, 1980). Supporting this view are reports that the effects of attention may spread across modalities to enhance processing of an attended object (Busse *et al.*, 2007). Alsius and colleagues (2007) also found that the McGurk effect may be reported less when experimental participants were simultaneously engaged in high attentional demand tasks. Using a visuo-tactile version of the sound-induced flash illusion paradigm, Werkhoven *et al.* (2009) reported that simultaneously attending to both sensory channels may enhance integration effects, relative to unimodal attention conditions. Finally, Mishra *et al.* (2010) noted that spatial attention appeared to modulate electrophysiological indices they previously reported to be critically involved in the experience of sound-induced illusory flashes. Together, these reports suggest that the selective manipulation of attentional resources may impact the degree to which stimuli are integrated across the senses.

The following two experiments each examine the effect of selective modal attention on audio-visual multisensory integration. The underlying hypothesis tested is that increasing attention across modalities can result in a greater incidence of multisensory integration. Reports of the sound-induced flash illusion (SFI) are used as the primary dependent variable to assess the degree to which multisensory integration has taken place. A novel bimodal 'matching' task is introduced in the first experiment to manipulate endogenous modal attention in the SFI paradigm. In the second experiment, we extend our new task to include near-threshold auditory stimuli. Comparisons between the two experiments allow us to evaluate the impact of exogenous, stimulus-generated attention on the SFI. We also examine multisensory response facilitation effects within both experiments to determine if they are experienced during the illusory SFI paradigm. This combination of tasks allows for the examination of both quantitative and qualitative multisensory effects within the same experimental paradigm. Finally, we ask whether individual differences in auditory sensitivity may help predict the incidence of experiencing the SFI. Together, these experiments provide a window into how endogenous and exogenous factors may alter multisensory binding perception.

2. Experiment 1

Experiment 1 examined the effect of endogenous modal attention on reports of the sound-induced flash illusion. Endogenous modal attention, defined here as the ability to selfselect the attended sensory modality, was manipulated by varying task instructions over the course of the experiment. Under the *unimodal* visual attention condition, participants were required to attend to only the visual stimuli and perform a visual recognition task. Participants were instructed to respond as quickly and accurately as possible by pressing a button whenever a small white circle flashed twice under a central fixation point. When a trial contained two visual flashes alone or two visual flashes accompanied by two auditory clicks (denoted V_V and AV_AV, respectively), these 'two flashes' were considered correct detections. If a participant responded to trials containing one visual flash and two auditory clicks (denoted AV_A), this false alarm indicated a second illusory flash had been experienced. This numerosity recognition task is similar to those performed in pioneering studies with this paradigm (Shams *et al.*, 2000; 2002). Held common in these experiments is the fact that participants were told to ignore the auditory modality and note the number of flashes present in each trial. A key difference is that previous studies asked participants to choose different responses depending on the number of flashes perceived (i.e. press different buttons depending on the number of flashes detected). Our go/no-go recognition task was used in order to reduce decision and planning time for the motor response component, allowing response time to be a stronger measure of the underlying perceptual processes.

Under our *bimodal* attention condition, the same participants were instructed to attend to both visual and auditory stimuli, responding when the number of stimuli presented in both modalities matched. Trials containing one flash and one click (AV) or two flashes and two clicks (AV_AV) were correct responses. In those trials presenting the AV_A stimuli, as in the unimodal task, participants reporting a 'match' were thought to have experienced an illusory second flash. In this case, a numerosity judgment was required in both attended modalities before a ‘matching’ decision was made. It is also possible that AV_A trials could have induced a ‘match’ response if the stimuli were experienced as a single flash and click (AV), with the second auditory stimulus ‘fused’ with the first audiovisual pair (Andersen *et al.*, 2004). This possibility will be addressed in the discussion.

The *unimodal* and *bimodal* attention tasks described above instruct participants to selectively attend to the visual and auditory modalities in a different fashion for successful completion of the tasks. The same stimuli were presented in both cases and the targets were equally likely across conditions. Our unimodal task, in which the other modality is ignored, is sometimes referred to as a *focused attention* paradigm (Diederich & Colonius, 2004). This is often contrasted with *redundant target* paradigms, where a valid target may appear in either modality. Our bimodal task, which can be characterized as a *conditional bimodal attention* paradigm, requires criteria across both modalities to be satisfied for a task-valid response. This is a novel design and was thought to require greater simultaneous attention to both modalities than the standard redundant target paradigm.

If integration of stimuli across sensory modalities is pre-attentive and automatic, then we should not expect to see a significant change in the number of visual stimuli perceived in the critical illusory trials. Comparison of reported SFIs *between* task attention conditions will help quantify the degree of multisensory integration present in each task. Response time differences also will be explored *within* the unimodal modal task to help elucidate the mechanisms of integration in perceptual illusions when compared with task-valid responses.

If simultaneously attending to both visual and auditory modalities enhances the sensory integration necessary for the sound-induced flash illusion, we would expect more illusory flashes to be reported in the bimodal condition than in the unimodal condition. As attention is thought to effectively increase the sensory gain of attended stimuli (Hillyard *et al.*, 1998; Martinez *et al.*, 2001), this increased baseline activity in both sensory channels may increase the likelihood of integrating auditory and visual stimuli. Such a finding would be consistent with a reported effect of task-directed attention in a numerosity task with illusory visuo-tactile stimuli (Werkhoven, van Erp, & Philippi, 2009).

If audio-visual multisensory integration is pre-attentive, and therefore unaffected by sensory focus, reports of the illusion should be similar across attention conditions. Such a finding would suggest that integrative processes responsible for the sound-induced flash illusion are automatic in nature and not alterable by fluctuating modal attention or task instruction.

A third possibility is that added attention to the number of stimuli present in both modalities may heighten awareness of the illusion in the bimodal attention condition. By forcing participants to determine whether the number of stimuli matched, we required explicit comparison of numerosity judgments across modalities. In this case, enhanced sensory processing and active comparison of the auditory and visual stimuli might make more information available to a post-sensory decision mechanism. Under this latter scenario, reports of the illusion in the bimodal task would decrease relative to the unimodal task and imply that multisensory integration is not entirely automatic.

The manipulation of endogenous task-attention reported here required the use of different behavioral tasks. Critically, both tasks required making correct judgments of the number of visual targets present. While the ‘two flash’ identification task and ‘matching’ task were similar, they did place different requirements on the participants. The explicit direction to compare auditory and visual stimuli in the matching task likely requires the engagement of additional cognitive resources. Therefore, we expected the bimodal task to be slightly more difficult and incur longer reaction times for all participants.

Along with reports of the illusion's frequency across both attention conditions, it was also important to assess relative perceptual sensitivity and biases where possible. For this reason, sensitivity (*d′*) and bias (*c*) were calculated for all conditions of interest (Macmillan & Creelman, 1991). As the illusion represents a false alarm in both tasks (i.e. perceiving AV_A trials as two flashes or matching stimuli), these responses were compared with the observer's affirmative 'hits' in veridical (AV_AV) trials. These signal detection measures allow for a more nuanced analysis of perceptual processing, allowing us to distinguish perceptual change from total response bias. As *d′* is a more sensitive measure of perceptual discrimination, it should provide a better index for the effect of attention on multisensory integration.

A *decreasing d′* is usually interpreted as indicating a reduced ability to accurately differentiate a signal from internal and external noise. Consistent with previous studies of this paradigm (Rosenthal *et al.*, 2009; Watkins *et al.*, 2006; 2007), we suggest that a decrease in sensitivity indicates an *increase* in the multisensory integration responsible for the illusion. The increase in sensory integration interferes with perceptual judgments, making the AV_A and AV_AV trials more difficult to differentiate. To ensure that any changes in responses are not uniquely due to bias inherent in the task, we also report decision criterion *c*. Task instructions are known to affect *c*, so we expect that some differences may exist between the unimodal and bimodal tasks. The critical comparison will be whether changes of *d′* accompany changes in the frequency of the false alarms (i.e. more AV_A trials incorrectly reported as AV_AV trials). If so, we can conclude that a perceptual change, with behavioral consequences, has occurred. This would be in keeping with previous findings in the sound-flash illusion paradigm (McCormick & Mamassian, 2008). However, if only *c* changes, then differential reports of an illusory flash may be attributable to the response biases for each task. Some authors (e.g. Witt *et al.,* 2015) have argued that *c* is a better measure of multisensory integration in the SFI as it quantifies a perceptual bias represented by the illusion. However, we note that *c* is a broad measure of total bias for a response, including strategic and decisional factors involved in the task. Therefore, we believe that *d′* has greater specificity for detecting changes in perception.[4]

In addition to the sound-induced flash illusion, we are also interested in how effective this experimental paradigm is in producing other reported multisensory effects. Of specific interest is the possible emergence of speeded response times (RT) in the unimodal visual task when irrelevant auditory stimuli are also present. In previous experiments (e.g. Hughes *et al.*, 1994; Stein & Meredith, 1993; Stein *et al.*, 1989), it has been found that target detection time can be lowered when a co-occurring stimulus is presented in an unattended modality. Although this has not been typically examined in work with the sound-induced flash illusion, Fiedler and colleagues (2011) have reported that responses to double flashes (V_V) are faster than single flashes (V). This is classified as a redundant signals effect (RSE), as their task was to respond to any visual stimulus. In our experiment, we expect to find a decrease in response times for the AV_AV condition, when compared with V_V stimuli in the 'two flash' unimodal task. If RT effects are seen in the present experiment, this would further establish that the paradigm is useful for simultaneously examining quantitative (RT) and qualitative (sound-induced flash illusion) multisensory effects. As RT response facilitation has been well-characterized and discussed within the context of competing models of multisensory

[4] This is discussed further in chapter 6.

interaction (Colonius & Diederich, 2004), this could also offer a point of departure for theoretical elaboration of the neural architecture responsible for the SFI.

2.1. Methods

2.1.1 Participants

Fifteen healthy adults (10 women, 5 men; mean age 22.2 years) participated in our first experiment after giving written informed consent, in accordance with the University of California, San Diego Human Research Protections Program. All participants had normal or corrected-to-normal vision, reported normal hearing, and were naïve to the purpose of the study. Individuals received course credit or monetary compensation for participation and were debriefed following completion of the study.

2.1.2 Apparatus and stimuli

The experiment was conducted in a darkened, sound-attenuated chamber where the only source of light was the 21-inch computer display monitor. Eight different stimulus trials were presented. Visual stimuli consisted of a uniform white circle 'flash' subtending 2 degrees of visual angle. The circle was presented 10 degrees below fixation for approximately 10 ms (one screen refresh at 100 Hz). Peripheral presentation of the visual stimuli follows Shams and colleagues' (2002) finding that the illusion is strongest outside central fixation. When a single flash (V) was followed by a second flash (V_V), the stimulus onset asynchrony (SOA) was 70 ms.

The auditory stimulus was a 1,000 Hz 'beep' played for 10 ms. The auditory beep was presented at 60 dB (A) SPL, measured from the source. Sounds were provided in a free field by speakers closely flanking the video monitor. The SOA between double auditory presentations (A_A) was 70 ms. When a single beep and flash were presented together (AV), their onsets were separated by 10 ms, with the auditory stimulus presented first. Illusion-inducing trials in which two beeps were presented with a single flash (AV_A) had an auditory stimulus SOA of 70 ms to match the A_A trials. The illusion-inducing trial presentation is illustrated in figure 2.1. All trials were presented in a pseudo-random fashion, separated by intervals of randomly jittered 1,400 – 1,900 ms.

All responses were recorded via button press on a computer mouse. Participants used the index finger of their dominant hand, or the hand they reported using a computer mouse most frequently.

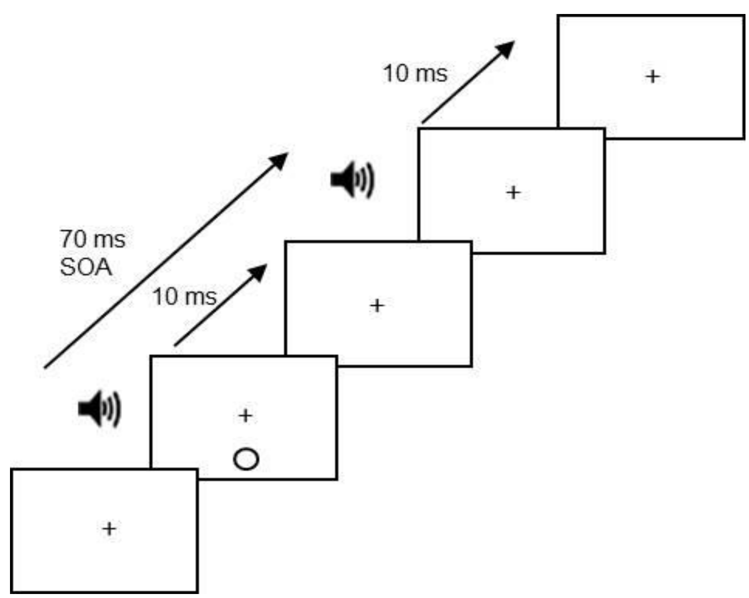

Figure 2.1 Illusion stimulus (AV_A) overview. Participants begin with a center fixation point that is always present. A 1,000 Hz tone is presented for 10 ms. A white circle appears 10 degrees below fixation and disappears after 10 ms. 70 ms after onset of the first tone, a second identical tone is played. The participant may or may not perceive a second flash at this point. Participants indicate whether 'two flashes' (unimodal attention) or 'match' (bimodal attention) was perceived via single button press. The inter-stimulus trial interval (ITI) is randomly jittered from 1,400 – 1,900 ms.

In addition to the five trial types above, we included a single beep (A), two beeps and two flashes (AV_AV), and an empty trial. The empty trial was used following Mishra *et al.* (2007) as a means of controlling for some EEG/ERP effects in comparative subtractions. EEG/ERP data were collected for all participants, but those results will be reported separately in chapter 5.

2.1.3 Procedure

The experiment was divided into two conditions, differing in task-attention instructions. In the visual attention condition, participants focused on a central fixation cross and were instructed to attend to the visual stimuli only, ignoring all auditory stimuli. In this go/no-go focused attention paradigm, the task was to respond via button press as quickly and accurately as possible any time two flashes appeared on the screen. No response was to be given when the target was absent. All participants performed three attend-visual 'two flash' blocks. In each of these blocks, 30 instances of each of the eight stimulus conditions were presented, for a total of 240 trials per block. Stimuli were presented in pseudo-random order, such that no stimulus was presented twice in a row. Over the three blocks, this yields a total of 90 trials for each of the eight stimuli in the attend-visual condition.

In the second part of the experiment, participants were again asked to focus on a central fixation cross. However, they were instructed to attend to both the auditory and the visual stimuli and respond as quickly and accurately as possible whenever the number of auditory and visual elements matched in number. This was a conditional bimodal attention paradigm, requiring criteria to be satisfied across both modalities for a successful response. As in the first condition, each participant completed three such blocks of 240 trials each, but with the audio-visual ‘match’ task. In total, participants viewed 720 trials per attention condition.

An equal number of valid targets were presented in the visual and bimodal attention conditions. In the visual-only ‘two flash’ task, two of the eight stimulus types were valid (V_V and AV_AV). In the bimodal attention ‘matching’ task, two stimulus types (AV and AV_AV) were valid. Given that one stimulus was a ‘blank’ trial in which no auditory or visual stimuli were presented, the chance of encountering a valid target was approximately 28.6% in any given trial. In both conditions, the illusion-inducing stimulus (AV_A) could be incorrectly perceived as a target.

In this experiment, all participants performed the focused visual attention task first, followed by the conditional bimodal matching task. While this may introduce order effects, it was felt that participants might find it difficult to fully ignore the auditory channel if the bimodal matching task was done first. Prior to each task, participants took part in a practice block until they felt comfortable with the task instructions. Throughout the experiment, participants were visually monitored via camera to ensure proper visual fixation, vigilance, and task response compliance.

2.2. Results

In the analyses below, we examine the impact of modal attention on both quantitative and qualitative multisensory effects. Our dependent measures were frequency of response, reaction times (RTs), and signal detection measures. These are reported in separate sections below, with specific accounts of sensitivity and bias measures provided for clarity. In accord with previous work (e.g. Mishra *et al*., 2007; Rosenthal *et al.*, 2009; Violentyev *et al*., 2005), repeated measures ANOVAs and planned pairwise *t*-tests are used to evaluate our manipulations, unless otherwise noted. We recognize that using ANOVAs and *t*-tests may be problematic with reported frequencies as the data are bounded by 0 and 1. We note that most of the critical illusion data lies in the center of the distribution and not near the end points. The transformed data in our Signal Detection measures will provide additional support to findings using the proportional data.

2.2.1 Frequency and reaction time

The first planned analysis was to determine if irrelevant, unattended auditory stimuli affected reaction times (RTs) to visual targets in the unimodal attention blocks. Such a facilitation effect due to the mere presence of an irrelevant auditory signal may improve the speed at which the ‘two flash’ targets were detected. The comparison of reaction times between V_V (*M* = 494 ms, *SD* = 69 ms) and AV_AV trials (M = 477 ms, SD = 71 ms), did show a moderate, significant ~17 ms decrease in RT speed for bi-modal stimuli [two-tailed pairwise *t*-test, $t(14) = 2.58$, $p < .05$, Cohen’s $d = 0.67$]. Correct detection of two flashes in the AV_AV condition (*M* = 97.8%, *SD* = 3.5%) appeared more accurate than V_V trials (*M* = 91.3%, *SD* = 14.1%); however, this difference did not reach statistical significance [two-tailed pairwise *t*-test, $t(14) = 2.11$, $p = .053$]. Results are summarized in Table 1. Given the variable response accuracy in the V_V trials, a difference in detection accuracy did not emerge. However, our initial RT finding supports the notion that typical multisensory response speed facilitation effects do occur in the current paradigm.

Table 2.1. Mean behavioral performance for reporting two flashes or matching beeps and flashes.

Stimulus	Mean percentage reporting 'two flashes' (SEM). Visual only.	Mean RT in ms. (SEM). Visual only.	Mean percentage reporting 'match' (SEM). Bimodal.	Mean RT in ms. (SEM). Bimodal.
V	21.9 (5.3)	667 (28)	0.5 (0.1)	217 (94)
V_V	91.3 (3.6)	494 (18)	3.2 (0.7)	329 (57)
AV	8.2 (3.3)	367 (83)	92.4 (2.3)	683 (25)
AV_A	71.3 (6.8)	559 (27)	75.8 (5.8)	662 (30)
AV_AV	97.8 (0.9)	477 (18)	94.9 (2.2)	554 (29)

We next wanted to determine whether endogenous attention affected the reports of targets and sound-induced flash illusions in stimuli common to both tasks. Accordingly, the proportion of responses indicating that participants experienced two flashes was analyzed with repeated measures ANOVA with factors stimulus (AV_A, AV_AV) and attention (unimodal, bimodal). Responses in the AV_AV condition are hits, while those in the AV_A are false alarms due to the illusion. Analysis revealed a main effect of stimulus condition [$F(1,14) = 15.5$, $p < .01$, $\eta^2_{partial} = .526$], and a trend for an interaction [$F(1,14) = 3.73$, $p = .074$; see figure 2.2]. Reports of the illusion increased numerically in the bimodal matching task (M = 75.8%, SD = 22.5%) versus the unimodal 'two flash' responses (M = 71.3%, SD= 26.2%), but did not approach significance [planned two-tail pairwise t-test, $t(14) = 1.18$, p = n.s.].

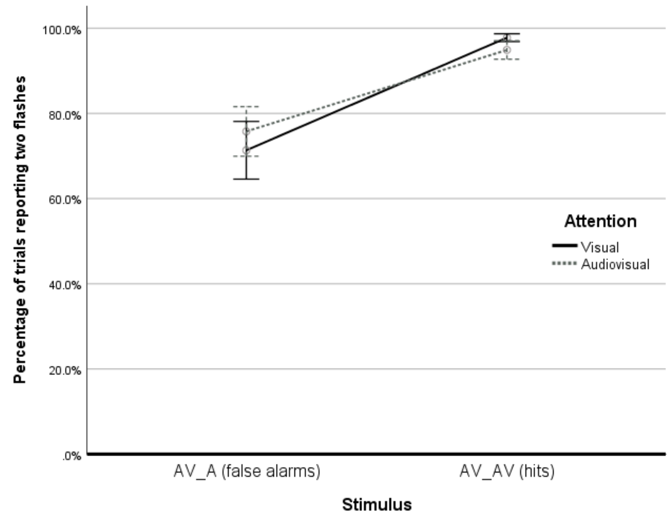

Figure 2.2. Reports of two flashes by task attention. There is a main effect for stimulus type, but not attention. The interaction did not achieve significance (p = .074). Error bars +/- 1 SEM.

Our initial analysis suggests that task attention did not significantly modulate reports of the illusion. As the present experiment extends the paradigm for testing the sound-induced flash illusion, we wanted to evaluate the relative times necessary to complete the two experimental tasks. A two (stimulus: AV_AV hits, AV_A false alarms) by two (attention: unimodal, bimodal) repeated measures ANOVA of reaction times found a main effect of stimulus condition, with reaction times to veridical instances of double flashes being approximately 105 ms faster than to induced illusory flashes [$F(1,14) = 40.0$, $p < .01$, $\eta^2_{partial} = .741$; see figure 2.3]. A main effect for task attention (unimodal vs. bimodal) was also found, with faster reaction times in the unimodal condition [$F(1,14) = 46.3$, $p < .01$, $\eta^2_{partial} = .768$].

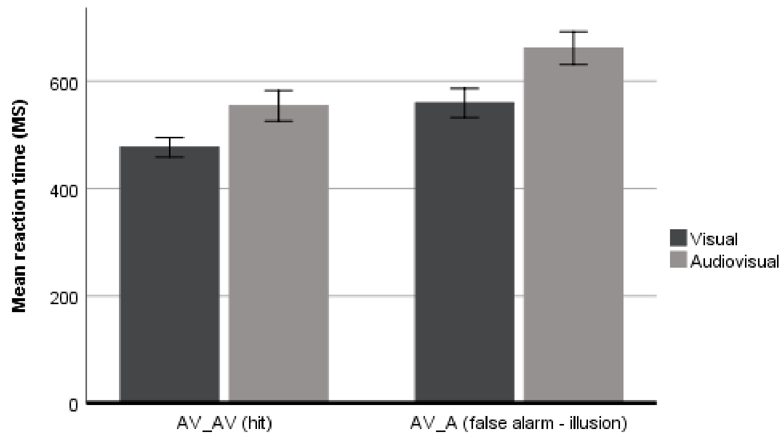

Figure 2.3. Reaction times for attention condition by stimulus type. This compares illusory false alarms (AV_A) and veridical hits for paired stimuli (AV_AV). Error bars +/- 1 SEM. The visual task was significantly faster than the audiovisual matching task. Reaction times to veridical stimuli were also faster than false alarms.

The above analyses suggest that (1) veridical responses tend to be faster than responses involving an illusory perception and (2) that the unimodal ‘two flash’ task takes less time than the matching task. Regarding the first point, we were curious to know if the asynchrony of the mismatching stimuli in the AV_A trials caused a delay in the sensory processing or decision-making of participants. We found that correct responses in the unimodal V_V condition were also faster than false alarms to the AV_A condition [*post hoc* two-tailed pairwise *t*-test, $t(14) = 3.5$, $p < .01$, Cohen’s $d = .90$]. Given that AV_AV and V_V stimuli were both responded to faster than AV_A stimuli, some additional processing seems to take place in the mismatching trials. The second point indicates a potential difference in the cognitive resources necessary to complete the two tasks in our experiment. By hypothesis, part of the difference lies in the allocation of attention across modalities. However, when we interpret the data above, we must be mindful that additional cognitive factors may also be in play.

To help isolate perceptual differences in task performance and determine whether the reaction time differences were due to sensory processing or decision factors, we next performed a signal detection analysis of the above data.

2.2.2 Sensitivity (d')

Our initial analysis using Signal Detection Theory (Macmillan & Creelman, 2005) measures examined the effect of irrelevant auditory stimuli on visual discrimination in our unimodal "two flash" task. To address this issue, we first ascertained the unimodal d' value for all participants when correctly identifying a visual double flash (V_V) target. False alarms were defined as those single flashes (V) erroneously reported as double flashes. d' was calculated as follows (Macmillan & Creelman, 2005):

$$d' = z(H) - z(FA), \tag{1}$$

where the z score for false alarm rates (FA) are subtracted from the z score for hits (H) in the above stimuli. When response rates of 0 or 1.0 were reported, these were replaced with values 0.5/*n* and (*n* – 0.5)/*n*, respectively, where *n* is the number of trials of that type. This assumes that, if twice as many stimuli had been presented, at least one different response would have been reported. While this correction may introduce a bias in sensitivity (Miller, 1996), it is the most common approach to this problem (Stanislaw & Todorov, 1999).

This unimodal d' measures the ability to perceptually discriminate a single flash from two flashes, without any influence from the auditory modality. These unimodal d' values were compared to the d' values of our multimodal targets. In our unimodal 'two flash' task, correct

AV_AV targets were identified as hits, while responses to illusory AV_A trials were labeled as false alarms (see Table 2). Analysis indicates greater sensitivity in the unimodal trials (mean d' = 2.72), when compared to the multimodal trials (mean d' = 1.45). This significant result [planned two-tailed, pairwise *t*-test, $t(14) = 5.61$, $p < .01$, Cohen's $d = 1.45$], is in keeping with the findings of Violentyev and colleagues (2005) who reported a similar reduction in sensitivity for multisensory stimuli when compared to unimodal trials in a visuo-tactile version of the same illusory paradigm.

Table 2.2. Categories employed in d' sensitivity analysis for multimodal stimuli.

	Response: "Two flashes"	No response
Stimulus: AV_AV	Hit	Miss
Stimulus: AV_A	False Alarm	Correct Rejection

Response bias was quantified using criterion *c*, a measure of bias that is statistically independent of d'. This was calculated as follows (Macmillan & Creelman, 2005):

$$c = -0.5\ [z(H) + z(FA)]. \tag{2}$$

For our unimodal task, we found a significant difference in response criterion *c* [planned two-tailed, pairwise *t*-test, $t(14) = 6.48$, $p < .01$, Cohen's $d = 1.67$], with the unimodal V_V visual stimuli ($M = -0.43$, $SD = 0.58$) seeing a more conservative response tendency than bimodal AV_AV stimuli ($M = -1.46$, $SD = 0.57$). This suggests that the mere presence of two auditory beeps increased the bias towards responding that two flashes were seen. There is some question as to how this should be interpreted within the SFI paradigm (Witt, Taylor, Sugovic, and Wixted, 2015). For the moment, we can say that the increased tendency to respond in the presence of task-irrelevant auditory beeps seems to accompany a change in perceptual sensitivity to visual flashes. This is discussed further in the general discussion and treated extensively in chapter 6.

Did simultaneously attending to both auditory and visual modalities decrease the ability to distinguish double flashes from single flashes in the presence of sounds? The comparison of d' measures for the multimodal AV_AV trials across task attention conditions, using AV_A as the false alarm, showed a moderate, significant difference [two-tailed pairwise *t*-test, $t(14) = 2.40$; $p < .05$, Cohen's $d = 0.62$]. In the visual attention condition, mean $d' = 1.45$ ($SD = 0.88$), while in the bimodal attention condition mean $d' = 1.06$ ($SD = 0.78$). The lowered sensitivity in the bimodal task is consistent with the notion that the multisensory interaction is greater in the matching task. No difference was observed in the response criterion c for the above responses [planned two-tailed, pairwise *t*-test, $t(14) = 0.67$, p = n.s.]. Mean $c = -1.46$ ($SD = 0.57$) in the unimodal attention task, compared to a mean $c = -1.41$ ($SD = 0.58$) in the bimodal attention task. There is some question as to how this should be interpreted in the SFI paradigm (Witt *et al.*, 2015). However, the present result suggests that total bias did not shift noticeably between tasks for these stimuli.

2.3. Discussion

In our first experiment, we found that the current go/no-go paradigm elicits multisensory effects in the form of speeded responses to 'two flash' stimuli when irrelevant auditory stimuli were also present. This is consistent to the findings of a related experiment (Fiedler *et al.*, 2011), which found a similar reaction time decrease when using a speeded response to any visual stimulus in a simple reaction time task. This extends the sound-induced flash illusion paradigm and helps establish the usefulness of the typical SFI stimuli for probing both quantitative and qualitative multisensory effects within the same study. The experiment also successfully replicated the SFI in a new experimental task, the *conditional bimodal attention* 'matching' task. Participants were asked to explicitly compare the number of stimuli appearing in both modalities, rather than to ignore the auditory channel. In this new bimodal attention task, participants reliably reported the illusion at a high frequency. Reports of the SFI within this task further demonstrate the strength of this illusion under conditions that should highlight the mismatch in the stimuli. Together, these two results extend the ways in which the SFI paradigm can be used to study multisensory integration.

Most importantly for the current study, we also found that directing attention to both auditory and visual stimuli seemed to decrease the ability to distinguish two flashes from a single flash when accompanied by two auditory beeps. This decrease in visual sensitivity, as measured by d', supports the hypothesis that endogenous, task-directed attention can modulate the amount of multisensory integration thought to be responsible for the illusion. To our knowledge, this is the first demonstration that task-directed modal attention can modulate the processes responsible for the SFI. These data imply that multisensory integration is not purely automatic, but may be affected by dynamic allocation of cognitive resources according to task demands.

The mechanism by which modal attention apparently impacts sensory integration remains unclear. In the simplest account, attending to the auditory channel may increase the gain of that signal, making later integration of stimuli in both channels more likely. In a slightly more complicated scenario, the focused attention visual task may allow participants to gate auditory information, reducing its contribution to multisensory integration compared to the bimodal attention task. It is also possible that increased auditory gain due to attention in the bimodal task might enhance the internal noise in the perceptual system. The greater noise could decrease the signal-to-noise ratio, decreasing the functional sensitivity required to correctly perform the numerosity task. This loss of functional sensitivity to the target stimuli may then, in some cases, increase the likelihood of perceiving the illusory flash.

Our experimental results, and possible interpretations, are somewhat tempered by the fact that the loss of sensitivity did not reliably increase the likelihood of reporting the sound-induced flash illusion in the bimodal task. While d' is a more sensitive measure of perceptual discriminability and should be better able to isolate changes in the multisensory integration thought to be responsible for the illusion, our initial prediction had been that participants would report the illusion more often in the bimodal attention task. To help explain the seeming inconsistency in our findings, we examined some of the factors involved in our two tasks. We note that participants performed exceptionally well when reporting two flashes in the unimodal AV_AV trials (M = 97.8%, SD = 3.5%) and identifying these same trials as matching pairs in the bimodal attention task (M = 94.9 %, SD = 8.4%). Additionally, the SFI was reported at a mean rate of 71.3% and 75.8% in the unimodal and bimodal tasks, respectively. This illusion rate appears much higher than has been reported in other studies. For example, Mishra and colleagues (2007) reported that participants saw the illusion an average of 37% in similar trials. Similarly, Watkins *et al.* (2006) found that illusory flashes were reported in 32% of comparable trials. Our high incidence of the flash illusion in both attention conditions may indicate that the stimulus parameters have maximized multisensory integration effects beyond the ability of selective modal attention to greatly modulate reports to the illusion. If this is the case, the absence of a significant effect, as measured by frequency of response, may be partially due to ceiling effects in integration.

Reviewing our debriefing notes, we also found that participants typically reported that the auditory stimuli were very clear and much easier to identify than the visual stimuli. It is possible that the intensity of the auditory stimulus acted as a strong exogenous cue that affected attention, making it difficult to completely ignore. Given these anecdotal reports, we reasoned that a difference in stimulus salience may have masked some of the potential effect of endogenous modal attention. That is, because the auditory stimulus was perceptually much stronger than the visual stimulus, the auditory beeps may have directed attention in the 'two flash' unimodal task. Consciously attending to these stimuli in the bimodal 'matching' task may have, therefore, added little perceptual 'gain' to the signal. This ceiling effect for stimulus salience, and any related difficulty in ignoring the auditory stimulus in the unimodal visual task, could reduce any measured difference in sensitivity due to endogenous modal attention.

One additional factor that may have come into play is an order effect caused by increasing fatigue over the course of the experiment. Since we were collecting EEG data along with our motor behavior, additional experimental time was spent applying the electrode cap and making adjustments throughout the experiment. Total time in the experiment exceeded two hours for each participant. The 'matching' bimodal attention condition always came second in our experiment, so any effects of fatigue that accrued would be largest for this condition. This may have diminished vigilance and reduced effect size in the experiment.

Finally, as we employed two different tasks in order to manipulate modal attention, it is also possible that unaccounted changes in cognitive demands could mask some underlying change in multisensory effects. Response times to the illusory flash were significantly faster in the unimodal task than the multimodal matching task, suggesting that the latter task was slightly more difficult or cognitively complex. To be sure, the two tasks in the current experiment demand different perceptual comparisons and may require diverging neurocomputational resources for their completion. In the simpler unimodal task, participants are only required to respond when they see two flashes. Any influence of the auditory signal on a visual judgment, such as being induced to experience an illusory flash in the AV_A trials, is purely incidental and unrelated to the task itself. In the matching task, the observer must additionally compare judgments from independent auditory and visual sensory streams. This further step could have reduced the likelihood of experiencing the SFI if the illusion was mainly a late decision-based effect. However, given the decreased ability discriminate between AV_AV and AV_A in the matching task relative to the unimodal task, this alternate hypothesis is not supported by the current evidence. We also noted that total response bias did not change between tasks, suggesting that differences in task requirements did not directly impact general response tendencies in the participants. Though we don't see differing task demands as an immediate limitation, we recognize that it is a potential source of interference in the current study.

To address the issues identified above, we performed a follow-up experiment to control for perceived stimulus intensity and attempt to reduce ceiling effects across tasks. The design follows the procedure used in experiment one, but employs an auditory stimulus that was titrated to be near-threshold for each participant. Making the auditory stimulus closer to the visual stimulus in lower perceptual salience should reduce potential masking effects of exogenous alerting cues and maximize any impact of endogenous attentional gain. We also replicated the experiment without EEG recordings, reducing the amount of time in the experiment and associated fatigue. While the order of conditions remains the same, the much shorter time in the experimental chamber (less than 1 hour per participant), should alleviate the impact of waning vigilance.

3. Experiment 2

This follow-up experiment examined the influence of endogenous modal attention on multisensory integration when using auditory stimuli of low perceptual salience. It used the same design and equipment as experiment 1, but first found the lowest intensity at which each participant could detect the 1,000 Hz auditory stimulus with at least 85% accuracy. At this lower subjective intensity, the auditory signal should be easier to ignore when performing the unimodal visual ‘two flash’ detection task. Compared with our previous experiment, any exogenous alerting or orienting induced by the auditory stimulus should be minimized. Additionally, when engaged in the bimodal ‘matching’ task, participants must more closely attend to the near-threshold auditory signal to maintain a high level of performance. The use of near-threshold stimuli maximizes any selective auditory sensory gain due to endogenous attention and provides for a more sensitive test of its effect on multisensory integration. Finally, because we are gathering information about the auditory sensitivity of each participant, this experiment will also allow us to explore one possible reason for the variability often reported in experiencing the SFI (Mishra *et al*., 2007; Watkins *et al*., 2006). In an exploratory analysis, we will evaluate whether individual auditory sensitivity may be related to susceptibility to the illusion.

In a similar experiment using near-threshold auditory stimuli, Andersen and colleagues (2004) found that the sound-induced flash illusion may be reported when employing low intensity auditory stimuli. While the illusion was not found to significantly decrease in frequency as auditory intensity decreased, Andersen and colleagues did report a trend in that direction. It is important to note that the near-threshold auditory stimulus in their study was set at a single value (10 dB) for all participants. Only afterwards was it determined by the authors that some participants were unable to detect the auditory stimulus. The present experiment adjusted near-threshold auditory stimuli for each individual instead of using a single intensity value for all participants. The study by Andersen and colleagues also delivered the auditory stimuli via headphones. In the current experiments, auditory stimuli were presented in free-field, from speakers positioned next to the video monitor.

If multisensory integration is largely pre-attentive and not affected by endogenous task-attention, we should see similar sensitivity to the sound-induced flash illusion as measured by their reported frequency and d' in both attention conditions. If multisensory integration can be modulated by modal attention, as suggested in experiment 1, we would expect both increases in the illusion and decreases in visual sensitivity in the bimodal condition as compared to the unimodal condition. In experiment 1 we also considered the possibility that active comparison of the number of auditory and visual stimuli in the bimodal 'matching' task might result in fewer reports of the illusion. The hypothesis was that added attention might overcome initial response biases, leading to more veridical judgments (i.e., uncovering the illusion). No evidence of this was found, but it may be possible that the lower auditory intensity could decrease a latent response bias. This would result in fewer illusions reported in the bimodal attention condition.

Motivating this follow-up experiment is the hypothesis that lowering the intensity of the auditory stimuli will better enable participants to dynamically allocate resources during the unimodal 'two flash' task. At lower intensities, participants should be able to ignore auditory stimuli, thereby reducing the amount of multisensory integration responsible for the SFI. Under the hypothesis that auditory stimuli serve as an exogenous alerting or orienting cue, it is also possible that other multisensory effects will be reduced. If this is the case, lowering the intensity of the auditory stimulus could lower the total incidence of illusory flashes reported in AV_A trials. Similarly, there may be a decrease in the reaction time advantage in AV_AV trials when compared to V_V trials found in experiment 1.

3.1. Methods

3.1.1 Participants

In our second experiment, twenty healthy adults (11 women, 9 men; mean age 21.8 years) participated after giving written informed consent, in accordance with the University of California, San Diego Human Research Protections Program. None of the participants took part in the previous experiment and all were naïve to its purpose. Two participants (1 male, 1 female) were excluded due to developmental visual problems reported after completion of the experiment. All remaining participants had normal or corrected-to-normal vision and reported normal hearing. Individuals received course credit or monetary compensation for participation and were debriefed after finishing the experiment.

3.1.2 Apparatus and stimuli

Experiment 2 employed the same equipment and stimulus categories used in experiment 1. However, the 60 dB SPL auditory stimulus was replaced by a near-threshold stimulus. The intensity of the auditory stimulus was set for each individual according to the procedure described below. Electrophysiological recordings were not collected as in experiment 1, significantly reducing the time each participant spent performing the experiment.

3.1.3 Procedure

The experimental procedure from experiment 1 was largely duplicated in experiment 2. A key difference was an additional test to find the intensity at which each participant would successfully detect a 1,000 Hz auditory stimulus, played for a duration of 10 ms, approximately 85% of the time. This was done in two steps. First, we used an adaptive staircase procedure (reviewed by Leek, 2001) to determine each participant's 50% performance level for the 1,000 Hz tone played from speakers. The threshold was established as the average of 6 runs. Each run was stopped after step-size was less than 1 dB or following 8 reversals. We then constructed a stimulus set around this average value by adding three louder stimuli, three quieter stimuli (each separated by 1 dB), and one silent catch trial. For example, if the threshold determined by the staircase method was 25 dB(A) SPL, the stimulus set would consist of sounds set at 22, 23, 24, 25, 26, 27, 28 and 0 dB SPL. These 8 stimuli were then randomly presented to the participant, 30 times each, using the method of constant stimuli. All auditory stimuli were presented in a free-field, from speakers flanking the screen used in the main experiment. Participants responded via button press only when they heard a sound. The value for our experimental auditory stimulus was set at the lowest level at which each individual participant reported hearing the sound at least 85% of the time. The average resulting sound intensity for participants was 24 dB(A) SPL, measured at the speaker. If a consistent auditory intensity level could not be identified, the individual was excluded from the experiment. No participants failed this criterion.

As in experiment 1, our second experiment was divided into two endogenous attention conditions, determined by task instructions. In the first condition, participants were instructed to respond as quickly and accurately as possible whenever two flashes appeared on the screen. In this go/no-go paradigm, all auditory stimuli were to be ignored. In the second conditions, participants were instructed to attend to both sensory channels, responding as quickly and accurately as possible anytime the stimuli presented in both channels matched in number. As before, the number of targets presented in each condition was identical. Participants were given practice blocks before each condition until they expressed comfort with the task.

3.2. Results

3.2.1 Frequency and reaction times

As was found in experiment 1, faster speeded responses were seen when concurrent auditory stimuli were presented in the visual task (see table 3 for a list of response times and accuracy rates). When responding to ‘two flashes’ in the unimodal visual task, participants were significantly faster when an irrelevant auditory signal was paired with the target visual stimuli (M = 454 ms, SD = 67 ms) versus trials in which visual stimuli were presented alone [M = 471 ms, SD = 72 ms; $t(17) = 2.19$, $p < .05$, Cohen’s d = 0.52]. Accuracy also increased in the multimodal (AV_AV) stimulus presentation trials when compared to unimodal (V_V) trials [M = 95.4%, SD = 4.7% versus M = 89.6%, SD = 8.5%, respectively; $t(17) = 4.31$, $p < .001$, Cohen’s d = 1.02].

We note that the above response facilitation for speed in the presence of an irrelevant sound was similar in magnitude to that reported in experiment 1. The response time advantage for AV_AV over V_V in experiment 1 had a relative effect size (as measured by Cohen’s d)

of 0.67, compared to an effect size of 0.52 in experiment 2. However, whereas the gain in accuracy was not statistically significant in experiment 1, it was significant and of fairly large magnitude (Cohen's d = 1.02) in experiment 2.

Table 2.3. Mean behavior for responding to two flashes or two click & two flash matching stimuli.

Stimulus	Mean percentage of trials reporting visual 'two flashes' (SEM).	Mean visual RT in ms. (SEM).	Mean percentage reporting bimodal 'match' (SEM).	Mean bimodal RT in ms. (SEM).
V	14.7 (2.4)	577 (32)	6.9 (1.5)	635 (36)
V_V	89.6 (2.0)	471 (17)	9.1 (2.0)	554 (40)
AV	14.1 (2.5)	438 (48)	80.4 (2.4)	641 (22)
AV_A	47.6 (6.4)	537 (27)	83.5 (3.2)	566 (25)
AV_AV	95.4 (1.1)	454 (16)	91.7 (1.5)	530 (24)

A two (stimulus: AV_AV, AV_A) by two (attention: visual, audiovisual) repeated measures ANOVA was performed on those responses indicating that participants experienced two flashes (hits in AV_AV and illusions in AV_A). This analyses revealed a main effect of stimulus condition [$F(1,17) = 49.0$, $p < .001$, $\eta^2_{partial} = 0.742$], a main effect of attention [$F(1,17) = 22.7$, $p < .001$, $\eta^2_{partial} = 0.572$], and an interaction between stimulus condition and attention [$F(1,17) = 50.8$, $p < .001$, $\eta^2_{partial} = 0.749$; see figure 2.4]. Inspecting figure 2.4, it appears that our main effects were driven by the interaction of attention and stimulus type.

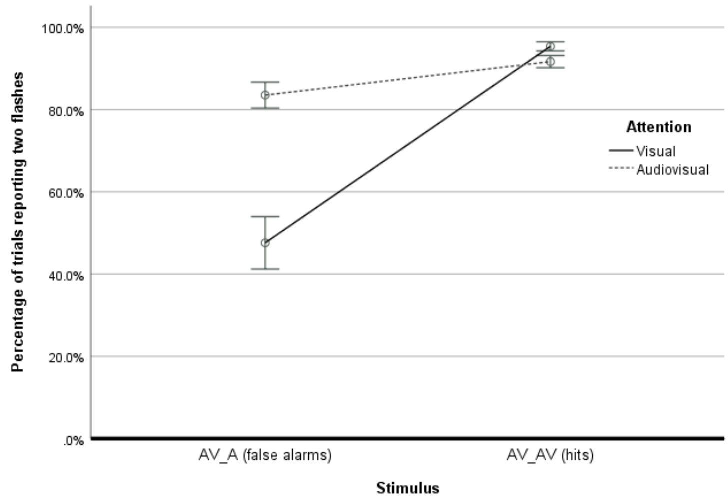

Figure 2.4. Repeated measures ANOVA found main effects for stimulus and task attention, and an interaction. The lower incidence of the sound-induced flash illusion in the visual attention task appears to drive all effects. Error bars +/- 1 SEM.

The sound-induced flash illusion was reported significantly less often in the unimodal visual condition than in the bimodal attention condition. On average, participants indicated the illusion was present in the 'two flash' visual task in 47.6% (*SD* = 27.0%) of the AV_A trials, while in the 'match' task, it was reported in 83.5% (*SD* = 13.4%) of AV_A trials. This was a significant and large result [two-tailed pairwise, *t*-test, $t(17) = 6.01$, $p < .001$, Cohen's $d = 1.42$] and suggests that task-directed endogenous attention did affect the rate at which the sound-induced flash illusion was experienced. Compared to experiment 1, this suggests that participants are better able to ignore the auditory stimuli in the focused visual attention condition when those stimuli are near-threshold.

For comparison, participants erroneously reported seeing two flashes in the single flash (V) and single flash and beep (AV) trials M = 14.7% (SD = 10.2%) and M = 14.1% (SD = 10.8%) of the time, respectively. The incidence of reported sound-induced flash illusions (AV_A) in the visual-only task was far greater than the false alarms reported in V trials [two-tailed pairwise t-test, $t(17) = 6.54$, $p < .001$, Cohen's $d = 1.54$] and AV trials [two-tailed pairwise t-test, $t(17) = 6.84$, $p < .001$, Cohen's $d = 1.61$].

As with analysis of the frequency of responses, analysis of reaction times involved a two (stimulus: AV_AV, AV_A) by two (task-attention: visual, audiovisual) repeated measure ANOVA. We found main effects for both stimulus condition [$F(1,16) = 25.4$, $p < .001$, $\eta^2_{partial} = 0.613$] and task-directed attention [$F(1,16) = 7.82$, $p < .05$, $\eta^2_{partial} = 0.328$]. The interactions did not reach significance [$F(1,16) = 3.697$, $p = 0.072$]. One participant was excluded from this analysis as they did not have any false alarms to the AV_A stimulus in the visual 'two flash' task. As seen in figure 5, participants were faster when responding in the visual-only attention condition than in the bimodal matching task. Participants were likewise faster when reporting veridical double flashes than responding to the illusion trials. This is consistent with experiment 1 and suggests that additional processing is required when responding to illusory stimuli. Furthermore, as a more complex cognitive task, the bimodal task requires more time to complete than the visual 'two flash' detection task.

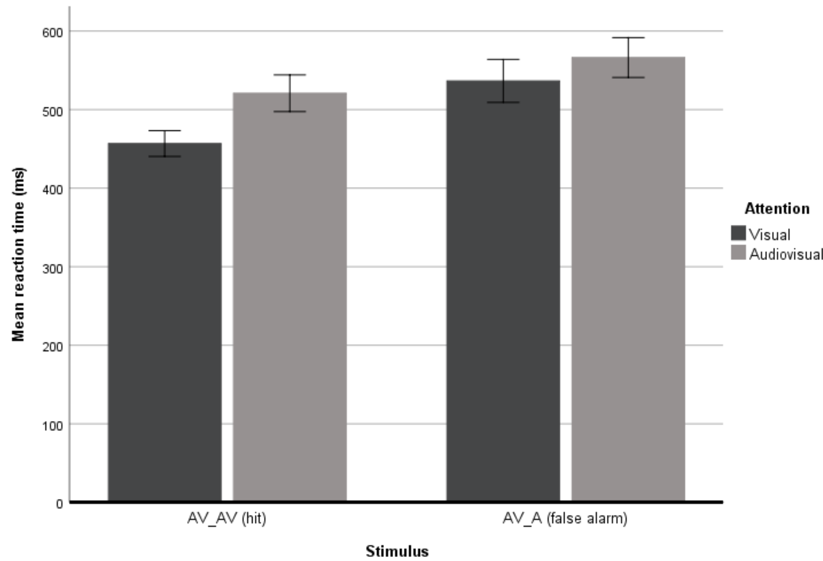

Figure 2.5. Reaction times for attention by stimulus type. This compares illusory false alarms (AV_A) and veridical hit for paired stimuli (AV_AV). Error bars +/- 1 SEM. Main effects were found for stimulus type and attention condition.

3.2.2 Sensitivity (*d′*) analysis

As in experiment 2, we examined whether visual sensitivity to a second visual flash varied within our unimodal "two flash" task and across attention conditions. The same formulae and categories from experiment 1 were again employed. Participants were significantly more sensitive to visual-only (V_V) stimuli ($d' = 2.58$) than multimodal (AV_AV) stimuli ($d' = 2.01$) in the unimodal 'two-flash' task attention condition [planned two-tailed, pairwise t-test, $t(17) = 3.53$, $p < .001$, Cohen's $d = 0.83$]. This suggests that the mere presence of auditory beeps made it more difficult to differentiate AV_A (flash illusion) from AV_AV, even when the near-threshold beeps were irrelevant to the task. Of particular interest, the average participant's sensitivity to the presence of double flashes and clicks (AV_AV) was much less ($d' = 0.37$) in the multimodal 'matching' attention condition than in the unimodal 'two flash' attention condition [planned two-tailed, pairwise t-test, $t(17) = 7.37$, p<.001, Cohen's *d* = 1.74]. This difference indicates a significant change in the ability to discriminate two veridical flashes, in the presence of near-threshold auditory stimuli, when the task requires bimodal attention to both visual and auditory stimuli. This result suggests that bimodal attention increases the sensory gain of the auditory signal and facilitates greater sensory integration. Further, the decrease in visual sensitivity in the bimodal task relative to the unimodal task supports the hypothesis that the illusion can be modulated by task attention.

In experiment 1, there was no recorded difference in total response bias for AV_AV stimuli across task attention conditions, as measured by criterion *c*. In the present experiment, however, we found that *c* changed from -0.85 in the visual only 'two flash' task to -1.31 in the audiovisual 'matching' task [pairwise two-tailed *t*-test, $t(17) = 3.32$, $p < .01$, Cohen's *d* = 0.78]. The fact that participants exhibited a shift in total bias, increasing the likelihood to respond in the 'matching' task, may reflect the increased difficulty in the task under near-threshold conditions. This will be explored further in the discussion.

3.2.3 Individual differences in the SFI

As an ancillary part of our project, we also have been interested to understand why some individuals seem to be more susceptible to experiencing the sound-induced flash illusion. As noted in previous work (e.g. Mishra *et al.*, 2007), there is notable between-subjects variability in reporting the illusion. In our visual-only task, participants ranged in reporting the illusion from not at all to 86.7% of AV_A trials. In the bimodal matching task, individual reports of the illusion varied from 53.3% to 98.3% of AV_A trials. To test the notion that inherent auditory sensitivity may be partially responsible for the illusory effects, we performed a *post hoc* correlation analysis on participants' auditory threshold levels and their later reports of the illusion. If the attention effect reported above is partly due to increased sensory gain in the auditory modality, individuals with naturally lower auditory thresholds in the pre-experiment auditory threshold task might have a greater likelihood of experiencing the illusion in the matching task. That is, more auditorily sensitive individuals may be more susceptible to experiencing the sound-induced flash illusion when attending to both auditory and visual targets. A Pearson correlation analysis found that individuals with lower auditory thresholds were more likely to experience the illusion in the matching task (*r* = - .468, n = 18, p < .05; see figure 6). The correlation is negative as auditory threshold scores tended to decrease as reported frequency of the illusion increased. This is interesting as the physical auditory stimulus was lower in intensity during the experiment for these participants, yet was still associated with a greater number of experienced illusions.

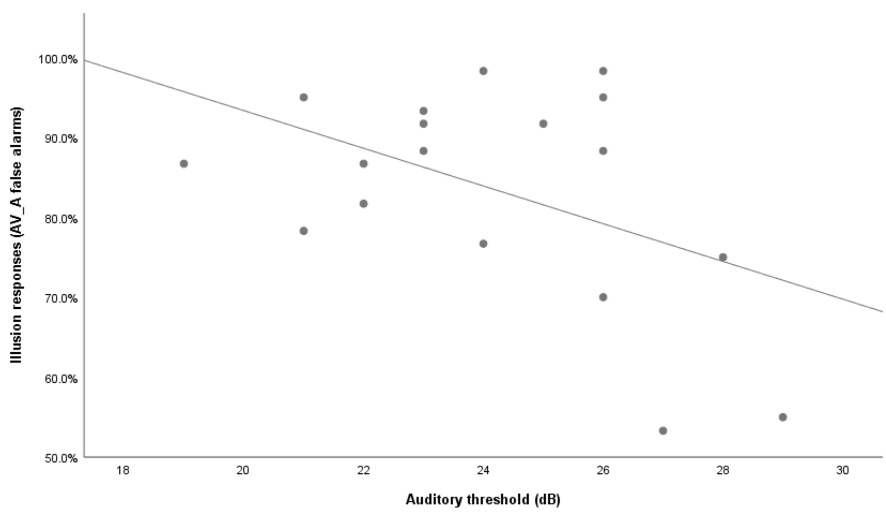

Figure 2.6. Correlation of auditory threshold and illusion. Lower auditory thresholds correlated moderately with experiencing more illusions in the bimodal attention condition, with approximately 22% of individual variance explained by their auditory threshold.

This correlation might be attributed to the multisensory principle of 'inverse effectiveness', in which the least effective stimuli are said to have the greatest proportional enhancement of multisensory interaction (Stein & Meredith, 1993). Under this interpretation, an individual's low perceptual threshold for auditory stimuli might result in greater enhancement of multisensory stimuli. However, no such correlation between auditory threshold and reports of the illusion was found in the unimodal 'two flash' identification task (r = .158, n =18, p = n.s.). This suggests that the increase in multisensory integration responsible for the illusion was not present in lower-threshold individuals when they were actively ignoring the auditory modality. Therefore, the 'inverse effectiveness' rule alone cannot explain why the lower intensity used by those participants with lower auditory thresholds correlated with increased incidence of the sound-induced flash illusion in the bimodal attention condition. It may be that some individuals with lower auditory thresholds have greater dynamic range in attentional gain, functionally decreasing their threshold during the auditory task, but increasing it when ignoring those stimuli. Of course, visual inspection of our scatter plot in figure 10 suggests that some of the correlation may be due to outliers. So, while this is a somewhat interesting exploratory analysis, the source of individual variation for the illusion remains unclear.

3.3 Discussion

When employing auditory and visual stimuli of similarly low perceptual salience, it appears that endogenous modal attention alters the multisensory integration processes responsible for the sound-induced flash illusion. In experiment 2, this was shown to hold true in both average reports of the SFI and in decreased ability to discriminate between AV_A and

AV_AV trials, as measured by *d′*. When observers attend to both auditory and visual channels, susceptibility to the illusion was found to increase, compared to a task in which they attended to only the visual channel. This finding suggests that, at low stimulus intensity levels, endogenous modal attention can serve to alternatively filter or enhance information that is temporally and spatially coincident with target stimuli. Attentional modifications in the unimodal channels are then carried through to the multisensory processes underlying the sound-induced flash illusion.

It is noteworthy that most of the difference in susceptibility to the illusion, measured both as frequency of reporting the illusory flash and *d′* sensitivity, can be attributed to its lower incidence in the visual-only ‘two flash’ task. Many participants reported that it was easy to ignore the quiet sounds during the visual task. While this is anecdotal, it is consistent with the notion that a selective attention mechanism could effectively gate extraneous, task-irrelevant stimuli. The fact that a number of illusions are still experienced in the unimodal task suggests a dynamic tension between the sensory ‘bottom-up’ and attentional ‘top-down’ processes involved in perceptual awareness. Attention can function to increase the gain of the bimodal stimuli during the ‘matching’ task, and also help gate that same auditory information during the visual ‘two flash’ task, each according to the instructional demands as consciously implemented by the participant.

4. General discussion

In the two experiments reported above, we tested the hypothesis that endogenous modal attention could alter the process of multisensory integration. Extending the sound-induced flash illusion paradigm (Shams *et al*., 2000; 2002), we introduced a novel ‘matching’ task that required participants to determine whether the number of stimuli presented in the auditory and visual modalities matched in number. This was compared with performance during a visual task in which participants reported instances of ‘two flashes’, ignoring the auditory stimuli. The mean frequency of SFIs reported and the ability to discriminate AV_A from AV_AV, as measured by d', served as indices of multisensory integration. In both experiments, it was found that participants were less able to discriminate the AV_A and AV_AV stimuli in the task that required bimodal attention. This effect was largest when near-threshold auditory stimuli were used. This is in keeping with a previous study which found that endogenous spatial attention could modulate electrophysiological components thought to be consistent with the sound-induced flash illusion (Mishra *et al*., 2009).

The use of lower intensity auditory stimuli in the second experiment appeared to allow endogenous task attention to have a greater effect on multisensory integration. Reducing auditory intensity primarily served to lower the rate at which participants reported the illusion in the visual ‘two flash’ task (M = 47.6%), compared to the SFI rate in the audiovisual “matching” task (M = 83.5%). To verify that this was a successful manipulation across experiments, we performed a *post hoc* comparison of the false alarms rates in AV_A stimuli in the ‘two flash’ unimodal tasks in experiment 1 (60 dB) versus experiment 2 (near-threshold) and found a significant difference [two-tailed t-test, $t(31) = 2.55$, $p < .05$, Cohen’s $d = 0.46$]. Although the experimental conditions across the two experiments were not identical, this result is consistent with the notion that the modulatory effect of endogenous modal attention was stronger in the presence of lower intensity auditory stimuli. Taken together, these results suggest that endogenous and exogenous attention dynamically interact during multimodal perception.

The current experiments also examined whether irrelevant sounds could produce a speeded facilitation effect for visual target identification in the SFI paradigm. A similar redundant signals effect had been found by Fiedler and colleagues (2011), in which participants were instructed to respond as quickly as possible to any visual stimulus presented. In our task, we employed a single button response task in both attention conditions to assay speeded responses in AV_AV versus V_V stimuli in our unimodal ‘two flash’ condition. A small facilitation effect was found in both experiments. While we believe this is likely due to generalized alerting caused by the arrival of the auditory stimulus, it is noteworthy that it also occurred in the presence of near-threshold stimuli employed in experiment 2. This supports the use of the SFI paradigm to simultaneously investigate both qualitative (illusory) and quantitative (reaction time) effects of multisensory integration.

While the above experiments support the hypothesis that endogenous attention can modulate multisensory integration, there are notable limitations in the design we employed. Our novel ‘matching’ task required participants to respond when the number of stimuli presented in each modality was the same. In the AV_A stimuli, a ‘match’ response was interpreted to indicate a false alarm that an AV_AV stimulus had been perceived. This is the defining perceptual experience of the sound-induced flash illusion.

Post-experiment debriefing interviews indicated that participants believed that the AV_AV trials were much more frequent than other stimuli. However, it is possible that some AV_A trials were misperceived as AV trials. This has been previously reported as a ‘fusion’ illusion (Andersen *et al.*, 2004), and tends to occur when a salient cue (e.g. loud beep), combines less salient stimuli (e.g. rapid flashes). Andersen reported the fusion of double flashes, when paired with a single loud beep (e.g. AV_V). However, they also reported that flashes did not seem to fuse two loud beeps. As this latter case is most similar to our first experiment, it suggests that AV_A would not be perceived as AV in our first experiment. More problematic for our second experiment was their finding that a single flash could fuse two near-threshold beeps. However, they found this to be a weak effect, only measurable when pooling data to include instances where one flash accompanied three beeps. Therefore, while the possibility of perceptual fusion cannot be excluded due to our single-button response method, we find it to be unlikely.

We have also pointed out that the order of our task-attention conditions remained fixed during both experiments, with the unimodal task always preceding the bimodal matching task. This was done as there were concerns that participants would have trouble ignoring the auditory stimuli once they had been asked to attend to both modalities. Although this was deemed a reasonable procedure while establishing our novel ‘matching’ task, it did introduce the possibility of order effects. Learning, fatigue, or demand characteristics in the task could have led to more false alarms in the AV_A trials during the second bimodal attention task. Fatigue was of special concern in the first experiment, which was longer in duration due to the addition of EEG setup and recordings.

Finally, we recognize that the greater complexity of the bimodal matching task may have introduced cognitive load as an experimental confound. As this task took more time to complete and was slightly more difficult, it could be argued that absolute cognitive load (i.e. increased task difficulty), rather than modal attention *per se*, was causally responsible for measured changes in multisensory integration. Against such a hypothesis, we note that Alsius and colleagues (2005) found that taxing visual attention during an audiovisual linguistic categorization task reduced the rate of *McGurk effect* misclassifications. In a separate study, Alsius *et al.*, (2007) also found a decrease in the McGurk effect when participants were asked to perform a tactile task while simultaneously making audiovisual linguistic perceptions. In that case, drawing endogenous attention away from the audiovisual stimuli reduced the generation of illusory speech perceptions characteristic of the McGurk effect. While we can't rule out an influence of task-related cognitive load in the present experiments, our hypothesis that endogenous attention enhanced multisensory integration reported here seems more consistent with findings using other qualitative illusory phenomena.

To further address the above issue and disentangle the effects of endogenous modal attention, exogenous stimulus intensity, and possible cognitive load differences between tasks, we designed two additional experiments. The two experiments separated and examined the effect of exogenous attention (e.g. stimulus intensity) on each endogenous modal attention condition in isolation. In chapter 3, we report on manipulation of auditory intensity within a visual-only, focused attention task. In chapter 4, we examine the effects of auditory intensity in our bimodal conditional matching task. In both experiments, we added response options to help disambiguate participant perceptions and allow for the inclusion of additional stimuli. Finally, we included controls for the possibility of order effects. These experiments will help further elucidate the role of attention in our multisensory integration paradigm.

Chapter 2 is co-authored with Townsend, Jeanne and Westerfield, Marissa. The dissertation author was the primary author of this chapter.

Chapter 3: Auditory intensity modulates multisensory integration in a focused attention visual task

1. Introduction

When a person is immersed in a busy sensory environment, sounds can be very effective at orienting attention towards important objects or regions of space. Drivers of new cars may have experienced this in the form of a beeping alarm that accompanies a visual icon in the side-view mirror when you attempt to change lanes with another car beside you. The visual icon becomes illuminated anytime a car is adjacent, helping you notice that something is in your car's 'blind spot.' However, as the icon occurs in your periphery when driving, you may remain unaware that it is activated. The beeping alarm occurs when you activate your turn signal while another car is present, cuing you to check your side mirror. The alarm is particularly effective as it can capture attention regardless of where you are looking. Although these signals provide redundant information in an attempt to make you aware of something unseen, their appearance together adds urgency that helps shift your attention to an immediate danger. Such audiovisual cues attempt to improve the speed and accuracy at which you detect an unseen vehicle and avoid an accident. If that fails, the much louder car horn of the driver next to you as you attempt to change lanes will almost certainly grab your attention.

The ability of sound to act as an effective exogenous cue for visual targets has received a great deal of study (Posner, 1980; Driver & Spence, 1998; McDonald, Teder-Sälejärvi, & Hillyard, 2000; Störmer, McDonald, & Hillyard, 2009). The above example is an especially interesting case in which your attention is drawn to an area you can't see (i.e. your car's 'blind spot'). The visual icon serves as the stand-in target and the audible alarm attempts to draw your attention to that visible icon. While each signal contains information, the combination of these two signals is specifically designed to make re-orientation even more effective (Ho, Santagelo, & Spence, 2009). Indeed, paired auditory and visual signals have been repeatedly found to speed overt sensory orientation to regions of space (Frens, Van Opstal, & Van der Willigen, 1995; Goldring, Dorris, Corneil, Ballantyne & Munoz, 1996; Spence & Santagelo, 2009). When both cues are informative, as in the redundant target driving task described above, enhanced performance makes intuitive sense. Each signal has value within the task, so the combination may add even more salience given the context. However, it has also been shown that *uninformative* auditory cues can speed or improve covert detection of visual targets (Lovelace, Stein, & Wallace, 2003; Spence & Driver, 1997; Vroomen & de Gelder, 2000).

In one remarkable study of uninformative auditory cues using a visual search task, participants were asked to find a vertical or horizontal line among diagonal visual distractors (Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). The color of the targets and distractors could alternate and the total set size varied. It was found that search times for the visual target greatly decreased when an auditory ‘pip’ was presented during a change of target color. Interestingly, performance improved even when the auditory ‘pip’ was uninformative, occurring only 20% of the time with targets and 80% of the time with non-target color changes. Reaction times to targets accompanied by ‘pips’ remained the same, even as the set size of distractors increased. Consequently, this result is known as the ‘pip and pop’ effect. Also critical to the above example is the fact that it was a focused attention task in which participants ignored the auditory ‘pips’ and only attended to the visual stimuli. Notably, the substitution of a visual cue for the auditory ‘pip’ did not produce the same performance gains

(Van der Burg, *et al.*, 2008). This was interpreted to mean that the reaction time reduction was not due to generalized alerting, but rather to the rapid crossmodal integration of the auditory cue and the visual target.

While the performance advantages of responding to multisensory signals over their unisensory components alone has been studied for some time (Stein & Meredith, 1993), the manner in which *attention* may interact with multisensory processes remains an area of intense interest (Koelewijn, Bronkhorst & Theeuwes, 2010 ; Talsma, Doty & Woldorff, 2006). The general goal of the current project is to investigate whether multisensory integration is an automatic process independent of attention or a dynamic process which can be modulated by endogenous and exogenous factors. Some experimental evidence has suggested that multisensory integration is largely automatic, occurring prior to spatial attention selection (Driver, 1996; Spence & Driver, 2000, Van der Burg et al., 2008), and independently of endogenously directed attention (Bertelson, Vroomen, De Gelder, & Driver, 2000; Vroomen, Bertelson & De Gelder, 2001). However, our first two experiments indicate that the processes responsible for multisensory integration may be modulated by endogenous modal attention, and that this modulation may be influenced by exogenous stimulus intensity.

In the experiments described in chapter 2, we examined the effect of endogenous, task-directed modal attention on the sound-induced flash illusion (Shams, Kamitani & Shimojo, 2000; 2002). In this audio-visual illusion paradigm, presenting two closely timed (< 100 ms inter-stimulus interval) auditory beeps with a single visual flash can induce participants to experience a second visual flash. The sound-induced flash illusion (SFI) is typically presented as a *focused attention* task, in which participants attend to the visual modality and ignore the auditory stream. In our first two experiments, we introduced a bimodal *conditional attention* task in which participants attended to both sensory channels in order to determine when the number of stimuli presented in each modality matched in number. When attending to both sensory channels, participants reported more flash illusions and were worse at discriminating illusory flashes from real flashes. This effect was more pronounced when a near-threshold auditory stimulus was used, producing a larger *reduction* of SFIs in the unimodal task condition. We hypothesized that individuals may have more success endogenously lowering the gain of the near-threshold stimuli, compared to louder irrelevant sounds, when performing a unimodal visual task. This may allow them to partially suppress the auditory stimulus and reduce multisensory integration. However, some amount of multisensory integration, as indicated by the tendency to experience the SFI and related sensitivity measures, still took place in the unimodal task. These findings suggest a dynamic interplay between endogenous and exogenous factors in speeded detection tasks.

To gain a better understanding of the role of stimulus factors in multisensory integration, we designed follow-up experiments to isolate the impact of auditory intensity on each endogenous modal attention condition. Rather than manipulate task-attention as in our previous experiments, the current experiment holds endogenous attention constant in a focused attention, unimodal visual task. We then present and compare two auditory intensity conditions to determine whether this exogenous stimulus factor influences multisensory integration. This will allow us to determine whether or not changing stimulus intensity in an unattended channel is sufficient to impact multisensory integration. The complement to this design, manipulating auditory intensity while holding endogenous attention constant in our bimodal task, is presented in chapter 4.

As in our previous experiments, one of our primary dependent measures used to quantify multisensory integration is the frequency of the sound-induced flash illusion (Shams *et al.*, 2000; 2002). While people vary greatly in their propensity to experience the illusion (Mishra, Martinez, Sejnowski & Hillyard, 2007), this illusion is robust in those who experience it and resistant to feedback (Rosenthal, Shimojo & Shams, 2009). As this audiovisual illusion paradigm relies on the multisensory binding of an initial pair of auditory and visual stimuli, it provides an interesting measure of multisensory integration. When a manipulation purports to increase multisensory integration, it should result in an increase in the incidence of sound-induced flash illusions. Factors that reduce integration should, likewise, reduce the frequency of the illusion.

We also expect changes in multisensory integration to be accompanied by shifts in Signal Detection Theory measures (primarily d') as an increase in the illusion should decrease the ability to discriminate illusory trials from their veridical counterparts. In addition to the above qualitative measures, we examine quantitative multisensory effects in the form of decreased reaction times and increased accuracy (Diederich & Colonius, 1987).

In the experiment presented below, we compare the effects of a near-threshold auditory stimulus with a more easily detected 60 dB (A) SPL stimulus. An individually titrated near-threshold stimulus was used successfully in experiment 2 and is similar to a study of the SFI by Andersen, Tiippana & Sams (2004). Andersen and colleagues reported the SFI (which they called a visual *fission*) in a focused attention paradigm using an auditory stimulus of 10 dB intensity. While they reported that the rate of the SFI was lower than that seen in a condition using 80 dB stimuli, the difference did not reach statistical significance.

Along with sound-induced flash illusions, we will also be interested in the occurrence of sound-induced visual *fusions.* Andersen *et al.* (2004) reported that visual flashes may 'fuse' together when presented with a lower number of auditory stimuli. This was reported with high intensity (80 dB) auditory stimuli, but disappeared with their 10 dB stimuli.

If multisensory integration can be affected by stimulus intensity, we would expect more sound-induced illusory flashes to be experienced in our higher intensity sound condition. We would also predict that more flash fusions would occur in the louder condition. However, if auditory and visual cues are automatically integrated, then stimulus intensity should have little effect on reports of the illusory flashes and fusions across auditory intensity conditions.

2. Methods

2.1 Participants

Twenty healthy adults (13 women, 7 men; mean age 21.1 years) participated in this experiment. All provided written informed consent, in accordance with the University of California, San Diego Human Research Protections Program. None of the participants took part in either of our earlier experiments. Two participants (1 male and 1 female) were excluded from analyses due to their reported prior experience with the illusion. All remaining participants had normal or corrected-to-normal vision, reported normal hearing, and were naïve to the purpose of the study. Individuals received course credit or monetary compensation for participation and were debriefed after the experiment.

2.2. Apparatus and stimuli

The same equipment and basic stimuli used in the current study were also employed in the experiments described in chapter 2. To summarize briefly, stimuli were presented in a darkened, sound-attenuated room. The visual stimuli were uniform white circle ‘flashes’ subtending 2 degrees of visual angle. A single flash (V) was presented 10 degrees below a fixation cross for 10 ms (one screen refresh at 100 Hz). Double flashes (V_V) were offset by a stimulus onset asynchrony (SOA) of approximately 70 ms. The auditory stimulus (A) was a 1,000 Hz ‘beep’ presented for 10 ms from two speakers flanking the 21-inch video monitor. When two auditory stimuli were presented (A_A), the SOA was approximately 70 ms. Auditory stimuli were presented at two levels: near-threshold and 60 dB SPL, as measured from the source.

The near-threshold auditory stimulus intensity was established for each participant using the two-step procedure described in chapter 2. The final value for our experimental auditory stimuli was set at the lowest sound level at which each individual participant reported hearing the sound at least 85% of the time. The average resulting sound intensity for participants was 23 dB (A) SPL, measured at the speaker. A consistent auditory intensity level was successfully identified for all participants.

In addition to the four unimodal auditory and visual stimuli outlined above, we included four multimodal stimuli. These were a single flash and beep pair (AV), double flash and beep pairs (AV_AV), a single flash and two beeps used to induce the SFI (AV_A), and a single beep with two flashes (AV_V). Paired stimuli (e.g. AV) shared the same onset timing. Stimuli following a pair (e.g. AV_A) had an SOA of 70 ms. The AV_V stimulus was used to balance the AV_A mismatch stimuli and prevent participants from adopting a default response strategy for ambiguous trials. It also provided an opportunity to explore sound-induced flash ‘fusions’ originally reported by Andersen and colleagues (2004).

2.3. Procedure

The experiment was divided into the two auditory intensity conditions. Each of the eight trial types outlined above was repeated 60 times. Trials were mixed in a pseudo-random style and presented over two blocks per auditory intensity condition. Trials were separated by randomly jittered intervals of 1,400 – 1,900 ms to help prevent anticipation of stimuli and began automatically. A short break was provided between blocks. For all participants, the near-threshold condition preceded the condition using 60 dB SPL stimuli. This order was kept constant as it was believed that participants might habituate to the louder auditory stimuli if presented in the first blocks and become less responsive to the near-threshold stimuli. A control condition for order is reviewed in the discussion section.

Participants were asked to indicate as quickly and accurately as possible whether they saw one or two flashes on each trial. Responses were recorded via alternative button press on a computer mouse. This was a focused attention paradigm and participants were instructed to ignore the auditory stimuli and withhold responses if no visual stimuli were present. If participants experienced a fusion illusion during AV_V trials, they should report only one flash. By contrast, if they experienced a sound-induced flash illusion in AV_A trials, they should respond that two flashes were present. Only trials with no visual component (A, A_A) would require no response. Single and double flashes were equally likely on any given target trial.

3. Results

A number of analyses were planned for the current experiment. We were broadly interested in the possible multisensory effects of auditory stimuli on accuracy and sensitivity to visual targets in our visual discrimination task. These potential multisensory interactions were compared with unimodal presentations of visual stimuli. As with our earlier experiments, our dependent measures were frequency of responses, reaction times (RTs), sensitivity (as expressed by *d´*), and bias (quantified by *c*). Below, the multisensory integration effects of irrelevant auditory stimuli on visual targets are presented separately for each intensity condition. We then report on between-condition comparisons to determine whether multisensory integration effects changed with auditory intensity, using our sound-induced flash and fusion stimuli (AV_A and AV_V, respectively) as critical trials. Comparisons employed repeated measures ANOVAs and planned pairwise *t*-tests, unless otherwise noted.

3.1 Near-threshold blocks: frequency and reaction time

Our first planned comparisons examined the possible effect of irrelevant sounds on the speed and accuracy of visual detections during our choice response task. We had two unimodal visual trials (V and V_V) which could be compared with bimodal stimuli (AV and AV_AV, respectively). In a check of the effect on response accuracy, a two (visual target: one flash, two flashes) by two (modality: unimodal, bimodal) repeated measures ANOVA of frequency of correct responses found a main effect of modality [$F(1, 17) = 4.58$, $p < .05$, $\eta^2_{partial} = .212$], but no main effect of number of flashes and no interaction (see figure 3.1). Individual planned comparisons revealed that accuracy improved for two flashes ($M = 79.8\%$)

when two irrelevant beeps were also presented [M = 86.5%; planned two-tailed, pairwise *t*-test, $t(17) = 2.41$, $p < .05$, Cohen's $d = 0.58$].

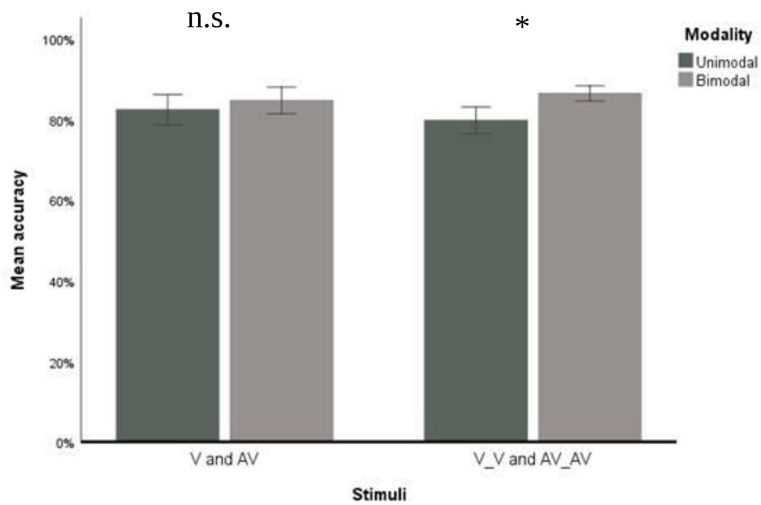

Figure 3.1. Accuracy by modality. There was a main effect of modality on the detection of visual flashes. This was largely due to the increased accuracy for detecting two flashes in the presence of two irrelevant beeps, compared to flashes alone [$t(17) = 2.41$, $p < .05$]. Error bars +/- 1 SEM.

Examining reaction times for unimodal and bimodal stimuli, we found a significant reduction in AV trials compared to V trials of ~17 ms [planned two-tailed, pairwise *t*-test, $t(17) = 2.23$, $p < .05$, Cohen's $d = 0.53$]. Reaction times and response frequencies for correct trials are provided in table 1. While AV_AV trials were also faster than V_V trials, this difference was not statistically significant [planned two-tailed, pairwise *t*-test, $t(17) = 0.53$, p = n.s.]. Reviewing these results, it appears that there is a trade-off of RT speed and accuracy in the significant effects above. We also note that while the RTs for our double flashes appear faster than single flash trials in Table 4, this is an artifact of our calculation method. RTs were derived by measuring responses from the onset of the last presented stimulus. For a proper comparison with single stimulus or single pair trials, you would add the 70 ms SOA time used for each double flash presentation.

Table 3.1. Mean behavior for correct visual target responses with near-threshold and 60 dB auditory stimuli.

Stimulus	Mean percentage of near-threshold trials reported correctly (SEM)	Mean RT in ms. (SEM)	Mean percentage of 60 dB trials reported correctly (SEM)	Mean RT in ms. (SEM)
V	82.4 (3.8)	678 (19)	79.6 (3.6)	654 (20)
V_V	79.8 (3.3)	614 (19)	72.2 (3.6)	595 (18)
AV	84.7 (3.3)	661 (19)	90.0 (2.4)	652 (28)
AV_V	63.1 (5.5)	625 (23)	40.2 (6.1)	703 (30)
AV_A	57.6 (4.4)	667 (22)	40.6 (6.3)	738 (36)
AV_AV	86.5 (1.9)	609 (22)	89.5 (2.2)	602 (23)

3.2 Near-threshold blocks: sensitivity and bias

To better understand the effect of task-irrelevant sounds on our flash discrimination task, we calculated Signal Detection Theory (Macmillan & Creelman, 2005) measures for perceptual sensitivity and total response bias in our critical comparisons. In order to have valid comparisons for both our illusion and fusion stimuli, it was necessary to calculate two unimodal visual baselines. When the comparison was to the AV_A sound-induced flash illusion trials (falsely perceived as AV_AV trials), we first ascertained the ability to correctly detect two unimodal flashes. Correct V_V trials were defined as unimodal hits, while V trials reported as two flashes were defined as unimodal false alarms. These values were used to calculate d' as our sensitivity measure and c as our indicator of bias.

When the comparison was to AV_V fusion trials (falsely perceived as AV trials), we determined the ability to correctly detect a single unimodal flash. Correct V trials were defined as unimodal hits, while V_V trials reported as a single flash was categorized as unimodal false alarms. The same Signal Detection Theory formulae used in our first two experiments for these measures were again used here. Values for reporting when two flashes were detected are shown in table 5.

Table 3.2. Mean behavior for reporting two flashes for near-threshold and 60 dB auditory stimuli.

Stimulus	Mean percentage of near-threshold trials reported as two flashes (SEM)	Mean RT in ms. (SEM)	Mean percentage of 60 dB trials reported as two flashes (SEM)	Mean RT in ms. (SEM)
V	15.9 (3.7)	677 (39)	18.7 (3.5)	714 (28)
V_V	79.8 (3.3)	614 (19)	72.2 (3.6)	595 (18)
AV	13.9 (3.3)	719 (47)	10.0 (2.4)	740 (47)
AV_V	63.1 (5.5)	625 (23)	40.2 (6.1)	703 (30)
AV_A	40.7 (4.5)	639 (25)	56.5 (6.0)	646 (23)
AV_AV	86.5 (1.9)	608(22)	89.5 (2.2)	602 (23)

Does the presence of irrelevant, near-threshold auditory stimuli impact visual discrimination? In our near-threshold condition, the *unimodal* visual sensitivity for two flashes (mean d' = 2.16) was significantly greater than our *multimodal* visual sensitivity to two flashes (d' = 1.44), using correct responses to AV_AV as hits and two-flash responses for AV_A trials as false alarms [planned two-tailed, pairwise t-test, $t(17) = 4.39$, $p < .001$, Cohen's d = 1.03]. Total bias was also significantly different [planned two-tailed, pairwise t-test, $t(17) = 5.72$, $p < .001$, Cohen's d = 1.35], with 'two flash' responses more liberal under multimodal conditions (mean c = -0.46) than unimodal conditions (mean c = 0.11).

In a similar analysis of flash-fusion responses, the *unimodal* visual sensitivity to a single flash (mean d' = 2.12) was compared to *multimodal* visual sensitivity, with correct AV responses recoded as hits and AV_V trials indicating a single flash present recorded as false alarms. Visual sensitivity under flash-fusion stimulus conditions yielded lower sensitivity (mean d' = 1.64) than the unimodal trials [planned two-tailed, pairwise *t*-test, $t(17) = 3.22$, $p < .01$, Cohen's $d = 0.76$]. There was also a shift in bias from the unimodal visual condition (mean $c = -0.03$), with multimodal stimuli (mean $c = -0.35$) reflecting a more liberal response profile for 'single flash' responses [planned two-tailed, pairwise *t*-test, $t(17) = 3.19$, $p < .01$, Cohen's $d = 0.75$]. Together, these results suggest that irrelevant, near-threshold auditory stimuli reduced participants' ability to accurately discriminate between one and two visual flashes. This effect was larger in magnitude for the sound-induced flash illusion (Cohen's d = 1.03) than the flash-fusion trials (Cohen's d = 0.76). The results also highlight a change in response bias, with a greater inclination to match visual discrimination responses to the number of auditory stimuli present.

3.3 60 dB blocks: frequency and reaction times

As above, we investigated the impact of the task-irrelevant 60 dB auditory signals on the correct, speeded discrimination of visual stimuli. Using a two (visual target: one flash, two flashes) by two (stimulus: unimodal, bimodal) repeated measures ANOVA on frequency of correct responses, we found a main effect of stimulus modality [$F(1, 17) = 45.49$, $p < .001$, $\eta^2_{partial} = .728$], and an interaction [$F(1, 17) = 4.58$, $p < .05$, $\eta^2_{partial} = .212$]. There was no main effect for number of flashes (see figure 3.2). Direct comparisons revealed that accuracy was improved significantly by the presence of irrelevant beeps. Unlike the near-threshold stimulus condition, a 60 dB beep increased accuracy (M = 90.0%) over a unimodal flash presented alone [M = 79.6%; planned two-tailed, pairwise *t*-test, $t(17) = 3.96$, $p < .001$, Cohen's $d = 0.93$]. Two flashes paired with two beeps were also more frequently identified

(*M* = 89.5%), than two flashes presented alone [*M* = 72.2%; planned two-tailed, pairwise *t*-test, *t*(17) = 6.71, p < .001, Cohen's *d* = 1.58].

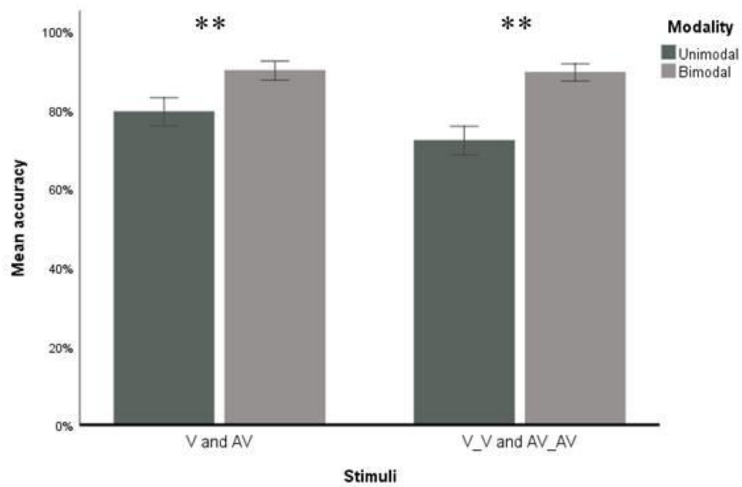

Figure 3.2. Mean accuracy for stimuli by attention modality. There was a main effect of modality on the detection of visual flashes, with irrelevant beeps increasing visual response accuracy [F(1, 17) = 45.49, p < .001, η_partial^2= .728]. This effect was larger for double flashes (see text). Error bars +/- 1 SEM.

In an analysis of reaction times, there was no statistically significant difference between RT for V trials (*M* = 654 ms) and AV trials (*M* = 652 ms) [planned two-tailed, pairwise *t*-test, *t*(17) = 0.11, p = n.s.]. There was also no notable difference in RT for unimodal (*M* = 595 ms) and bimodal (*M* = 602 ms) presentations of double flashes [planned two-tailed, pairwise *t*-test, *t*(17) = 0.60, p = n.s.]. As was noted in the near-threshold condition, there appears to be a tradeoff between speed and accuracy.

3.4 60 dB blocks: sensitivity and bias

As in our analysis of the near-threshold auditory condition, we began by finding baseline values for sensitivity and bias in our unimodal visual conditions (V_V and V).

Measured sensitivity for distinguishing two from one visual flash revealed a mean *d′* of 1.71. This was significantly higher than visual sensitivity when stimuli were accompanied by two irrelevant auditory beeps (mean *d′* = 1.23) in illusory AV_A trials [planned two-tailed, pairwise *t*-test, $t(17) = 3.77$, $p < .01$, Cohen's $d = 0.89$]. As in the near-threshold condition, we also found a shift in response criterion [planned two-tailed, pairwise *t*-test, $t(17) = 9.97$, $p < .001$, Cohen's $d = 2.35$], with unimodal visual responses (mean $c = 0.20$) becoming more lenient when beeps were present (mean $c = -0.82$). Though both effects were significant, it is noteworthy that the shift in bias (Cohen's $d = 2.35$) was much larger in magnitude than the change in sensitivity (Cohen's $d = 0.89$).

A comparison of unimodal visual sensitivity was also performed on fusion stimuli, using AV_V as a false alarm for hits in the AV condition. Analysis of baseline visual discrimination for *unimodal* one and two flashes found a mean *d′* of 1.69. Visual sensitivity was significantly lower in the *bimodal* condition (mean *d′* = 1.23), indicating that a single 60 dB beep made it more difficult to distinguish one flash from two flashes [planned two-tailed, pairwise *t*-test, $t(17) = 3.52$, $p < .01$, Cohen's $d = 0.83$]. Total bias was also affected by the presence of a single beep [planned two-tailed, pairwise *t*-test, $t(17) = 6.53$, $p < .001$, Cohen's $d = 1.54$],, with a much greater tendency to respond that one flash had been seen in the AV trials (mean $c = -0.85$) when compared with the unimodal V trials (mean $c = -0.13$).

3.5 Comparison of auditory intensity conditions: frequency and reaction times

Here we examined the effects of auditory stimulus intensity on reports of the sound-induced flash illusion (AV_A) and veridical two flash and two beep (AV_AV) trials. A two (stimulus: AV_A, AV_AV) by two (intensity: near-threshold, 60 dB) repeated measures

ANOVA was used to compare the proportion of trials in each condition responded to as two flashes. This analysis revealed a main effect of stimulus type [$F(1, 17) = 67.56$, $p < .01$, $\eta^2_{partial} = .799$]. A main effect was also found for auditory intensity between near-threshold and 60 dB conditions [$F(1, 17) = 8.20$, $p < .05$, $\eta^2_{partial} = .325$]. A significant interaction between these two factors was additionally reported [$F(1,17) = 8.02$, $p < .05$, $\eta^2_{partial} = .321$]. The interaction and main effect of intensity appear largely driven by the reports of the SFI, with a lower incidence in the near-threshold condition (M = 40.7%, SD = 19.3%) than in the 60 dB condition (M = 56.5%, SD = 25.3%). This was a statistically significant difference [planned two-tailed, pairwise t-test, $t(17) = 3.07$, $p < .01$, Cohen's $d = 0.72$; see figure 3.3]. By contrast, the intensity manipulation did not significantly affect veridical judgements of the AV_AV stimuli.

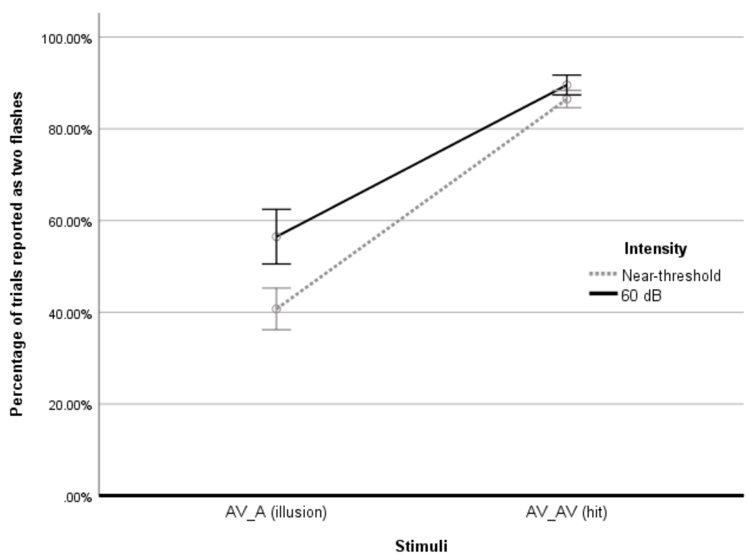

Figure 3.3. Reports of two flashes in AV_A (illusion) and AV_AV (hit) by auditory intensity condition. There is a main effect for stimulus and intensity. The interaction was also significant (see text). The intensity and interaction effects appear driven by the difference in AV_A illusions seen in near-threshold and 60 dB conditions [$t(17) = 3.07$, $p < .01$]. Error bars +/- 1 SEM.

We next examined the effect of auditory intensity on reports of a sound-induced flash fusion (SFF). We conducted a two (stimulus: AV_V, AV) by two (intensity: near-threshold, 60 dB) repeated measures ANOVA as the critical comparison. The ANOVA revealed main effects of both stimulus [F(1, 17) = 51.96, p < .001, $\eta^2_{partial}$= .753] and auditory intensity [F(1, 17) = 34.26, p < .001, $\eta^2_{partial}$= .668]. The ANOVA additionally indicated an interaction between stimulus and intensity [F(1, 17) = 18.25, p < .001, $\eta^2_{partial}$= .531]. The interaction is driven (see figure 3.4) by the difference in reports of the SFF between near-threshold (M = 34.8%, SD = 23.0%) and 60 dB (M = 58.0%, SD = 26.5%) intensity conditions [planned two-tailed, pairwise t-test, t(17) = 5.87, p < .001, Cohen's d = 1.38]. A relatively smaller difference was also found between AV trials in the lower and higher auditory intensity conditions [two-tailed, pairwise t-test, t(17) = 2.49, p < .05, Cohen's d = 0.59].

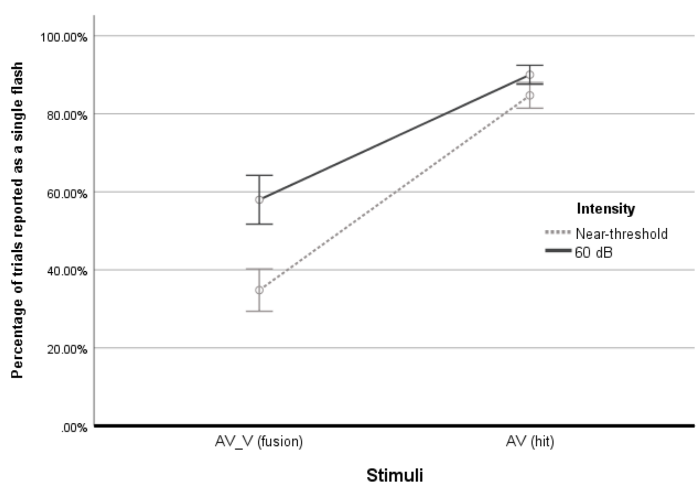

Figure 3.4. Reports of one flash in AV_V (fusion) and AV (hit) trials by auditory intensity condition. There is a main effect for stimulus and intensity. The interaction was also significant (see text). The largest effect was the difference in AV_V fusions seen in near-threshold and 60 dB conditions [t(17) = 5.87, p < .001]. Error bars +/- 1 SEM.

In the above ANOVA, we chose to compare AV_V fusion stimuli with AV trials as these are perceptually similar. In the case of a fusion in AV_V trials with a response indicating a single flash was seen, the stimuli are mistaken for AV trials. This is consistent with the logic used when comparing sound-induced flash illusion trials (AV_A) with veridical AV_AV trials. The line graphs are useful for comparing the perceptual continuity between the two discrete stimulus classes. These comparisons are also employed in our sensitivity and bias analyses below. While other comparisons are also valid, our approach appears to be the most stringent comparison for purposes of stimulus discriminability. This is discussed further in chapter 6.

In section 3.1 and 3.3, we have reported quantitative multisensory effects (e.g. reaction time and accuracy) within each intensity condition. We became curious to know whether changes in quantitative multisensory effects could be induced by louder auditory stimuli, as seen for the SFI and SFF. In a between-condition comparison of reaction times and accuracy for trials where the number of number of auditory and visual stimuli were the same (e.g. AV or AV_AV), we found few statistically significant differences. Although reaction times decreased and accuracy increased in the louder 60 dB condition for AV and AV_AV trials, the only statistically significant difference was for accuracy in reporting near-threshold AV trials (M = 84.7%, SD = 14%) vs. louder 60 dB AV trials [M = 90.0%, SD =10.3%; *post hoc* two-tailed, pairwise t-test, $t(17) = 2.49$, $p < .05$, Cohen's $d = 0.59$]. It is difficult to know if this quantitative multisensory effect is a meaningful result as the number of unplanned comparisons inflates the family-wise false alarm rate and renders this not significant when adjusted. The quantitative multisensory effects persist across auditory intensity conditions, but they appear to remain fairly constant. That is, they do not change in the same fashion as the qualitative effects reflected in the illusory perceptions.

3.6 Comparison of auditory intensity conditions: sensitivity and bias

To further characterize the effect of task-irrelevant sounds on our flash discrimination task, we compared d' and *c* values across sound intensity conditions. To our surprise, for our baseline two flash (V_V) valid targets, we found a difference between the unimodal visual d' in our near-threshold (mean d' = 2.16) and 60 dB (mean d' = 1.71) conditions [two-tailed, pairwise *t*-test, $t(17) = 3.47$, $p < .01$, Cohen's $d = 0.82$]. This is notable as these stimuli did not have an auditory component. No difference in response criterion *c* was found between auditory intensity conditions [two-tailed, pairwise *t*-test, $t(17) = 0.93$, p = n.s.]. We then compared the unimodal visual d' for single flash (V) valid targets in the near-threshold (mean d' = 2.12) and 60 dB (mean d' = 1.69) conditions. These were also found to be significantly different [two-tailed, pairwise *t*-test, $t(17) = 3.62$, $p < .01$, Cohen's $d = 0.85$]. No difference in response criterion *c* was uncovered for unimodal flash discrimination between auditory conditions [two-tailed, pairwise *t*-test, $t(17) = 0.91$, p = n.s.]. These are interesting findings as they indicate that visual discrimination becomes worse in the higher auditory intensity condition, even in trials where the auditory stimuli are not present.

Unimodal d' and *c* measures were compared with multimodal illusion conditions. The first comparison was a two (stimulus: visual only, SFI) by two (intensity: near-threshold, 60dB) repeated measures ANOVA of d' values. This examines whether the presence of auditory beeps makes it more difficult to determine the number of visual flashes present. We found main effects for both stimulus condition [$F(1,17) = 51.99$, $p < .001$, $\eta^2_{partial} = .754$] and intensity [$F(1,17) = 15.3$, p < .001, $\eta^2_{partial} = .474$]. No significant interaction was present. Reviewing the individual pairings across intensity conditions (see figure 3.5), we found that d' decreased in the louder 60 dB sound condition. However, this sensitivity decrease was significant for visual-only stimuli, and not for the sound-induced flash stimuli [near-threshold *M* = 1.44, 60 dB *M* = 1.23; planned two-tailed, pairwise *t*-test, $t(17) = 1.30$, p = n.s.].

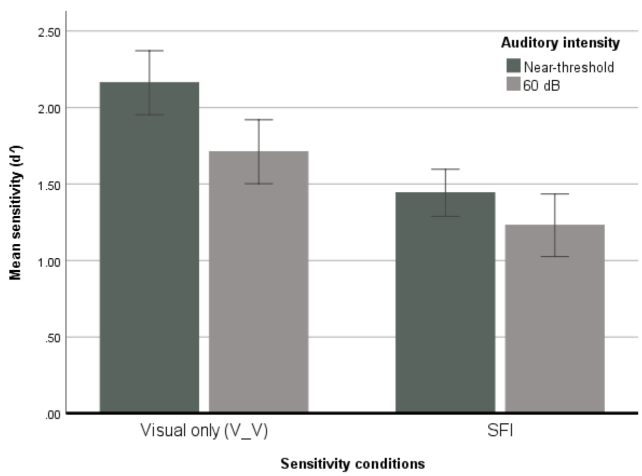

Figure 3.5. Comparison of d' values across sound intensity conditions. Participants were worse at distinguishing single from double flashes in the 60 dB condition, even when sounds were absent. However, differences in SFI sensitivity across intensity conditions were not significant (see text). Error bars +/- 1 SEM.

We performed an analogous repeated measure ANOVA on bias values. This revealed a main effect of stimulus [$F(1,17) = 97.99$, p < .001, $\eta^2_{partial} = .852$] and an interaction between stimulus and auditory intensity[$F(1,17) = 13.83$, p < .01, $\eta^2_{partial} = .449$]. There was no main effect of auditory intensity. Inspecting figure 3.6, the interaction can be seen in the larger difference in bias measure found for the SFI across conditions. The ‘two flash’ responses in near-threshold SFI trials (*c* = -0.46) were significantly more conservative than 60 dB trials [c = -0.82; planned two-tailed, pairwise t-test, $t(17) = 3.13$, $p < .01$, Cohen's d = 0.74]. Taken together, these measures indicate that participants were similarly able to distinguish AV_AV trials from AV_A trials across auditory intensity conditions, but had an increased tendency to respond that AV_A trials contained two flashes in the louder 60 dB condition. The reverse pattern was seen in visual only stimuli, with V_V and V trials more difficult to distinguish in the louder 60 dB condition, but no significant change in response tendency.

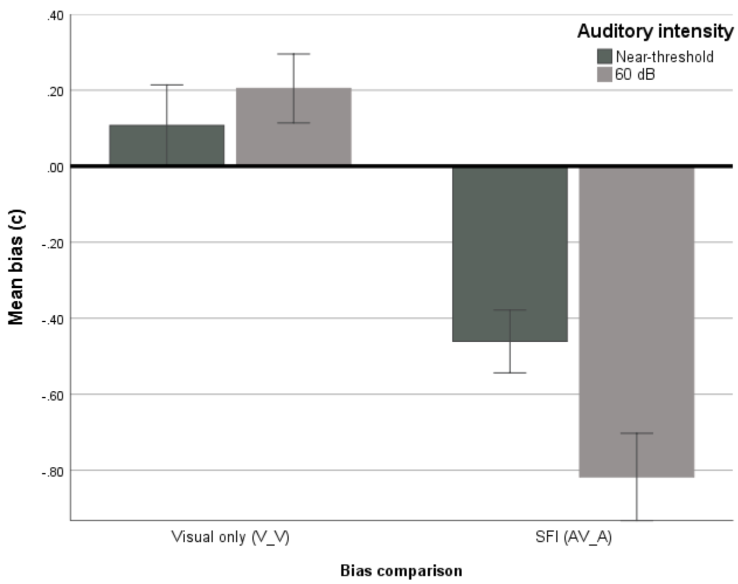

Figure 3.6 Comparison of bias (c) across sound intensity conditions. Bias was similar for unimodal visual stimuli across intensity condition. However, participants were more liberal in responses to the SFI in the 60 dB condition. Error bars +/- 1 SEM.

In a similar examination of flash-fusion sensitivity across intensity conditions, a two (stimulus: visual only, fusion) by two (intensity: near-threshold, 60 dB) repeated measure ANOVA found main effects for stimulus [$F(1,17) = 27.5$, $p < .001$, $\eta^2_{partial} = .618$] and intensity [F(1,17) = 32.7, p < .001, $\eta^2_{partial}$= .658], but no interaction (see figure 3.7). A direct comparison of fusion stimulus d' values found a significant difference between near-threshold (mean d' = 1.64) and 60 dB (mean d' = 1.23) conditions [planned two-tailed, pairwise *t*-test, $t(17)$ = 2.95, p < .01, Cohen's d = 0.70]. This indicates that fusion stimuli (AV_V) were harder to differentiate from AV stimuli in the louder trials.

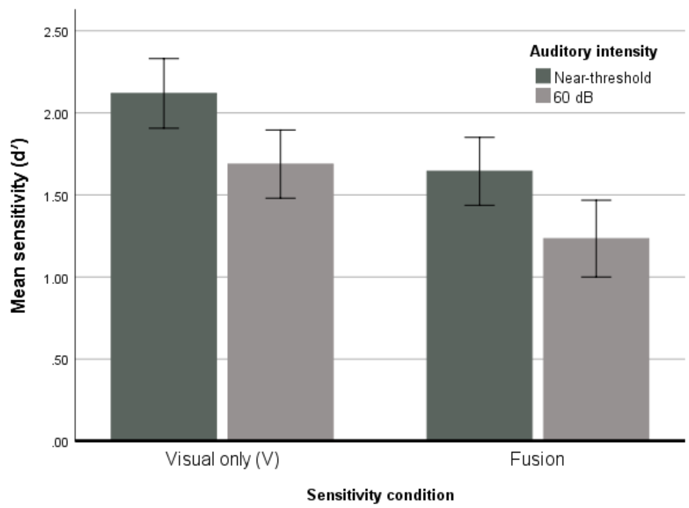

Figure 3.7. Comparison of d′ values across sound intensity conditions for visual only and fusion conditions. Participants were worse at distinguishing double from single flashes in the 60 dB condition. This was true when auditory sounds were present (flash fusion trials) or absent (visual only). Error bars +/- 1 SEM.

A comparison of mean bias in fusion trials across intensity conditions using repeated measure ANOVA on bias values revealed a main effect of stimulus [F(1,17) = 29.95, p < .001, $\eta^2_{partial}$= .638], a main effect of auditory intensity [F(1,17) = 13.71, p < .01, $\eta^2_{partial}$= .446], and an interaction [F(1,17) = 21.17, p < .01, $\eta^2_{partial}$= .555]. Inspecting figure 3.8, the interaction is most evident in the SFF difference [near-threshold c = -0.35, 60 dB c = -0.85; planned two-tailed, pairwise *t*-test, $t(17)$ = 6.27, p < .001, Cohen's d = 1.48]. These analyses suggest that participants adopted a more lenient criterion to respond that one flash was present in AV_V trials, with the tendency becoming greater in the 60 dB auditory condition.

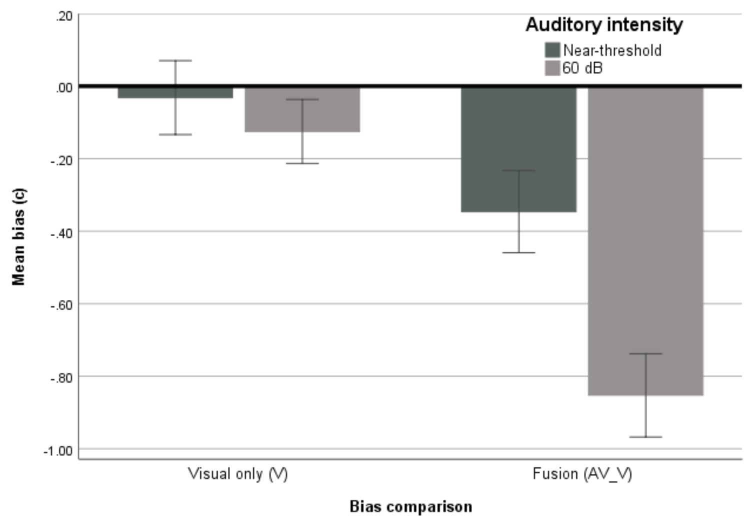

Figure 3.8. Bias comparison across intensity conditions for visual only and SFF trials. A more liberal bias was evident in the Fusion trials, and this significantly increased in the 60 dB intensity condition. Error bars +/- 1 SEM.

3.7 Individual differences in experiencing the sound-induced flash illusion

In our earlier two experiments and in previous work with this paradigm (e.g. Mishra *et al.*, 2007), there has been tremendous individual variability in the rate at which individuals report the sound-induced flash illusion. In the current experiment, participants ranged from reporting the SFI in 3% of near-threshold AV_A trials to 95% in 60 dB AV_A trials. In chapter 2, we explored the notion that auditory sensitivity may be related to susceptibility in experiencing the SFI. In our second experiment, we noted that individuals with lower auditory thresholds were more likely to experience the SFI in a bimodal matching task using near-threshold stimuli ($r = -.468$, $n = 18$, $p < .05$). This correlation was not present in the unimodal visual task, and was only found when participants had to attend to both sensory channels. We speculated that individuals who exhibit a lower auditory threshold may have greater endogenous control over auditory gain, and that this might make them more susceptible to the SFI when the auditory channel was task-relevant. It may also allow them to more easily ignore near-threshold auditory stimuli in a unimodal visual task where the auditory elements are irrelevant.

In the present study, we examined the relationships between individual auditory thresholds and reports of the sound-induced flash illusion. Consistent with the findings of experiment 2, a Pearson correlation analysis of individual auditory thresholds and reports of the SFI in the near-threshold auditory condition was not statistically significant in our unimodal visual task (r = .310, n = 18, p = n.s.). However, in the 60 dB SPL condition, there was a significant correlation between auditory thresholds and reports of the SFI (r = .488, n = 18, p < .05; see figure 3.9). It is noteworthy that the relationship in the unimodal task is in the opposite direction as that seen in the bimodal task in experiment 2.

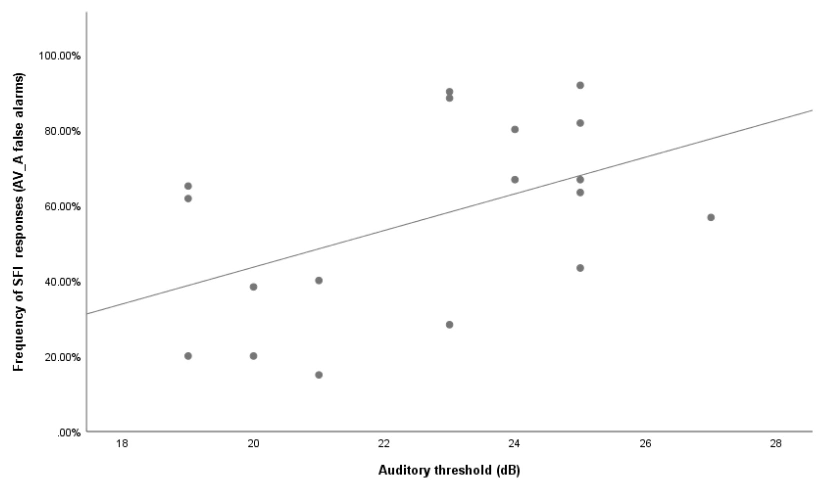

Figure 3.9. Correlation of SFI with auditory threshold. Lower auditory thresholds correlated moderately with experiencing fewer SFIs in a unimodal visual discrimination task, but only in the 60 dB intensity condition. Approximately 24% of individual variance was explained by a participant's auditory threshold.

4. Discussion

The current experiment tested the hypothesis that exogenous attention, manipulated through changes in the intensity of a task-irrelevant signal, could modulate multisensory integration. Participants were asked to perform a unimodal visual discrimination task, indicating the presence of one or two visual flashes. Irrelevant auditory beeps could be presented during the visual task, but were to be ignored. The beeps were delivered in two blocked intensity conditions, near-threshold for each individual or 60 dB SPL. When performing the unimodal visual discrimination task, we found that irrelevant auditory stimuli could increase either the accuracy or decrease reaction time to concurrently presented visual stimuli within each auditory intensity condition. This is consistent with the quantitative multisensory integration effects reported in chapter 2.

These quantitative multisensory effects were accompanied by qualitative multisensory effects (i.e. illusions) in each auditory condition. When the number of auditory stimuli did not match the number of flashes present, as in AV_A and AV_V trials, participants frequently reported sound-induced flash illusions (SFI) and sound-induced flash fusions (SFF), respectively. Importantly, the rate at which illusory flashes and fusions were reported was found to *increase* in the higher auditory intensity condition. This change was accompanied by a *decrease* in perceptual sensitivity (d') in the case of flash fusions and a change in response bias (c) for both the sound-induced flash illusion and flash fusions. Surprisingly, a decrease in unimodal visual sensitivity was also found in conditions in which louder 60 dB auditory stimuli were possible, but not actually present (i.e. V and V_V trials). This suggests that exogenous auditory intensity can modulate both multisensory integration and unimodal visual discrimination.

The current work significantly expands upon an important study of the effect of auditory intensity on the sound-induced flash illusion by Andersen and colleagues (2004). While they reported a reduction in experienced SFIs in a 10 dB sound condition when compared to a louder 80 dB condition, their result was not statistically significant. The current study finds a significant reduction in reported sound-induced flash illusions in our near-threshold condition when compared with a louder 60 dB condition. It is difficult to make direct comparisons to Andersen and colleagues' analysis as they pooled results over a number of stimuli that may induce a flash illusion (e.g. one flash paired with two beeps or three beeps). However, it is likely that our use of individually titrated, near-threshold stimuli provided a less variable floor for multimodal effects in the lower intensity condition. We also note that our auditory stimuli were presented via speakers flanking the visual stimuli, while auditory stimuli in the other study were delivered via headphones. This difference in stimulus delivery may impact the spatial coherence of the crossmodal binding necessary for the illusion. Finally, their lower number of analyzed participants (n = 5) may have resulted in insufficient power to reveal the effect.

Additionally, while Andersen *et al.* (2004) found flash fusions in their louder condition, no fusions were reported in their lower intensity 10 dB condition. The present study found significant evidence of the fusion illusion in both of our intensity conditions. The reasons for this difference likely include the factors noted above, but may also be due to the inability of many of their participants to detect the 10 dB auditory stimulus. Though Andersen *et al.* (2004) report dropping participants from analysis for this very reason, it may be that the auditory stimulus intensity was too low for the remaining participants to detect in a consistent fashion.

A final point of comparison with Andersen and colleagues' (2004) study comes in a common finding of reduced performance in visual target identification for *unimodal* stimuli in the higher intensity auditory condition. They noted this change in performance, but could only attribute it to random fluctuations in group performance. The present study also found reduced accuracy in reporting unimodal stimuli (V and V_V) in the louder 60 dB condition. We additionally reported a change in visual discriminability, as measured by *d′*, between auditory intensity conditions. It is puzzling why a louder condition should negatively impact visual discrimination in trials where no auditory stimuli are present. We suggest that the expectation of the louder stimulus causes participants to allocate attention resources to inhibit the auditory channel. This requires more effort in the louder condition and would be consistent across all trials as they are presented randomly. Such inhibitory effort would also help explain why illusions are less frequent in lower auditory intensity conditions as those stimuli would be easier to ignore.

One curious result in our experiment is a potential disconnect between our sensitivity measure (*d′)* and the changes in report frequency of the SFI and flash fusions across auditory intensity conditions. While we found *increases* in the reported frequency of the SFI and flash fusions in the louder intensity conditions, this was only accompanied by a *decrease* in perceptual sensitivity in the case of the flash fusion effect. There are a few possible explanations for the failure to find a significant decrease in *d′* in our SFI stimuli. First, and most parsimoniously, it is possible that the change in reported frequency of SFIs in the near-threshold condition is largely due to the noted shift in response bias. Changes in bias measure *c* were found in both cases, with a greater tendency to make responses consistent with the number of beeps present in the louder 60 dB condition. While this interpretation is not consistent with the results of our first two experiments, this must remain a possibility. Second, the manipulation may not have been powerful enough for our specific participant sample to yield a significant difference in *d′*. Reviewing experiment 1, we found that SFIs were reported at a higher rate (*M* = 71.3%) than in the 60 dB condition in the current experiment (*M* = 56.5%). This raises the possibility a larger manipulation using a greater range of auditory intensities may be necessary to consistently reveal a difference in perceptual sensitivity. This latter possibility seems more persuasive and is worth pursuing.

An additional issue regarding the lack of a significant difference in our *d′* values between auditory intensity conditions revolves around questions of its appropriateness to measure changes in illusory phenomena in the first place. Witt, Taylor, Sugovic & Wixted (2015) have argued that as the SFI is a visual illusion, it is inherently a product of bias in the perceptual system. They have suggested that this makes criterion *c* the more appropriate Signal Detection Theory measure of perceptual change in experiencing the illusion. If true, this would render the absence of a significant difference of *d′* values to be unimportant. While this is an intriguing idea, we still note that *c* is a measure of *total* bias in a class of judgments. It is affected by factors such as motivation, reward structure, perceived risk, and other ‘top-down’ influences (Macmillan & Creelman, 2005). Even if you allow that the SFI and flash fusions may reflect a kind of perceptual bias, criterion *c* must additionally be influenced by these other elements. Furthermore, as *d′* is a measure that quantifies the ability to tell two stimuli apart, it seems very well suited to determining the power of an illusion. Due to its greater specificity for the quality we wish to measure in our multisensory illusions, we find that *d′* remains an important measure for the current project. We will return to this issue in chapter 6.

In the current experiment, it is important to note that the near-threshold auditory condition always preceded the louder, 60 dB SPL condition. This order was initially deemed necessary over a concern that participants might habituate to a louder stimulus if presented first, and become less responsive to near-threshold stimuli later. While interesting in its own right, this might mask the effects of exogenous attention we sought to examine. However, we had previously identified this design as a potential problem as it introduces a possible order effect in which fatigue, practice, or other time-based confounds might explain our results. To check for this possibility, we ran a small control group ($n = 7$; 2 females, 5 males) on the same experimental set-up, but with auditory intensity conditions presented in the reverse order. The louder blocks were always presented before the near-threshold blocks. The only difference in our control condition is that, due to an unanticipated change in auditory settings, our louder stimuli were presented at 65 dB SPL.

Comparing the critical SFI and fusion trials across intensity conditions, we saw trends similar to those uncovered in the main experiment. The SFI was reported more often in the louder auditory condition ($M = 55.5\%$) than in the near-threshold condition [$M = 20.7\%$; *post hoc* two-tailed, pairwise *t*-test, $t(6) = 3.08$, $p < .05$, Cohen's $d = 1.16$]. Similarly, fusions in AV_V stimuli were reported more frequently in louder conditions ($M = 52.1\%$) than with near-threshold auditory stimuli [$M = 29.3\%$; *post hoc* two-tailed, pairwise *t*-test, $t(6) = 3.89$, $p < .01$, Cohen's $d = 1.47$]. While the pattern of responses is consistent with our findings in the main experiment, it is noteworthy that the SFI was reported to a lower degree in the near-threshold conditions. This is consistent with our expectation that participants would adapt to the louder stimuli if presented first, and be better able to ignore near-threshold stimuli when presented in the second condition.

When the louder auditory condition was presented first, our Signal Detection Theory bias measure (c) showed similar patterns of results to that seen in the main experiment, with both SFI and fusion stimuli showing a shift towards more liberal responses in the louder 65 dB condition. However, notable differences were found in a direct comparison of fusion and SFI d' values across auditory intensity conditions. Sensitivity to fusion stimuli did not differ significantly across conditions when the louder stimulus conditions was presented first [*post hoc* two-tailed, pairwise *t*-test, $t(6) = 1.89$, p = n.s.]. Notably, near-threshold (mean $d' = 2.34$) and 65 dB (mean $d' = 1.43$) comparisons of sensitivity to the SFI were significantly different, unlike the main experiment [*post hoc* two-tailed, pairwise *t*-test, $t(6) = 2.98$, $p < .05$, Cohen's $d = 1.13$]. It is difficult to know whether the difference in d' values was due to the louder intensity stimulus used in the control experiment, natural variability in the smaller group, or some other factor. The accidental use of a different intensity in the louder condition makes direct comparisons to the main experiment problematic. However, it does provide evidence that order effects alone were not responsible for the results in our main experiment.

As part of our project, we also asked whether an individual's auditory sensitivity may influence the degree to which they experience the sound-induced flash illusion. In a previous experiment, we found a negative correlation between auditory thresholds and reported frequency of the illusion. This was only found in a bimodal attention condition using near-threshold stimuli. In the current unimodal attention task, we found a positive correlation between auditory threshold and likelihood of reporting the illusion in a higher intensity (60 dB) condition. To reconcile the different directions of the correlations, we speculate that a lower auditory threshold reflects some ability to endogenously alter the gain of auditory stimuli. When attending to the stimuli, as in the bimodal attention condition, a lower threshold would result in more sensitivity to the incoming stimulus and result in a greater degree of multisensory integration. When attempting to ignore the auditory stimulus, most participants were equally successful in ignoring the near-threshold stimulus. However, at higher intensities (e.g. 60 dB), those with lower auditory threshold were able to endogenously lower the gain of the auditory stimuli. This would result in a tendency to experience fewer illusions in those with lower auditory thresholds. It would also result in more illusions in those individuals with higher thresholds and less ability to endogenously gate auditory information. Of course, given that this is based on two correlation analyses of modest effect size, this remains a highly speculative possibility. However, this seems worthy of pursuit, possibly with the inclusion of more auditory levels for comparison and more sensitive auditory pretesting.[5]

Finally, it is worth noting that the use of the phrase ‘exogenous attention’ in this project to describe the effect of stimulus intensity on multisensory integration may be conceptually problematic. Exogenous attention is often discussed in terms of a spatial or object-oriented exogenous cuing in which one stimulus alerts the participant, causing a near automatic orienting response to a following stimulus (Posner, 1980). The timing involved in such priming is typically greater than 200 ms and is, therefore, longer than the multisensory integration window for our SFI and flash fusion stimuli (< 100 ms). Because of the different timing and evidence that multisensory integration occurs prior to, and independent of, exogenous spatial cuing (Driver, 1996; McDonald, Teder-Salejarvi & Ward, 2001), there may be some resistance to applying ‘exogenous attention’ to the effects we have reported. Though we recognize the differences, we argue that the stimulus intensity effects revealed in the

[5] A natural question is to ask whether auditory thresholds were correlated to experiencing the flash fusion. We did find a number of significant correlations to sensitivity to both the flash fusion and unimodal discriminability. Additionally, visual sensitivity to unimodal stimuli was highly, negatively correlated to sensitivity to all multisensory illusions. While this raises interesting possibilities for future research, it is a bit afield from our original theoretical questions.

above experiment and related work (Andersen *et al*., 2004) suggest a longer-duration process of channel monitoring across conditions. As the task conditions change by intensity, participants may also change the manner in which they process incoming information. Lower intensity auditory stimuli are easier to ignore and require less sustained effort over the course of a block of trials. In a sense, it is easier to endogenously reduce the gain of these signals. Louder intensity stimuli, being more intrinsically salient, are more difficult to ignore and require more sustained effort. This may distract from the primary visual task and redirect common cognitive resources, giving rise to more perceptual errors. Such a shift could reasonably be called an effect of 'exogenous attention.'

We have found that irrelevant 60 dB auditory stimuli reduce visual discrimination in unimodal visual stimuli, when compared to conditions featuring near-threshold auditory stimuli. The incidence of sound-induced flash illusions and flash fusions were found to increase in the louder condition. These changes in illusory perceptions are sometimes accompanied by changes in perceptual sensitivity, and always associated with changes in total response bias. We also found suggestive evidence that these effects may be sensitive to order of presentation. Taken together, we see this as supporting the hypothesis that the processes underlying multisensory integration can be broadly influenced by exogenous factors. This appears true even if stimulus intensity is modulated in an unattended sensory channel. It remains to be seen how these exogenous stimulus factors may influence multisensory integration when both modalities are relevant to the task. We turn to that question in the next experiment.

Chapter 4: The effect of auditory intensity on bimodal attention

1. Introduction

In the normal course of daily activity, we engage with our multisensory environments in a variety of ways. Sometimes we focus solely on a particular item within a sensory channel, as when reading an article and trying to shut out the conversation at a nearby table. Other times, we need to attend to multiple channels simultaneously in order to achieve our goal, such as matching speech and lip movements when following a conversation in a noisy room. Because each of these activities is so commonplace, we sometimes assume that they take place automatically and require little effort. However, research into the role of selective attention during crossmodal tasks has suggested that complex interactions take place during these simple events (Koelewijn, Bronkhorst, & Theeuwes, 2010; Talsma, Doty, & Woldorff, 2006).

In previous chapters, we have investigated the degree to which endogenous and exogenous attention can influence reports of multisensory integration effects in unimodal and bimodal target detection tasks. The broad goal is to describe the extent to which multisensory processes can be modulated by task-directed attention or stimulus-generated processing. In chapter 2, we presented experiments that compared multisensory integration under unimodal *focused attention* conditions where the targets were visual-only versus a bimodal *conditional attention* task in which the targets had both auditory and visual components. We found that attending to both auditory and visual channels in the bimodal task increased multisensory effects and that this difference was larger when using near-threshold auditory stimuli. In chapter 3, we explored the role of stimulus intensity in a unimodal visual target detection task and found that multisensory effects *increased* in the presence of louder auditory stimuli when compared to conditions involving near-threshold auditory stimuli. Taken together, these experiments suggest that endogenous task-attention can alter multisensory processing, but that these changes are also sensitive to the exogenous factors of the task stimuli.

The current experiment provides a complement to our earlier studies. We now ask whether auditory stimulus intensity can affect multisensory processing when participants are required to attend to both auditory and visual channels. While chapter 3 described the effects of auditory intensity on crossmodal integration in a *focused attention visual task*, the present study manipulates auditory intensity in the *bimodal conditional attention* matching task introduced in chapter 2. Auditory intensity was previously found to impact multisensory integration when participants were instructed to ignore all auditory stimuli in the former task. We now explore the effect of auditory intensity when participants are required to actively integrate auditory information during the bimodal task. This important extension will enable us to determine whether requiring full attention to both modalities can impact the effects of auditory intensity found earlier. This will also assist in a discussion of our results in terms of competing functional hypotheses that have been offered for our illusory effects (Andersen *et al.,* 2004; 2005).

In our bimodal conditional attention task, participants are instructed to attend to both auditory and visual modalities to determine whether a matching number of stimuli are presented in each trial. As in chapter 3, this will be done under two possible intensity conditions, a louder 60 dB (A) SPL intensity and a softer intensity adjusted to the individually determined threshold of each participant. Unlike the experiments with this bimodal task in chapter 2, however, we include a new response option that allows participants to indicate the number of matching pairs of stimuli encountered. This allows for evaluation of additional multisensory effects and will enable us to better characterize the ways in which stimulus intensity may interact with modal attention.

We have used the *sound-induced flash illusion* (Shams, Kamatani & Shimojo, 2000; 2002) and sound-induced *flash fusion* (Andersen, Tiippana & Sams, 2004) as primary indices of multisensory integration in previous experiments. In the sound-induced flash illusion (SFI), when two auditory beeps are presented with a single visual flash (AV_A), they can be experienced as two pairs of beeps and flashes (AV_AV). In the sound-induced flash fusion illusion (SFF), the AV_V combination is perceived as a single matching pair (AV), though typically when the auditory signal is loud in intensity (Andersen *et al*., 2004). The same group also found that two flashes and a single beep (AV_V) could be experienced as two flashes and two beeps (AV_AV). This latter effect, the flash-induced sound illusion (FSI), was only experienced at low auditory stimulus intensities. Finally, AV_A stimuli could be perceived as AV trials, a phenomenon described as an *auditory fusion*. While the unimodal visual task used in chapter 3 allowed us to investigate SFIs and flash fusions, the bimodal task in the current experiment additionally allows for measurement of the FSI and auditory fusions.

An additional change in the current experiment is the inclusion of order of auditory intensity condition as a between-subjects variable. In previous experiments, we were concerned that participants might habituate to louder stimuli if presented before near-threshold auditory conditions. Indeed, this seemed to emerge as a possibility in a small control experiment reported in chapter 3. It additionally seemed possible that the order of the auditory conditions affected sensitivity and bias in our conditions of interest. With the inclusion of order of presentation in randomized and balanced groups, these issues will be examined in more detail in the current experiment.

In our previous experiments, we have hypothesized that sensory gain control is one of the factors influencing changes in illusory perceptions. For example, in the visual focused attention task, we speculated that participants were gating some auditory signal when instructed to ignore the irrelevant sounds. This had the effect of partially suppressing multisensory integration and reducing the SFI and flash fusions. When stimulus intensity increased, as in experiment 3, this gating was less successful and the SFIs and flash fusions increased. In experiments 1 and 2, when endogenously shifting from focused attention to the bimodal task, participants would functionally increase the gain of the auditory signal by attending to the auditory channel. This had the effect of increasing SFI effects.

If multisensory integration in our bimodal task can be affected by stimulus intensity, as demonstrated in the unimodal task in chapter 3, we would expect *increases* in reports of the SFI and flash fusions in the louder intensity condition, relative to the near-threshold condition. These effects may be smaller than seen in experiment 3, since the bimodal task should promote a higher baseline of multisensory activity.

The flash-induced sound illusion (i.e. AV_V perceived as AV_AV) was not tested in our earlier experiments. For this to occur, the auditory stimulus intensity must be low (Andersen *et al.,* 2004). At higher auditory intensities, a flash fusion may occur for these same trials. Similarly, we predict that reports of auditory fusions in the AV_A condition (perceived as AV) should be largest at lower auditory intensities. At higher intensities, the sounds would become more likely to induce an illusory flash.

Predictions on the basis of order of auditory intensity condition may vary from the small changes seen in the unimodal task control condition described in the Discussion section in chapter 3. In that task, auditory stimuli were to be ignored. It was found that participants who heard the loud auditory condition first reported fewer SFIs with the softer stimuli presented second. We interpret this as habituating to the higher intensity sound and becoming better able to suppress the lower intensity sound that followed. This was something of a surprise given that previous work had found the SFI to be resistant to experience in the form of explicit feedback (Rosenthal, Shimojo & Shams, 2009). As participants must attend to both stimulus channels in the current experiment, the effect of order of presentation is unclear. However, if multisensory integration is largely automatic in nature, the default is to expect no change of behavior due to order.

2. Methods

2.1 Participants

Thirty-one adults (18 women, 13 men; mean age 21.3 years) participated after giving written informed consent. None of the participants took part in any of our earlier experiments and all were naïve to its purpose. One male participant was excluded from analysis due to extremely poor performance detecting veridical matching stimuli. The remaining 30 participants had normal or corrected-to-normal vision and reported normal hearing. Participants received course credit or monetary compensation for participation, and were debriefed following the experiment. All procedures were approved by the University of California, San Diego Human Research Protections Program.

2.2. Apparatus and stimuli

This experiment used the same equipment and basic stimuli as the experiments described in chapter 2 and 3. As reported in chapter 3, auditory stimuli were presented at either 60 dB(A) SPL, as measured from the source, or at near-threshold levels customized for each individual. Near-threshold auditory stimuli were determined for each participant using the two-step process described in chapter 2. For each individual, we established their near-threshold stimulus as the minimal intensity at which they responded with at least 85% accuracy. The overall average sound intensity was 21 dB (A) SPL, measured at the speaker.

Stimuli included a single flash (V), a double flash (V_V), a single beep (A), and a double beep (A_A). In addition to the four unimodal visual and auditory stimuli, we included four other critical trials. A single flash was paired with a single beep (AV), double flash and beep pairs (AV_AV), a single flash with two beeps (AV_A), and a single beep with two flashes (AV_V). The AV_A trials indicate the *sound-induced flash illusion* when responded to as AV_AV or an *auditory fusion* if responded to as AV. Likewise, the AV_V trials signal a *flash-induced sound illusion* when responded to as AV_AV or a *flash fusion* when responded to as AV.

2.3. Procedure

As in chapter 3, the experiment was divided into two auditory intensity conditions. Each of the eight trial types described above was presented 60 times in each condition. Trials were produced in a pseudo-random order and presented over two blocks of equal size. Trials were presented automatically following randomly jittered intervals of 1,400 – 1,900 ms. To control for possible order effects, participants were randomly assigned to one of two possible orders of intensity conditions, with either the near-threshold (‘soft’) or 60 dB (‘loud’) conditions coming first.

The experiment used the novel ‘matching’ task introduced in chapter 2, with one important response modification. As before, participants were instructed to maintain fixation on the center cross. When stimuli were presented, participants were to determine whether the number of stimuli in each sensory modality matched in number. Unlike the earlier experiment, however, participants were instructed to respond with their index finger to indicate a single matching pair of stimuli (AV) and their middle finger to signal a double matching pair of stimuli (AV_AV). Speed and accuracy were equally emphasized, as in our earlier experiments. This allowed participants to specify how many pairs of stimuli were encountered on each trial, allowing tracking of both induced illusory stimuli (e.g. AV_A experienced as AV_AV) and fused stimuli (e.g. AV_A experienced as AV). If stimuli were only presented in a single modality (e.g. V or A), they were to withhold responses. Likewise, if the stimuli mismatched in number, (e.g. AV_A or AV_V), they were instructed to withhold responses. All stimuli were equally likely on any given trial. Participants were given practice trials prior to beginning the experiment until they expressed comfort and displayed competence with the task.

3. Results

As in chapter 3, we were primarily interested in measuring the effects of exogenous stimulus intensity on multisensory integration. Because of a concern over order effects, we also include between-group comparisons of the order in which loud (60 dB) and soft (near-threshold) conditions were presented. Our dependent measures include frequency of responses, sensitivity (d'), and bias (c). Below, we organize our analyses first by comparing potentially induced-illusion trials (AV_A and AV_V) with the trials with which they are mistaken (AV_AV). We then turn to evaluations of 'fusion' trials in which the same mismatching stimuli are mistaken for AV trials. Unless otherwise noted, comparisons rely on mixed measures ANOVAs with order as a between-subjects factor and both trial type and intensity as within-subjects factors. Planned pairwise *t*-tests were also used to clarify interactions.

3.1. Sound-induced flash illusion trials: AV_A reported as AV_AV

Our first planned analysis examined the effect of sound intensity on reports of the sound-induced flash illusion (SFI) in trials containing two beeps and one flash (AV_A). Accordingly, we compared the proportion of AV_AV trials judged as double-match with the proportion of AV_A trials judged as double-match (SFI) across both intensity conditions. Analysis involved a two (order: soft/loud, loud/soft) by two (intensity: soft, loud) by two (trial type: AV_A, AV_AV) mixed ANOVA, with order as the between-subjects factor and trial type and intensity as within-subjects factors. This analysis revealed significant main effects for trial type [$F(1, 28) = 67.4$, $p < .01$, $\eta^2_{partial} = .707$] and intensity [$F(1, 28) = 41.0$, $p < .01$, $\eta^2_{partial} = .594$]. There was no main effect for order [$F(1, 28) = 1.04$, $p = .32$]. Most relevant to our interest in the effect of sound level on multisensory integration, we found a significant interaction of auditory intensity and trial type [$F(1, 28) = 9.74$, $p < .01$, $\eta^2_{partial} = .258$]. Reporting two matching pairs of stimuli for both AV_AV and AV_A occurred less often with softer sounds, with a larger reduction seen in SFI responses. As can be seen in figure 4.1, participants reported significantly more SFIs in the loud condition (M = 70.4%, SD = 27.3%), than in the soft condition [M = 50.5%, SD = 20.8%; planned two-tailed, pairwise *t*-test, $t(29) = 5.61$, $p < .001$, Cohen's $d = 1.02$]. Additional response frequencies are shown in table 6.

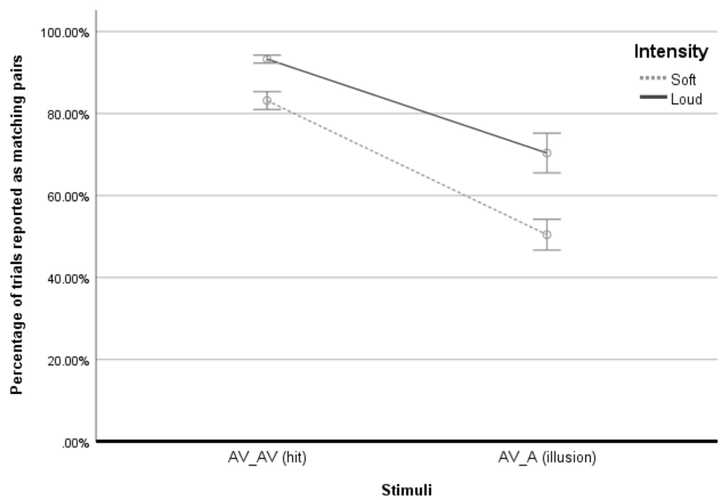

Figure 4.1. Proportion of trials in each intensity condition in which participants responded 'double match.' There were main effects of trial type and intensity on reports of double matching pairs of stimuli. An interaction between stimulus and intensity resulted from fewer reports of the SFI in the softer auditory condition [$F(1, 28) = 9.74$, $p < .01$, $\boldsymbol{\eta^2_{partial}} = .258$]. Error bars +/- 1 SEM.

The overall analysis also revealed an interaction between trial type and order of intensity conditions [$F(1, 28) = 5.23$, $p < .05$, $\eta^2_{partial} = .157$]. The three-way interaction of order, trial type, and intensity was not significant [$F(1, 28) = 0.53$, $p = .82$]. Inspecting the interaction in figure 4.2, participants who had experienced the louder (60 dB) intensity condition first were more likely to report the sound-induced flash illusion overall. This suggests that experience with different auditory intensities over the course of the experiment impacted participants' reports of the illusion (see table 6 for response frequency by order of intensity condition). Whether this is the result of perceptual priming or settling on a response preference is unclear. It is now important to see if differences caused by auditory stimulus intensity are associated with changes in perceptual discrimination or overall bias.

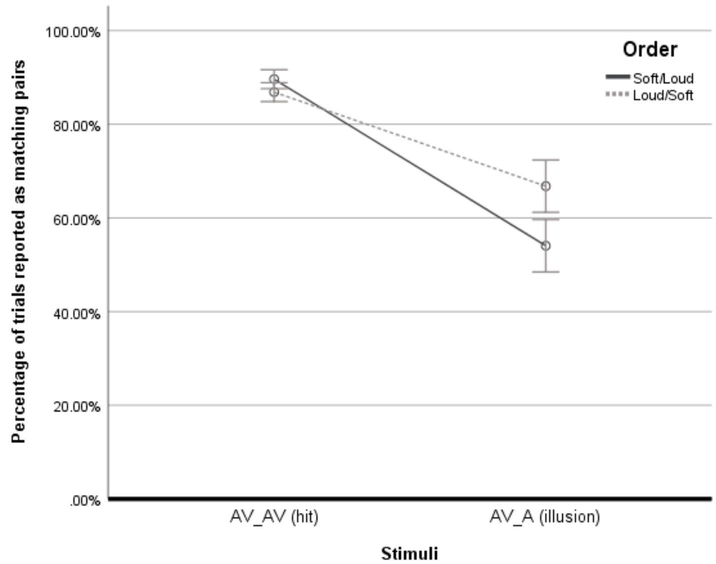

Figure 4.2. Order and stimuli interact. There was an interaction between stimulus and order, resulting in fewer overall reports of the SFI when the softer auditory condition was presented first. Error bars +/- 1 SEM.

To evaluate changes in the ability to distinguish illusory AV_A trials from veridical AV_AV trials, we used Signal Detection Theory (Macmillan & Creelman, 2005) to calculate measures for perceptual sensitivity (d') and bias (c) using the formulas described in chapter 2. Testing for the ability to discriminate between AV_AV and AV_A, correct responses to AV_AV trials were considered 'hits' and 'double-match' responses to AV_A trials were considered 'false alarms.' These values were used to calculate d' scores in both soft (near-threshold) and loud (60 dB) auditory intensity conditions. Scores were then analyzed via mixed ANOVA with between-subjects factor order (soft first, loud first) and within-subjects factor intensity (soft, loud). Analysis revealed only a main effect for order [$F(1, 28) = 4.60$, $p < .05$, $\eta^2_{partial} = .141$], with participants in the soft-first order having a higher mean d' (1.22) that those in the loud-first order (0.77). Neither intensity [$F(1, 28) = 0.98$, $p = .331$] nor the interaction of order and intensity [$F(1, 28) = .023$, p = .88] were significant. This suggests that order of auditory intensity conditions influenced the ability to perceptually distinguish AV_AV from AV_A, consistent with figure 4.2. Those who experienced the soft intensity condition first were better able to discriminate between veridical AV_AV trials and AV_A SFIs. Manipulation of auditory intensity alone seemed insufficient to change the ability to differentiate these stimuli.

Bias (c) was analyzed in the same manner as d'. This mixed ANOVA revealed a main effect for auditory intensity [$F(1, 28) = 59.6$, p < .001, $\eta^2_{partial}$= .68]. There were no significant effects for order or the interaction of order and intensity. Response bias was greater in the loud condition (*M* = -1.13, *SD* = 0.56) than in the soft condition (*M* = -0.53, *SD* = 0.47). This suggests that the auditory intensity effects on the sound-induced flash illusion seen in figure 4.1 are largely driven by an increase in the willingness to respond that two flashes are co-present when two loud beeps are played.

Table 4.1. Mean behavior in near-threshold (soft) and 60 dB (loud) conditions reporting 'match', by auditory intensity condition presentation order.

Stimulus	Near-threshold trials (SEM), presented first – order soft/loud	Near-threshold trials (SEM), presented second – order loud/soft	60 dB trials (SEM), presented second – order soft/loud	60 dB trials (SEM), presented first – order loud/soft
V	9.6 (3.7)	4.7 (1.1)	0.9 (0.3)	1.3 (0.7)
AV	86.5 (1.6)	82.0 (2.2)	97.2 (0.6)	96.3 (0.6)
AV_V (fusion)	48.7 (5.4)	66.7 (6.1)	81.6 (5.1)	91.7 (1.8)
AV_V (FSI)	26.4 (4.0)	13.0 (3.8)	2.6 (0.8)	1.5 (0.8)
AV_A (fusion)	34.6 (5.4)	28.8 (4.2)	7.2 (2.6)	6.1 (0.8)
AV_A (SFI)	45.7 (5.2)	55.2 (5.5)	62.4 (8.2)	78.3 (5.2)
AV_AV	85.8 (2.0)	80.6 (3.9)	93.4 (1.4)	93.1 (1.3)

3.2. Flash-induced sound illusion trials: AV_V reported as AV_AV

The second item of interest was the potential effect of sound intensity and order on the flash-induced sound illusion (FSI). As above, we examined effects of intensity and order of intensity condition on our trials of interest. In a two (order: soft/loud, loud/soft) by two (intensity: soft, loud) by two (trial type: AV_V, AV_AV) mixed ANOVA, with order as the between-subjects variable, we found significant main effects of trial type [$F(1, 28) = 1670.7$, $p < .001$, $\eta^2_{partial} = .984$], intensity [$F(1, 28) = 5.24$, $p < .05$, $\eta^2_{partial} = .158$], and order [$F(1, 28) = 4.97$, $p < .05$, $\eta^2_{partial} = .151$]. As with our other induced illusion comparison, we also found a significant interaction of auditory intensity and trial type [$F(1, 28) = 89.6$, $p < .001$, $\eta^2_{partial} = .762$; see figure 4.3].

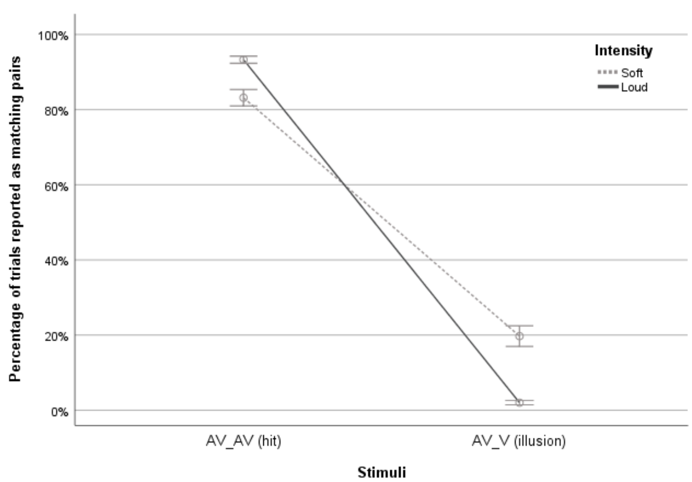

Figure 4.3. Proportion of trials in which participants responded 'double match.' There was an interaction between stimulus and intensity, resulting in fewer reports of the flash-induced sound illusion in the louder auditory condition. See text for details. Error bars +/- 1 SEM.

Reporting two matching pairs of stimuli in AV_V trials occurred less often with louder sounds, with FSI responses (*M* = 2%, *SD* = 3.1%) becoming indistinguishable from the false alarm rate to single flash trials (*M* = 1.1%, *SD* = 2%) in the loud intensity condition [two-tailed, pairwise *t*-test, $t(29) = 1.34$, p = .20]. This suggests that the FSI is virtually absent when two visual flashes are paired with a single, 60 dB beep. Collapsed across the order manipulation, participants reported significantly more FSIs in the soft condition (*M* = 19.7%, *SD* = 16.3%), than in the loud auditory condition [planned two-tailed, pairwise *t*-test, $t(29) = 6.40$, p < .001, Cohen's *d* = 1.17].

As in our earlier analysis of auditory intensity order as a between-subjects factor, we found an interaction between trial type and order of intensity conditions [$F(1, 28) = 6.75$, p < .05, $\eta^2_{partial} = .194$]. Ignoring the impact of auditory intensity, participants who had the louder (60 dB) intensity condition first were *less* likely to see the flash-induced sound illusion overall. As with the SFI, experience with different auditory intensities over the course of the experiment impacted participants' reports of the FSI. However, the trend for the FSI goes in the opposite direction seen in figure 4.2 for the SFI. There was no three-way interaction of order, trial type, and intensity.

To better characterize the effect of sound intensity on performance in our matching task, we again turned to Signal Detection Theory measures of perceptual sensitivity (d') and bias (c) for our critical trials of interest. AV_V trials reported as two matching pairs were considered 'false alarms' and AV_AV trials reported as matching pairs were labeled 'hits.' Scores were analyzed via mixed ANOVA with between-subjects factor order (soft first, loud first) and within-subjects factor intensity (soft, loud). Analysis of d' scores revealed a main effect of intensity [$F(1, 28) = 147.2$, p < .001, $\eta^2_{partial} = .840$], a marginal main effect of order

[*F*(1, 28) = 4.06, p = .054], and a non-significant interaction between order and intensity [*F*(1, 28) = 3.23, p = .08]. To determine the influence of auditory intensity alone, we compared the d' scores within each ordered group. For those who had the soft intensity first, there was a significant difference between the soft intensity d' (*M* = 1.78, *SD* = 0.61) and loud d' [*M* = 3.66, *SD* = 0.59; two-tailed, pairwise *t*-test, *t*(14) = 10.42, p < .001, Cohen's *d* = 2.69]. Similarly, participants who had the loud intensity first had a significant difference between the soft intensity d' (*M* = 2.39, *SD* = 0.73) and loud d' [*M* = 3.78, *SD* = 0.52; two-tailed, pairwise *t*-test, *t*(14) = 6.94, p < .001, Cohen's *d* = 1.79]. These analyses suggest that greater auditory intensity increased the ability of participants to perceptually discriminate between AV_V and AV_AV trials, although this improvement was more noticeable in the group that had the soft intensity first.

Examining the role of total response bias (*c*) in our FSI trials, a two (order: soft first, loud first) by two (intensity: soft, loud) mixed ANOVA found a main effect for auditory intensity [*F*(1, 28) = 19.4, p < .001, $\eta^2_{partial}$= .41], a main effect for order [*F*(1, 28) = 4.60, p < .05, $\eta^2_{partial}$= .141], and an interaction between the two factors [*F*(1, 28) = 7.52, p < .05, $\eta^2_{partial}$= .212; see figure 4.4]. Within-group comparisons uncovered differences in response bias to AV_V trials in the soft first [two-tailed, pairwise *t*-test, *t*(14) = 7.10, p < .001, Cohen's *d* = 1.83], but not in the loud first [two-tailed, pairwise *t*-test, *t*(14) = 0.96, p = 0.35] orders. This suggests that the auditory intensity effects on the flash-induced sound illusion are partially driven by an increased willingness to respond that two beeps are co-present when sound is presented in the near-threshold (softer) condition, but primarily when the soft beep condition comes first.

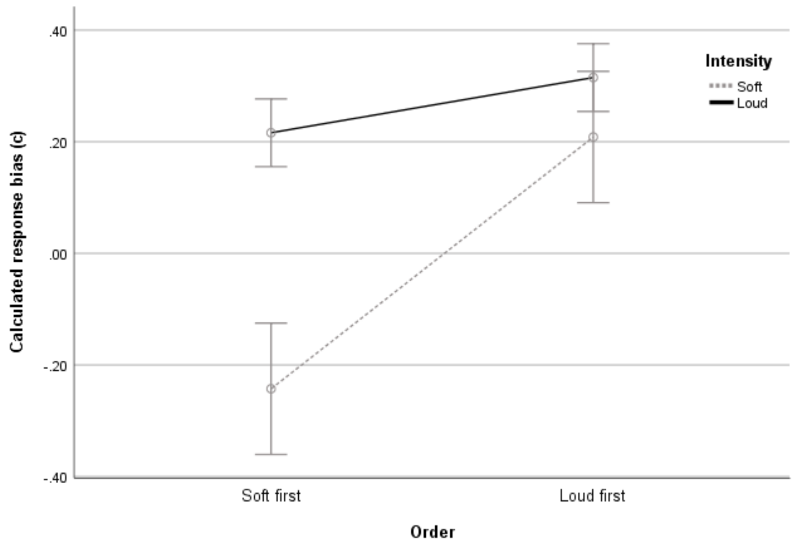

Figure 4.4. Response bias plotted as a function of sound intensity and order of intensity condition. There was an interaction between intensity and order. Participants in the soft first condition set a more liberal criterion for responding to AV_V trials as matching pairs. See text for detail. Error bars +/- 1 SEM.

3.3. Sound fusion trials: AV_A reported as AV

Our mismatching stimuli (e.g. AV_A) can trigger different types of false alarms. When experiencing the sound-induced flash illusion, the false alarm is to respond that two matching pairs of stimuli (AV_AV) were present. However, participants can also inaccurately indicate that a single pair of matching stimuli (AV) was present. To examine the influence of sound intensity on this latter type of response, we again used a two (order: soft/loud, loud/soft) by two (intensity: soft, loud) by two (trial type: AV, AV_A) mixed ANOVA, with order as the between-subjects factor and both intensity and trial type as within-subjects factors. The analysis revealed significant main effects for trial type [F(1, 28) = 835.7, p < .001, $\eta^2_{partial}$= .968] and intensity [F(1, 28) = 22.15, p < .001, $\eta^2_{partial}$= .442]. There was no main effect for order [F(1, 28) = 1.81, p = .19]. There was one significant interaction between trial type and intensity [$F(1, 28) = 137.6$, p < .001, $\eta^2_{partial}$= .831], which can be seen in figure 4.5.

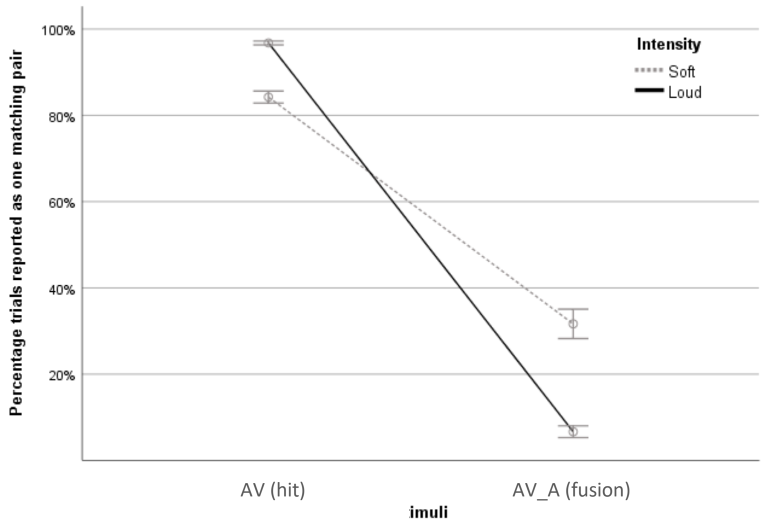

Figure 4.5. Proportion of trials reported as a single matching pair. There was an interaction between intensity and stimulus type. In the louder auditory intensity condition, participants were much less likely to miss the added beep. See text for detail. Error bars +/- 1 SEM.

Focusing on the effect of intensity on each trial type, we note that participants reported significantly more correct hits of the AV trials in the loud condition (*M* = 96.8%, *SD* = 2.4%), than in the soft auditory condition [*M* = 84.2%, *SD* = 7.8%; planned two-tailed, pairwise *t*-test, $t(29) = 10.16$, p < .001, Cohen's *d* = 1.85]. In the loud condition, participants also falsely reported AV_A trials as AV matching targets at a mean rate of 6.7% (*SD* = 7.4%). This was significantly less than the 31.7% of AV_A trials reported as AV targets in the soft condition [planned two-tailed, pairwise *t*-test, $t(29) = 9.38$, p < .01, Cohen's *d* = 1.71].

As is our earlier analyses, we investigated the degree to which changes in response behavior could be attributed to changes in perceptual discriminability (d') or bias (c). To address the first of these possibilities, d' scores were analyzed with a mixed ANOVA with between-subjects factor order (soft first, loud first) and within-subjects factor intensity (soft, loud). This analysis revealed a main effect of intensity [$F(1, 28) = 353.9$, $p < .001$, $\eta^2_{partial} = .927$], with no main effect for order and no interaction. This suggests that auditory intensity caused a change in the ability of participants to perceptually discriminate between AV_A and AV trials and accounts for at least some of the behavioral change seen across auditory intensity conditions.

Response bias (c) was analyzed using the same approach as with d'. This mixed ANOVA revealed a main effect of auditory intensity [$F(1, 28) = 9.34$, $p < .01$, $\eta^2_{partial} = .250$], but no significant main effect of order. Though not significant, there was a trend towards an interaction between intensity and order [$F(1, 28) = 3.48$, $p = .073$; see figure 4.6]. Within-group comparisons found a significant difference in response bias in the 'soft first' order condition [two-tailed, pairwise *t*-test, $t(14) = 3.05$, $p < .01$, Cohen's $d = 0.79$], but no difference in the 'loud first' condition [two-tailed, pairwise *t*-test, $t(14) = 1.01$, $p = 0.33$]. These results suggest that part of the increase in AV_A fusions is due to a more liberal response criterion adopted by those participants who hear the soft auditory condition first.

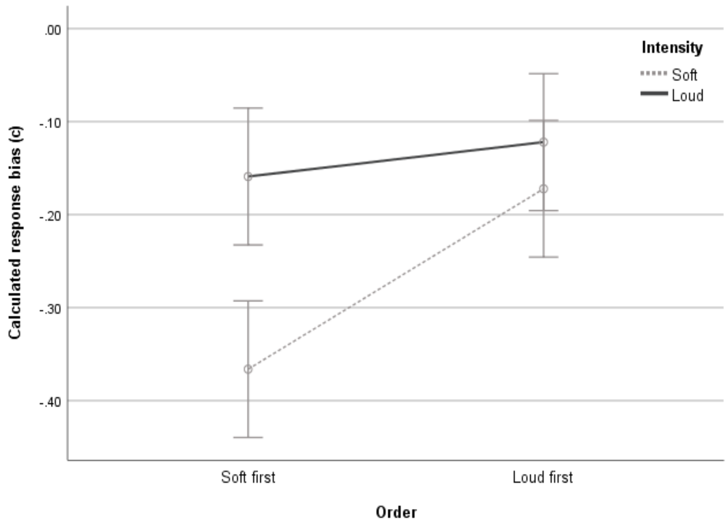

Figure 4.6. Response bias (c) values in reported a single pair of AV stimuli. There was main effect for intensity and a marginal interaction of intensity and order. This appears driven by the tendency to be more liberal in 'match' responses when the soft intensity is presented first in the experiment. When the loud intensity is played first, this may create an expectation for fewer 'match' responses. See text for detail. Error bars +/- 1 SEM.

3.4. Flash fusion trials: AV_V reported as AV

To investigate the effect of our experimental manipulations on the flash fusion effect, we again used a two (order: soft/loud, loud/soft) by two (intensity: soft, loud) by two (trial type: AV, AV_V) mixed ANOVA, with order as the between-subjects factor and both intensity and trial type as within-subjects factors. The analysis revealed significant main effects of trial type [$F(1, 28) = 40.7$, $p < .001$, $\eta^2_{partial} = .593$] and intensity [$F(1, 28) = 119.4$, $p < .001$, $\eta^2_{partial} = .810$]. A significant interaction between trial type and intensity was found [$F(1, 28) = 31.14$, $p < .001$, $\eta^2_{partial} = .527$], and can be seen in figure 4.7. Differences in reporting a single match in AV vs. AV_V trials were significant in both the soft [two-tailed, pairwise *t*-test, $t(29) = 6.25$, $p < .001$, Cohen's $d = 1.14$] and loud [two-tailed, pairwise *t*-test, $t(29) = 3.73$, $p < 0.01$, Cohen's $d = 0.68$] intensities, but were larger for the soft stimuli.

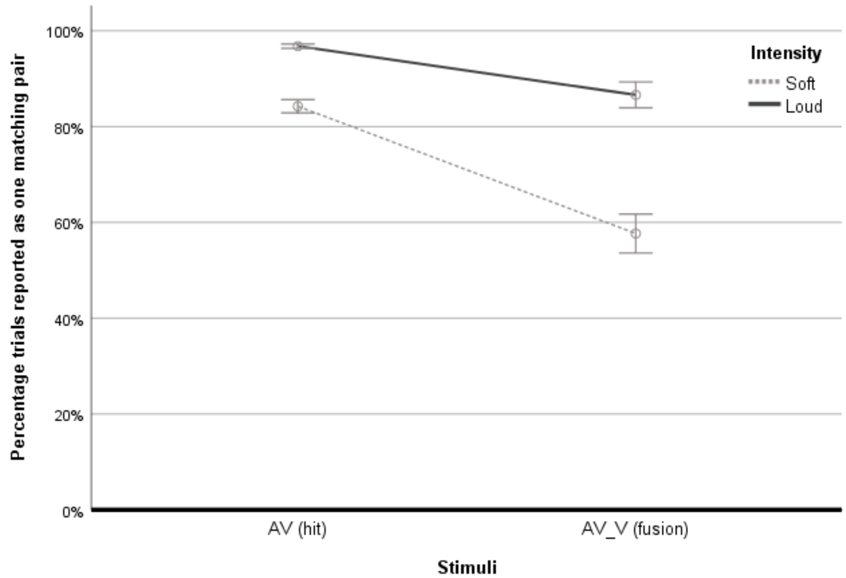

Figure 4.7. Proportion of trials reported as a single matching pair. There was an interaction between intensity and stimulus type. In the louder auditory intensity condition, participants were more likely to miss the added flash. See text for detail. Error bars +/- 1 SEM.

Although there was no main effect for order [$F(1, 28) = 2.73$, p = .11], the analysis revealed a significant interaction of stimulus and order [$F(1, 28) = 8.42$, p < .01, $\eta^2_{partial}$= .231]. The interaction occurs because the order manipulation had a larger impact on the fusion illusion than the AV stimulus. Similar to the pattern seen earlier in figure 4.2 for AV_AV and AV_A stimuli, figure 4.8 suggests that those who experience the softer auditory stimulus first will report the visual fusion illusion less often than those who encounter the loud intensity first.

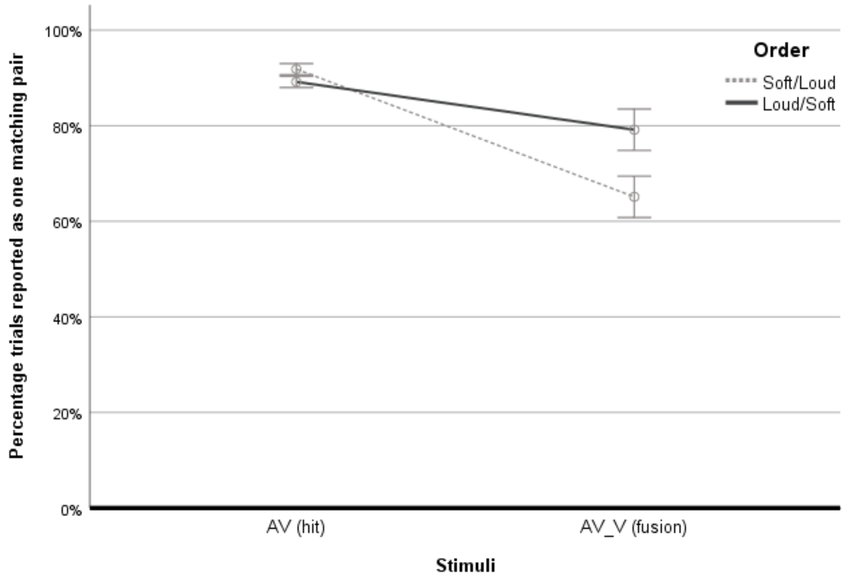

Figure 4.8. Proportion of trials reported as a single matching pair, with interaction between order and stimulus type. Those who encountered the loud condition first were more likely to report the flash fusion. See text for detail. Error bars +/- 1 SEM.

Again turning to our Signal Detection Theory measures, we calculated d' scores and c labeling AV_V trials reported as matching as 'false alarms' and AV trials reported as matching as 'hits.' A mixed ANOVA with between-subjects factor order (soft first, loud first) and within-subjects factor intensity (soft, loud) revealed only a main effect of order [$F(1, 28) = 61.89$, p < .001, $\eta^2_{partial}$= .689], with neither intensity nor the interaction of order and intensity significant. Those participants encountering the loud intensity condition first had a much reduced ability to distinguish AV trials from AV_V trials (see figure 4.9). This mirrors our earlier findings in the perceptual sensitivity comparison done for AV_AV and AV_A, where order of intensity had the same effect.

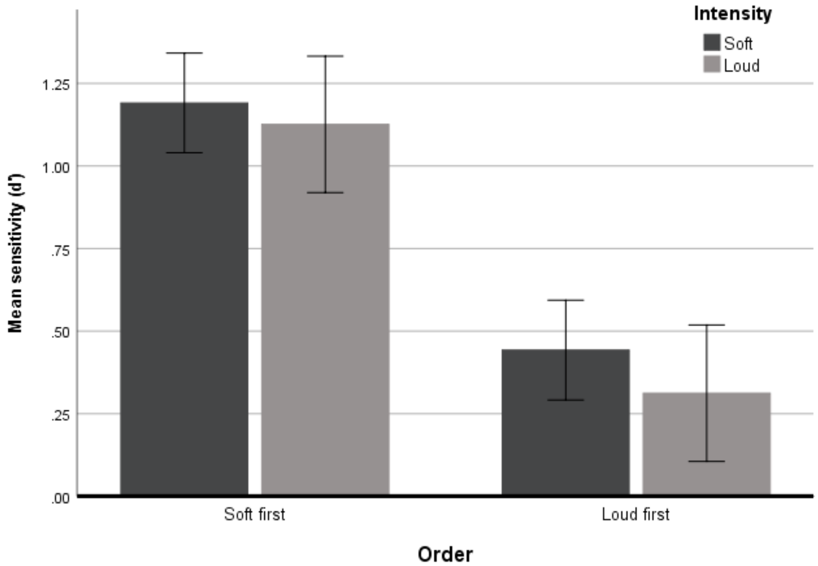

Figure 4.9. Sensitivity scores by order of intensity condition. Those who encountered the loud condition first were worse at discriminating stimuli. See text for detail. Error bars +/- 1 SEM.

Analysis of bias scores (*c*) revealed a main effect of auditory intensity [$F(1, 28)$ =99.45, p < .001, $\eta^2_{partial}$= .780]. There were no significant effects of order or interaction of intensity and order. Response bias was more liberal in the loud condition (*M* = -1.55, *SD* = 0.56) than the soft condition (*M* = -0.65, *SD* = 0.43). Similar to results in the sound-induced flash illusion analysis (section 3.2), this analysis suggests that part of the flash fusion effect is driven by an increased willingness to respond that one flash is co-present when a single loud beep is present, regardless of auditory intensity order.

4. Discussion

In the present experiment, we tested a number of hypotheses regarding the relationship between exogenous attention and multisensory integration. The overall goal was to determine how changes in auditory intensity may impact reports of audiovisual illusions in a bimodal target classification task. If multisensory integration is a largely automatic and obligatory process, then manipulations of auditory intensity and the order in which these changes occur should have little effect on reports of our multisensory targets. However, consistent with the findings in our previous experiments, we found that manipulating auditory intensity greatly affected the rates at which sound-induced flash illusions, flash-induced sound illusions, and related audiovisual fusions were reported in our bimodal matching task. We also uncovered complex relationships between the order in which these auditory intensities were presented and the perceptual sensitivities and response biases underlying those reports.

Focusing first on the effects of sound on visual perception, we found that louder auditory stimuli consistently increased reports of the sound-induced flash illusion in AV_A trials and flash fusions in AV_V trials. This effect of auditory intensity on SFIs and flash fusions matches the general pattern found in our unimodal visual task experiment described in chapter 3. These findings are also in accord with work by Andersen and colleagues (2004), who uncovered the flash fusion effect. This is particularly important as it suggests these findings extend across differing endogenous task-attention requirements and different experimental environments.

Our comparison of order of intensity condition also revealed that participants who received the softer auditory condition first reported fewer sound-induced flash illusions and flash fusions over the course of the entire experiment. This suggests that participants who saw the illusions and fusions *less* in the near-threshold condition, maintained the same perceptual discrimination ability in the louder intensity condition. This notion is supported by Signal Detection Theory analyses that show a change in total response bias (*c*) with order of presentation, but no parallel change in discriminability (d'). Response bias changed with intensity, regardless of order, even as perceptual discrimination remained the same once established in the first condition encountered.

In instances where visual stimuli impacted perception of auditory stimuli, we found a similar trend of sound intensity interacting with perceptual ambiguity. In AV_V trials, *flash*-induced sound illusions were reported in the near-threshold auditory condition, but were eliminated in the louder condition. In AV_A trials, auditory fusions were more frequently indicated in the near-threshold condition. The simplest interpretation is that the louder auditory stimulus has greater salience in the task. Greater salience would make it more difficult to falsely perceive a second beep in AV_V trials or miss the second beep in AV_A trials. In both cases, d' values for the critical stimuli were lower in the softer condition and higher in the loud condition. Additionally, bias scores became more liberal toward the number of auditory stimuli when d' was lowest (i.e. in the softer condition). We note that the louder auditory stimuli were much more noticeable and alerting than the visual stimuli in this experiment. This suggests that the ability of visual stimuli to drive auditory perception is greatest when they are at similar levels of salience.

Putting these findings together, we suggest that the perceptual system actively adapts to the bimodal task as it tries to resolve the perceptual ambiguity posed by incongruent AV_A and AV_V trials. This emerges in two separate stages during our experiment. In participants' first experience with the task, regardless of the order of intensity, they establish a perceptual set based on initial intensity (and salience) level. In the case of low intensity auditory stimuli, sensitivity (*d′*) is set higher and participants retain this discriminability throughout the experiment. In essence, the perceptual system establishes an expectation for what 'matching' stimuli look like. Once the perceptual set is established in the first condition, exogenous stimulus intensity can still affect responses. When auditory intensity changes in the second condition, it shifts response bias in the direction of the change. When the auditory intensity is higher, the response bias shifts further towards the more salient stimuli.

To place these findings in a broader context, it is helpful to briefly discuss them in terms of a few of the competing hypotheses that have been offered for the sound-induced flash illusion and related effects. Andersen and colleagues (2004, 2005) have evaluated and found evidence for three such hypotheses. The *discontinuity hypothesis*, originally offered by Shams *et al.* (2002), suggests that the discontinuous (i.e. repeating) stimulus will tend to dominate the continuous (i.e. unchanging) stimulus. We agree with Andersen and colleagues (2004) that this is true in a weak form, as the incongruity of the stimuli (e.g. AV_A or AV_V) creates the perceptual uncertainty necessary as a precondition for the effects noted above. However, additional factors are necessary to explain how this uncertainty is resolved in favor of fusions and may change in different experimental conditions.

The *information reliability hypothesis* (Andersen *et al*., 2004, 2005) suggests that one modality will dominate another if it provides more reliable information in the context of the task. Reliability can be determined in a few ways. Exogenous stimulus factors (e.g. intensity) can make a stimulus more immediately salient, giving it a greater impact in the task. In the case of the SFI, a louder auditory stimulus is more salient than the flash, and should increase the incidence of the illusion. We have described such salience in terms of sensory gain, but find the explanations compatible. History with similar stimuli should also have an effect as the participants may have a prior expectation that rapidly presented stimuli will tend to occur together. Together, this creates a bias to report incongruent stimuli (e.g. AV_A or AV_V) in the direction of the most salient stimulus modality. Our findings support this hypothesis as both stimulus intensity and order of intensity presentation strongly affected the SFI, flash fusions, and related effects in our experiment.

Finally, the *directed attention hypothesis* (Andersen et al., 2004, 2005; Warren, 1979) suggests that task instructions (endogenous attention) or stimulus intensity (exogenous attention) may engage different cognitive resources, changing responses to the incongruent stimuli. We agree with Andersen and colleagues (2004) that this appears to be an important factor in resolving uncertainty. As seen in our four experiments, endogenous task attention (experiments 1 and 2) and exogenous stimulus intensity (experiments 3 and 4) impact perceptual sensitivity and total bias when responding to the incongruent AV_A and AV_V stimuli. We have expressed such attention effects as modulating sensory gain, alternatively increasing signal strength or selectively gating a sensory signal.

These functional hypotheses help frame some of the results in our experiments. While discontinuity creates the preconditions for the illusions, information reliability and directed attention interact to resolve perceptual ambiguity in the context of the task conditions. We now turn to some of the electrophysiological data that can speak to the brain behavior associated with our results.

Chapter 5: Decision and response monitoring in illusory perception

1. Introduction

In everyday life, we make decisions and take actions based on a multitude of sensory experiences. These decisions are often made with little thought, and our awareness of the events may be fleeting. At the grocery store, you may reach for a familiar brand of cereal or coffee almost automatically. If it is readily identifiable and expected, the perceptual decision and action are easy. Other decisions take more time and attention. To determine if a particular melon is ripe, you might look at its color, feel its firmness, or tap to hear a hollow sound. You then make a decision (ripe or not ripe) and take the appropriate action (buy that melon or try another). Such decisions are made under uncertainty, as you won't know whether your judgement and subsequent action were correct until later. The best you can do is to use all the information available and update your decision process when you get the answer.

The broad goal of the current project has been to examine the effects of endogenous and exogenous attention on the perception of multisensory stimuli. Each individual perception reflects a decision as to how to classify the stimuli presented. Some of these stimuli call for difficult perceptual distinctions, leading to uncertain responses. The primary means for testing the perceptual judgements in our experiments has been through the use of an ambiguous audiovisual stimulus. When two beeps are presented with one flash (AV_A), this can be perceived as containing a second visual flash (AV_AV). We have used this *sound-induced flash illusion* (SFI; Shams, *et al.*, 2000; 2002) as an index of multisensory processing and its effects on perceptual decisions. In this chapter, we will examine event-related brain potentials (ERPs) collected during our first experiment with the SFI (reported in chapter 2). Our objective is to determine whether participants treated the real stimuli (i.e. AV_AV) and illusory stimuli (i.e. AV_A) differently during target and response evaluation. We will also compare these evaluations and responses across different task-attention conditions. This is an important step in better understanding how endogenous attention can impact outcomes of multisensory perception. Additionally, it can provide some insight as to the role of multisensory integration during perceptual discrimination and learning.

When making perceptual decisions under ambiguous circumstances, the ability to learn from experience and reduce future uncertainty is of critical importance. Recent work has highlighted the advantages multisensory processing provides for learning and performance in a variety of tasks (Shams & Seitz, 2008). Our earlier experiments demonstrated that exogenous and endogenous attention factors can influence the frequency of false alarms in AV_A stimuli, and additionally alter perceptual sensitivity and response bias to those stimuli. One question that remains is whether or not people have any awareness that they are making judgment errors during these tasks. Does exposure to the SFI provide information to the participant that could be used to overcome the perceptual ambiguity caused by the mismatching stimuli? If so, how does task-directed attention impact the use of such information? During our experimental debriefings, participants sometimes said they felt as though some of the target trials looked different, but they didn’t indicate any awareness of the specific illusion. Though we did not directly assay awareness of the experiential quality of the illusion or judgement confidence during our experiments, other groups have made some effort in this direction.

In one study that examined the phenomenological experience associated with ‘seeing’ an illusory flash, McCormick & Mamassian (2008) found that the illusory flash had a measureable visual contrast. When preceded by a high-contrast white flash, an illusory flash triggered similar responses as those seen with a second, low-contrast white flash. Although it was not as salient as a real flash, the illusory flash did provide enough sensory information to satisfy the detection of a second flash on a significant number of trials. The authors also noted that this led to a decrease in perceptual sensitivity (measured with *d′*) to a single flash when accompanied by two beeps. Additional work suggests that people can change their responses to SFI trials under heightened monetary reward conditions, but that they continue to mistakenly respond to the SFI stimuli once those rewards are taken away (Rosenthal *et al.*, 2009). This finding suggests that there is some detectable difference in the SFI from the veridical AV_AV, but that it does not provide a consistent enough signal to enable the participant to always 'see through' the illusion.

In chapter 2, our first experiment found that endogenous task-directed attention could modulate the degree to which participants would experience a multisensory illusion. In one task, participants were asked to attend to only the visual stimuli (*focused attention*) and respond when they saw two flashes. In a second task, the same participants attended to both visual and auditory stimuli and responded when the number of visual stimuli matched the number of auditory stimuli (*conditional bimodal attention*). When individuals attended to both auditory and visual channels in our bimodal matching task, they reported more sound-induced flash illusions than when ignoring auditory stimuli in a unimodal visual task. This effect was larger when the auditory stimuli were presented at near-threshold levels for each individual (experiment 2).

One of the hypotheses we evaluated was that bimodal attention to the auditory channel could increase sensory gain of auditory information and make participants more aware of the differences in the auditory and visual stimuli presented. An increase in sensory gain might enable better accuracy in the task and reduce the number of SFIs reported. Instead, the spread of attention across sensory channels appeared to enhance the multisensory processes responsible for the illusion. This resulted in more perceptual errors. We also found that participants had greater difficulty in distinguishing valid targets (AV_AV) from illusory targets (AV_A) in the bimodal task, as indicated by reduced perceptual sensitivity (d'). While these are interesting psychophysical results, it might be illuminating to evaluate the brain behavior associated with these findings.

In the first of the two experiments described in chapter 2, we recorded electroencephalographic (EEG) data associated with the unimodal and bimodal tasks. The high temporal resolution of ERP recordings is an excellent tool for investigating both the onset of sensory integration processes and their subsequent effects on response selection and evaluation. Below, we present these data as they relate to decision making and response processing involved in distinguishing valid from illusory targets. Previous EEG/ERP studies of the sound-induced flash illusion have tended to focus on the sensory activity associated with perception of the illusion. This generally occurs within the first 300 ms following stimulus presentation. Before turning to brain activity related to decision making, it is useful to review some of these findings.

Shams and colleagues performed the first ERP study of the SFI by recording visually evoked potentials (VEPs) to the illusion and valid double flashes (Shams *et al.*, 2001). In that study, the intent was to determine whether the perception of the illusory flash could be correlated with neurophysiological activity in primary visual areas. The authors suggested that perception of the illusory flash implied activation of a new visual token or representation, which was induced by the auditory stimulus. If correct, it may be possible to view such activation in early unimodal visual area such as V1. Shams and colleagues concluded that the VEP to the illusory flash strongly resembled the VEP to the physically present flashes. This was taken as evidence that the illusory flash could be caused by activation occurring as early as V1. However, their recordings used a small set of three electrodes (O1, Oz, and O2) over the occipital cortex, so spatial resolution was very limited. Arden *et al.* (2003) used a Laplacian derivation of ERPs to suggest that integrative effects associated with the SFI were localized to area V1. Finally, an fMRI study reported activity in V1 and additionally implicated the right superior temporal sulcus and superior colliculus in this multimodal illusion (Watkins *et al.*, 2006). Although these are intriguing results, some methodological concerns and limitations of this early work, most notably limited electrode coverage in the ERP studies and poor temporal resolution in the MRI study, limit the interpretation of these results.

In a more extensive series of ERP studies, Mishra and colleagues found a number of interesting electrophysiological correlates to the SFI (Mishra, Martinez, Sejnowski, & Hillyard, 2007). The first study isolated an early positive component over occipital areas, termed the PD120, which was seen most prominently in individuals who regularly experienced the SFI. This component, along with later positive (PD180) and negative (ND270) deflections, were also found in non-illusory multisensory interactions. Additional negative components localized to auditory cortex (ND110) and superior temporal cortex (ND130) were associated with the SFI in a trial-by-trial analysis. Importantly, cortical activity associated with veridical flashes was seen to have a different scalp topography from the SFI.

This suggests that the cortical activation uniquely associated with the illusory flash is generated differently than physically present flashes.

In a second study that is especially relevant to the present project, Mishra, Martinez, & Hillyard, (2010) explored the effect of spatial attention on the ERP components associated with the SFI. Participants were asked to attend to either an upper or lower visual field, responding only when stimuli were presented in the attended region. When SFI stimuli were present in the attended field, participants had larger PD120, PD180, and ND250 components compared to those triggered within the unattended field. These SFI-related components were temporally and topographically very similar to those found in their earlier 2007 study. Although there was no motor response to the unattended stimuli which would make greater association of the ERP components with the observed versus unobserved illusory flashes, this is strongly suggestive evidence that the multisensory integration responsible for the SFI can be manipulated by endogenous spatial attention.

The studies by Mishra and colleagues were done using a larger electrode montage and a greater number of participants than previous ERP experiments within this paradigm. Additionally, they used different analysis criteria for indexing the presence of a multisensory effect in the ERP data. These issues will be discussed in the Methods section. This electrophysiological work and follow-up studies with the SFI paradigm are particularly noteworthy as they isolate between-subject neurological measures which appear to correspond to the subjective experience of the illusion.

While Shams and colleagues have done additional research with this illusion (Shams, Iwaki *et al.*, 2005; Shams, Ma *et al.*, 2005), most of this work focused on establishing the SFI as a robust sensory illusion and confirming its early (<300 ms) impact on the visual cortex.

This experimental paradigm has not been fully exploited to determine the stimulus- and attention-modulated parameters of this multimodal illusion. The later (>300 ms) EEG effects concurrent with decision and response factors could also benefit from additional study. To this end, we now turn to a discussion of the ERP components that can help evaluate these decision and response factors.

There are two ERP measures typically used to quantify error-related activity following a motor response. The first is the error negativity (Ne; Falkenstein, et al., 1990), also known as the error-related negativity (ERN; Gehring et al., 1990, 1993). The ERN is described as a frontocentral negative deflection that peaks within the first 100 ms after an incorrect response. It is generally understood to be generated as a result of a mismatch between a correct response required by a task and the incorrect response actually produced (Falkenstein et al., 1990; Scheffers & Coles, 2000). However, it has been shown that the ERN may also be produced following correct responses (Carter et al., 1998; Vidal et al., 2000). It may be more broadly interpreted as indicating a response conflict monitoring process that is often a precursor to explicit error detection (Yeung et al., 2004), though the exact interpretation is still debated (Falkenstein et al., 2000; Gehring et al., 2018).

A second measure, the error positivity (Pe; Falkenstein, Hohnsbein, Hoorman, & Blanke, 1991) is a centro-parietal focused positivity occurring approximately 200 – 500 ms following an incorrect response. The Pe is thought to reflect an error monitoring system that is independent of the ERN (Di Gregorio, Maier, & Steinhauser, 2018; Falkenstein et al., 2000; Overbeek, Nieuwenhuis, & Ridderinkhof, 2005). Pe amplitude has been found to increase with subjective error awareness and more consistently indexes explicit recognition of a motor error than the ERN (Overbeek et al., 2005; Steinhauser & Yeung, 2010). The Pe has also been related to confidence in responses, with amplitude increasing with the certainty that an error has been made (Boldt & Yeung, 2015).

Though there are competing hypotheses regarding the function of the Pe (see Wessel, 2012 for a recent review), one account suggests that the Pe reflects the degree to which *accumulating evidence* of an error exists (Steinhauser & Yeung, 2010). This evidence may come from a variety of sources, including the sensory, cognitive, and autonomic activity following a series of responses. The uncertainty reflected by the Pe can occur even during a correct response (Hewig et al., 2011). For our purposes, it is important to note that Pe amplitude will rise with an increasing sense that an error is occurring, even if the earlier processing underlying the ERN does not indicate that a response error has occurred (Di Gregorio *et al.*, 2018).

Another related ERP component, the stimulus-locked P3 (or P300), is a positive-going waveform that tends to peak 250-500 ms following target onset. The P3 has a midline distribution, increasing in amplitude toward posterior electrodes. It has long been known to index the processing time required for target categorization and is relatively independent of response selection (Kutas, McCarthy, & Donchin, 1977). Amplitude of the P3 is also known to increase with confidence in a target categorization, showing a generally positive relationship with the discriminability (d') of targets (Hillyard, Squires, Bauer, and Lindsay, 1971). The P3 is also sensitive to task demands, with tasks requiring greater attentional resources decreasing amplitude of the potential (Kok, 2001; Polich, 1987). While the P3 doesn’t directly signal the occurrence of a perceptual or response error, it can indicate the degree of confidence and effort in a perceptual judgment.

What can these ERP measures tell us about perceptual decisions involving the sound-induced flash illusion? If participants are unable to distinguish AV_AV from SFIs in the AV_A stimuli, we would expect to see similar P3 potentials to detections in both stimuli and across both unimodal and bimodal attention conditions. This would reflect equal confidence in the detections made. If, however, some perceptual differences are still detectable even when the illusion has been identified as a target, then we should see a diminished P3 for AV_A stimuli compared to valid AV_AV targets. When comparing across attention conditions, we know that perceptual sensitivity decreased in the bimodal attention condition. This suggests that we should find lower amplitudes for P3 potentials of both valid and illusory stimuli in the bimodal attention condition. Additionally, the greater cognitive complexity of the matching task may contribute to a reduced P3.

It is less clear what would be expected in our response error monitoring measures. An ERN should be present when there is a response error, but only if the actual response does not match the desired or required response. The power of the SFI is that it can trigger response behavior similar to the valid target. If the participants are responding to a true perceptual event in the SFI, its ERN may be similar in amplitude to the valid target. However, if the error monitoring system that generates the ERN has sufficient information to identify the SFI as a false alarm, then we would expect to see a larger ERN to those stimuli.

The Pe, however, is sensitive to confidence that a response error has been made. If the *accumulating evidence* account of the Pe is correct and participants have some awareness that the illusion looks different from the valid target, we would expect to see an increase in amplitude for the SFI responses relative to the valid AV_AV responses. As sensitivity decreased in the bimodal task, we expect Pe amplitude to again increase relative to the valid trials, reflecting a growing belief that responses to the SFI are incorrect.

Together, these ERP potentials can help us understand the motor responses generated in our manipulation of endogenous attention in the SFI paradigm. We will also explore what these ERP results mean for the *directed attention hypothesis* (Andersen et al., 2004, 2005) discussed at the end of chapter 4.

2. Methods

2.1 Participants

Fifteen healthy adults (10 women, 5 men; mean age 22.2 years) participated in the experiment after giving written informed consent, in accordance with the University of California, San Diego Human Research Protections Program. Three participants were excluded from analysis due to noisy EEG data and excessive loss of trials during artifact rejection. All remaining individuals had normal or corrected-to-normal vision, reported normal hearing, and were naïve to the purpose of the study. Participants received course credit or monetary compensation for participation and were debriefed following the study.

2.2. Apparatus and stimuli

The display equipment and basic stimuli used in this experiment is described in chapter 2. Seven stimulus trials were presented over the course of the experiment (see figure 5.1). An eighth 'blank' trial was also presented in order to facilitate some subtractions in our ERP analyses. When presented together as a pair (e.g. AV), the visual stimulus would follow the auditory stimulus with an onset asynchrony of 10 ms. In instances where only visual stimuli were presented (e.g. V and V_V), we included a blank 'null' stimulus to ensure the timing would match onset of visual stimuli in other trials. Repeated stimuli (e.g. V_V and A_A) were separated by an SOA of 70 ms.

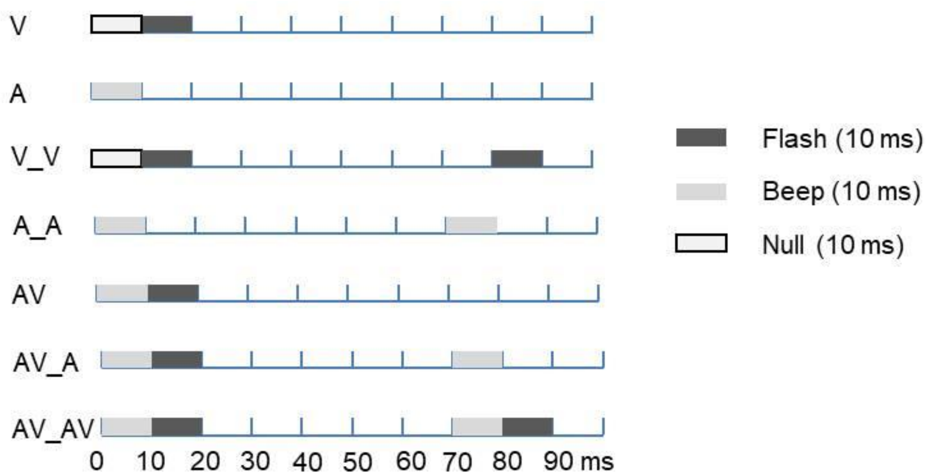

Figure 5.1. Stimulus timing in experiment 1. A blank 'null' stimulus was placed in front of unimodal visual stimuli to maintain temporal register with audiovisual stimuli. A 'blank' trial was also included to allow some subtractions in the ERP analysis.

2.3. Procedure

The experimental procedure is described in the first experiment in chapter 2. Briefly, participants were asked to perform a go/no-go task under two different sets of task instructions. In the first, they were asked to respond via button press as quickly and accurately as possible whenever two flashes (i.e. V_V and AV_AV) appeared on the screen. They were asked to ignore the auditory stimuli, focusing only on visual stimuli. In the second condition, participants were asked to press the response button whenever the number of flashes matched the number of beeps presented (i.e. AV, AV_AV). This was the bimodal task-attention condition. In both cases, the AV_A stimuli might elicit a response if perceived as the SFI. All stimulus trials were presented in pseudo-random order 90 each, for a total of 720 trials per task-attention condition. All stimuli were equally probable on a given trial.

2.4. EEG/ERP recordings

Continuous EEG data were recorded from 68 channels using an ActiveTwo data acquisition system (BioSemi B.V., Amsterdam, Netherlands). The ActiveTwo system employs “active electrodes” which amplify signals at the scalp and reduce the impact of movement and environmental artifacts. The system has an input range of -264 to 264 mV and uses a 24-bit 4th order Delta-Sigma AD converter with a dynamic range of 115 dB. The BioSemi system uses a driven right leg (DRL) circuit to provide a baseline measure of electrode-skin conductance (Common Mode voltage). All recording electrodes were kept within 25 µV of this reference value. EEG data was digitized at 256 Hz for offline analysis. Sixty-four channels of scalp data were recorded from electrode sites according to the International 10-10 system. Horizontal and vertical electro-oculograms were monitored via electrodes placed beneath each eye and at the outer canthi.

EEG recordings were imported into EEGLAB for signal processing (Delorme & Makeig, 2004). The data was high-pass filtered at 0.1 Hz with a 2-way least squares FIR filter and referenced to the average of the left and right mastoids. An independent component analysis (ICA) blind source separation algorithm was used to identify eye movement artifacts in the continuous data. Following visual inspection of the isolated components, these artifacts were removed from the data (Jung et al., 2000a, 2000b). Data were then processed in ERPLAB, a Matlab-based plugin for EEGLAB (Lopez-Calderon & Luck, 2014). Continuous EEG data were time-locked to stimuli and responses of interest and extracted into one-second epochs (-200 ms pre-event and 800 ms post-event). Automated artifact rejection was performed to remove trials containing muscle artifact or excessive noise. The remaining epochs were low-pass filtered with a Gaussian finite impulse response function (3 dB attenuation at 50 Hz) and averaged for each participant.

2.5. ERP analysis

Following earlier work with the SFI (Mishra et a., 2007), we examined cortical activity associated with perceived sound-induced flash illusions by creating difference waves from the illusory (AV_A) and unimodal trials (A_A, V). The subtraction was performed as follows: [(AV_A) + blank] – [(A_A) + V], where the 'blank' was an empty trial during which no other stimuli were presented. It is believed the blank trial captures activity that is present in all trials, such as the contingent negative variation (CNV) or other anticipatory waveforms (Teder-Salejarvi et al., 2002; Talsma & Woldorff, 2005). By including the blank trials, the subtraction balances such activity in equal amounts in both additive elements. Previous work did not include this control measure. The remaining cortical activity reflects the 'super-additive' multisensory activity associated with the binding of the AV stimuli, plus the activity associated with the perception of the illusory flash.

For analysis of the SFI difference ERPs, mean voltages were calculated for clusters of electrodes over the deflections analogous to Mishra et al.'s (2007) PD120, PD180, and ND270. For the PD120, we computed the mean voltage over a window of 100-132 ms for PO7, PO3, POz, PO4, PO8, O1, Oz, O2, and IZ. For PD180, we examined the window of 195-220 ms over F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2, P1, Pz, and P2. Finally, for ND270, we took the mean voltage of 252-284 ms over FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2. The means were compared by *t*-tests with the 100 ms baseline average.

P3 mean amplitude was measured within the time window of 300 – 500 ms for targets (AV_AV) and illusory false alarms (AV_A) common to both attention conditions. The P3 was compared a 100 ms pre-stimulus baseline. For our analyses, we focused on electrodes around the midline in the following groupings: frontal (F1, Fz, F2), fronto-central (FC1, FCz, FC2), central (C1, Cz, C2), centro-pariental (CP1, CPz, CP2), parietal (P1, Pz, P2), parieto-occipital (PO3, POz, PO4), and occipital (O1, Oz, O2). Repeated measure ANOVA was used to compare stimuli across task-attention condition and electrode groupings. The Greenhouse-Geisser correction was applied when assumptions of sphericity were violated.

Analysis of error-related activity was conducted on response-locked averages associated with targets and false alarms. Epochs were extracted and averaged over a 200 ms pre-response and 500 ms post-response window. In keeping with previous studies, we quantified the ERN as the largest negative peak occurring at Fz, FCz, or Cz in the first 100 ms following the response (Falkenstein et al., 2000). Pe was measured as the mean amplitude over an interval 300 to 500 ms post-response at CPz and Pz. Difference ERPs (error minus correct) were used to highlight both ERN and Pe waveforms. Repeated measures ANOVA and pairwise *t*-tests were used for statistical analyses on the most appropriate electrode sites following topographical inspection.

3. Results & Discussion

Response times, accuracy, and Signal Detection Theory measures were reported in chapter 2, experiment 1. Below, we present analyses for early sensory interactions occurring < 300 ms for comparison with earlier studies of the SFI (Mishra *et al*., 2007). We then turn to components primarily occurring after 300 ms which reflect decision making and response evaluation.

3.1. Early ERP characteristics

In keeping with a study by Mishra and colleagues (2007), we examined the early sensory components associated with the SFI (AV_A) and its unimodal elements (A_A and V). Figure 5.2A presents grand averages (over 12 participants) at representative auditory and visual electrode sites. The unimodal auditory stimulus A_A evoked typical N1 (~110ms) and P2 (~200 ms) components. Likewise, the unimodal visual ERP to the V stimulus displayed typical N1 (~190 ms) and P2 (~220 ms) components over occipital sites, though these were slightly shifted in time due to the 10 ms null marker preceding the V stimulus.

We calculated the SFI difference ERP (see ERP Methods), measuring mean amplitudes against the pre-stimulus baseline. In figure 5.2B, we find a negative peak similar to the ND270 first identified by Mishra *et al.* (2007). This negative deflection (M = -3.45 μV, SEM = 1.48 μV) was significantly different from the baseline [two-tailed, pairwise t-test, $t(11) = 3.01$, $p < .05$, Cohen's $d = 0.87$]. The scalp map shows a comparable distribution over central electrode sites (figure 5.2C). The ND270 was said to be an indicator of multisensory integration and was seen in other audiovisual stimuli examined by Mishra and colleagues (2007). The appearance of a similar component in our experiment suggests that we are reproducing multisensory interactions like those seen in their study.

We also find a positive deflection similar to their PD180, although it is about 30 ms later in latency. However, this peak did not achieve a significant difference from baseline [two-tailed, pairwise t-test, $t(11) = 0.36$, $p = .724$]. We also did not find a clear P120 component as identified by Mishra and colleagues. These differences may be due to the different stimuli used across experiments and the greater power of their earlier study. Their study had a greater number of participants (n = 34) and stimulus trials, allowing for better resolution of early sensory components. However, the similar morphology of the difference ERP and the later negative component finding suggests that our manipulation produced similar cortical activity to that seen by Mishra *et al.* (2007).

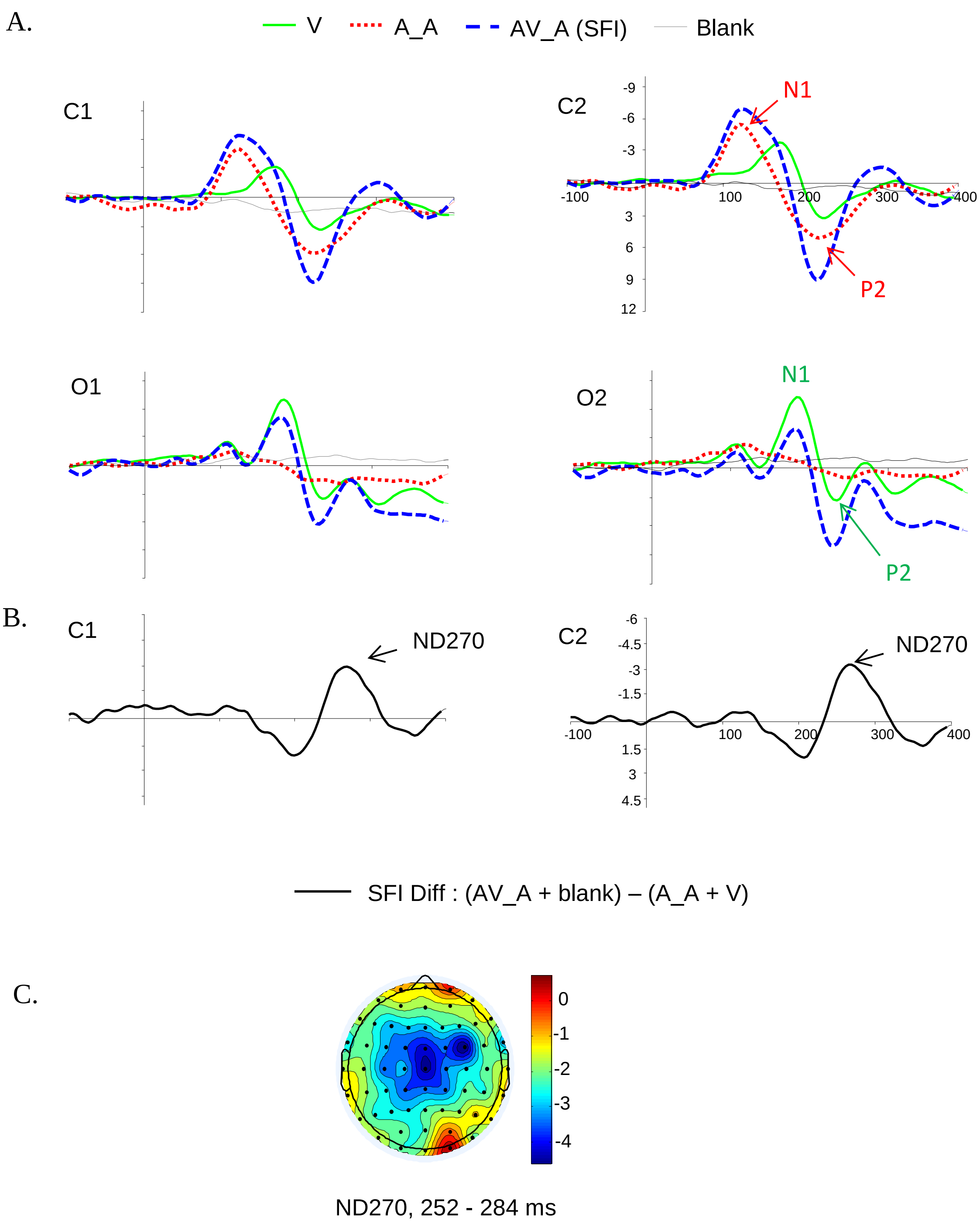

Figure 5.2. Grand average ERPs to SFI and unimodal components in visual task. A. ERPs of AV_A illusion trials, along with A_A, V and blank trials. B. Difference waves with unimodal components subtracted from SFI trials. The negative deflection appears similar to the ND270 found by Mishra *et al.*, 2007. C. Scalp map showing the voltage distribution for the negative deflection similar to the ND270.

3.2. P3 to targets and SFI illusions

Comparing P3 response to valid targets (AV_AV) and illusions (AV_A) across attention conditions, we found main effects of stimulus [$F(1,11) = 13.96$, $p < .01$, $\eta^2_{partial} = 0.558$] and a main effect of task attention condition [$F(1,11) = 9.06$, $p < .05$, $\eta^2_{partial} = 0.452$]. Valid AV_AV targets had larger P3 mean amplitudes than illusory stimuli, and stimuli in the visual-only attention condition had larger P3 amplitudes than the bimodal attention condition (see figure 5.3). The P3 response to both AV_AV targets and AV_A illusions were largest over posterior sites, so we next focused on a nine electrode cluster composed of P1, Pz, P2, PO3, POz, PO4, O1, Oz, and O1. A follow-up repeated measures ANOVA with factors of stimulus type and attention condition on this smaller cluster again found main effects of stimulus [$F(1,11) = 18.61$, $p < .001$, $\eta^2_{partial} = 0.628$] and attention condition [$F(1,11) = 6.44$, $p < .05$, $\eta^2_{partial} = 0.369$], but no interaction of these factors. A stimulus by electrode interaction [$F(2,22) = 7.07$, $p < .01$, $\eta^2_{partial} = 0.391$] and attention by electrode interaction [$F(2,22) = 5.16$, $p < .05$, $\eta^2_{partial} = 0.319$] showed activity to be lateralized left of midline (figure 5.4).

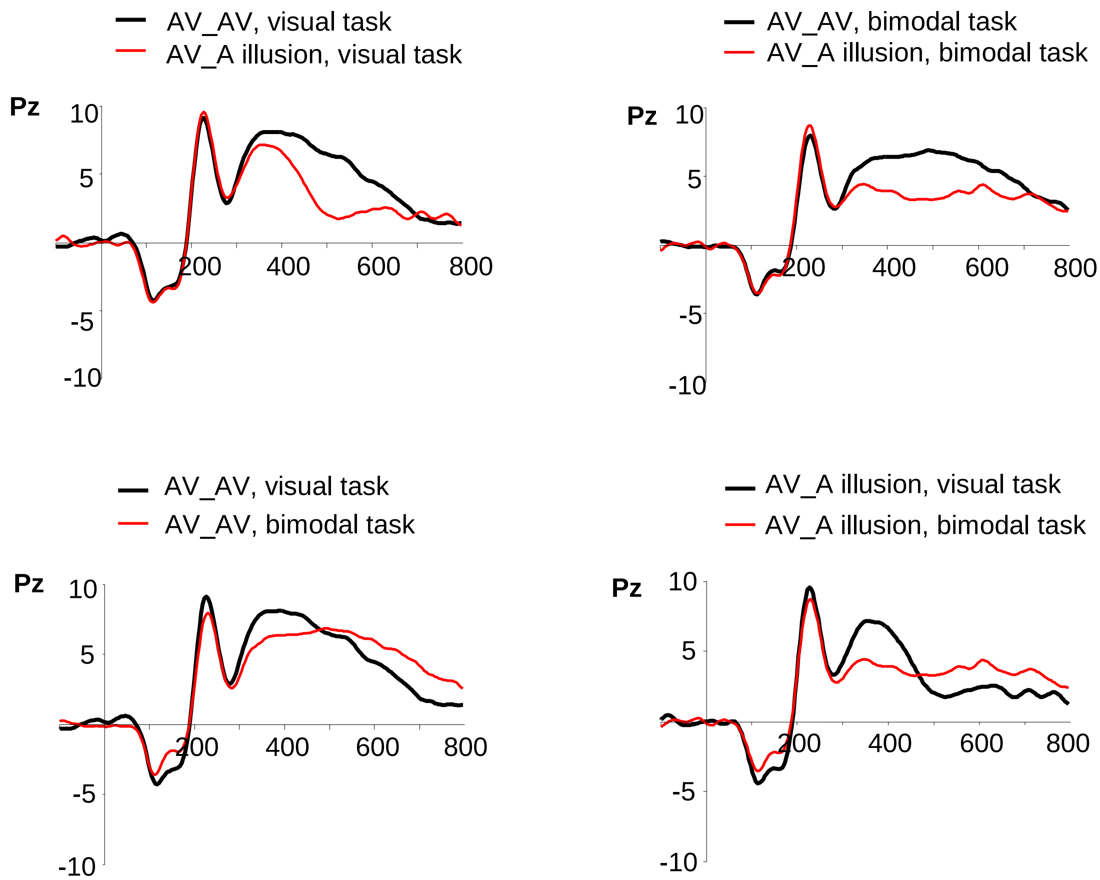

Figure 5.3. Grand average P3s (across 12 participants) to targets and illusions. Note that these plots are positive up. P3 amplitude is larger for valid trials than illusions. The visual task also had a larger mean amplitude than the bimodal task. Data was low-pass filtered at 25 Hz for plotting.

Our analyses suggest that the cortical processes reflected in the P3 responded independently to the nature of the target (or illusion) and the attention condition under which it was experienced. The valid targets were accompanied by a larger amplitude P3, suggesting greater confidence in the perceptual judgements associated with those targets over the illusory stimuli. Smaller P3 amplitudes in the illusory AV_A stimuli could also reflect more variable categorization processing time, as the RTs for these stimuli were approximately 100 ms later than valid targets. The overall smaller P3 amplitudes in the bimodal attention condition likely reflect the greater engagement of cognitive resources in the more difficult bimodal matching task. The reduced amplitude in the bimodal task could additionally reflect reduced confidence in the responses as the incidence of the SFI increased as valid target accuracy decreased slightly.

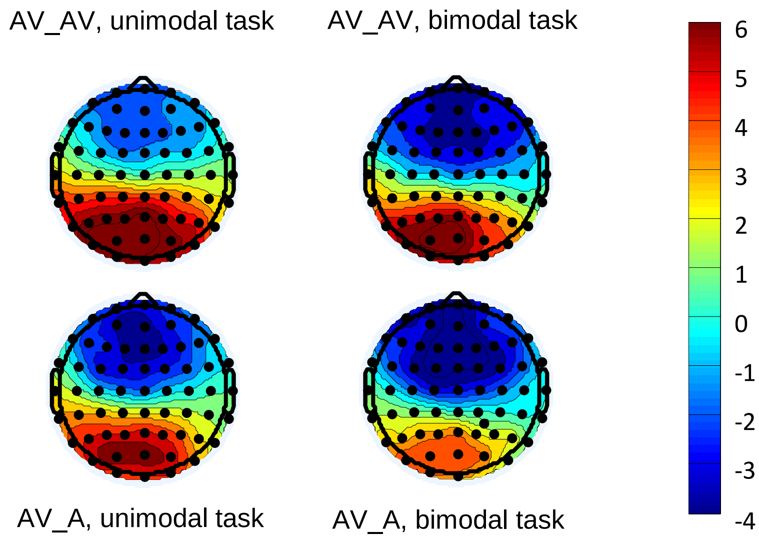

Figure 5.4. Scalp distributions of the P3 for target and illusory stimuli. Distributions, averaged over 300 - 500 ms, were similar across stimuli and conditions. Larger P3 amplitudes were seen in valid AV_AV targets (top).

3.3. Error-related components to targets and SFI illusions

For our first planned analysis of ERPs associated with response monitoring, we examined ERN and Pe components in our unimodal attention condition. To establish a baseline, we compared correct response to valid visual targets (V_V) and false alarms to a single flash (V). However, due to the relatively low false alarm rate in the V condition, we had insufficient trials across all participants to perform a reliable test of significance.

For our double pairs of audiovisual stimuli (AV_AV) and illusory false alarms (AV_A), we were able to compare across attention conditions as AV_AV was a valid target in both tasks. Visual inspection showed the ERN to be maximal at FCz, so this was used for statistical comparisons. In a two (stimulus: AV_AV, AV_A) by two (attention: unimodal, bimodal) repeated measures ANOVA at FCz, we found a significant main effect of attention [$F(1,11) = 8.56$, $p < .05$, $\eta^2_{partial} = 0.438$], but no main effect of stimulus and no interaction of stimulus and attention. The size of the ERN in the bimodal matching task (M = -1.71 μV, SEM = 1.95 μV) was approximately half the size of the ERN in the unimodal task (M = -3.51 μV, SEM = 2.84 μV). The reduced ERN in the bimodal matching task suggests that it is more difficult to distinguish correct from erroneous responses in that task. This is consistent with the behavioral findings of more SFIs in this condition, along with a reduction of perceptual sensitivity (d').

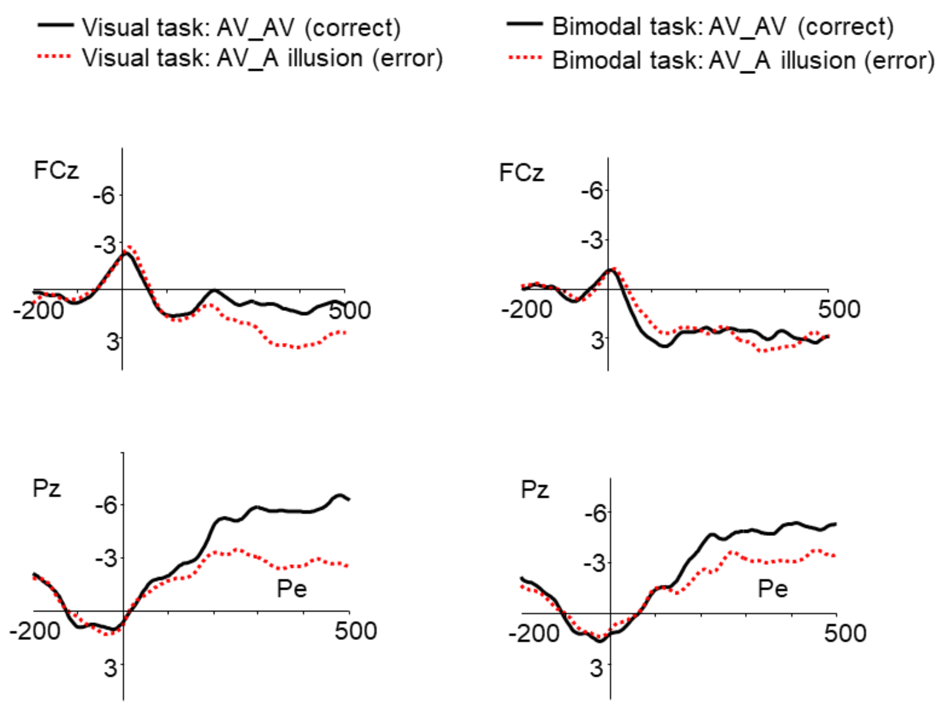

Figure 5.5. ERN and Pe to audiovisual stimuli. The unimodal visual task is on the left, bimodal task on the right. On the top left, a small ERN can be seen for both valid AV_AV and illusory AV_A trials. This ERN was reduced in the bimodal task on the right. The Pe for the illusion was larger than seen in the valid AV_AV trials in both attention conditions.

As seen in figure 5.5, there appears to be a small ERN for both the valid target (AV_AV) and the false alarm (AV_A) mistaken for that target. Vidal and colleagues (2000) have previously found that that correct responses can be accompanied by small ERN components. Notably, incorrect trials still exhibited larger ERN components than correct trials in their experiments. Our results suggest that the conflict monitoring reflected by the ERN is unable to distinguish between the appropriateness of responses to our veridical and illusory stimuli.

Comparison of the Pe was done at Pz, the site of maximal activity 300 to 500 ms post-response. As with the ERN, we performed a two (stimulus: AV_AV, AV_A) by two (attention: unimodal, bimodal) repeated measures ANOVA. The analysis revealed a main effect of stimulus [$F(1,11) = 13.90$, $p < .01$, $\eta^2_{partial} = 0.558$], no main effect of attention, and no interaction. The larger Pe seen for SFIs compared to valid AV_AV trials (see figure 5.6) suggests participants may have had less confidence in the accuracy of their SFI judgments than for the valid trials. Though the Pe for AV_AV and AV_A remained statistically different in the bimodal attention condition, the narrowing in figure 5.6 may reflect the decreased perceived perceptual difference between the two, as reflected in the lower d' in that condition. In accord with the accumulating evidence hypothesis (Steinhauser & Yeung, 2010), participants may have a sense that some of their responses in the SFI trials are incorrect, but don't yet recognize them as explicit response errors.

4. Conclusions

In this chapter, we have evaluated the ERP data collected in experiment 1. We found early sensory components to illusory AV_A stimuli similar to those reported by Mishra and colleagues in an earlier study (Mishra et al., 2007). Specifically, we found a negative deflection comparable to the ND270 that indexes the multisensory processing found in the audiovisual stimuli used in these experiments. This is important as it helps to both confirm the previous findings in a different experimental setting and suggests that our additional findings may have similarly broad external validity.

Our first examination of decision-related brain activity focused on the endogenous P3 component during evaluation of valid targets (AV_AV) and illusory stimuli (AV_A). We found that the P3 was larger in amplitude to valid AV_AV targets in both task-attention conditions. We interpret this as demonstrating greater confidence that the valid targets were properly categorized in our go/no-go tasks. We find this supporting evidence for the previous findings by Rosenthal and colleagues (2009) that the illusory flash in the SFI has a perceptually distinct contrast from a veridical second flash. It is also consistent with the ERP results reported by Mishra *et al.* (2007) demonstrating that the illusory flash has a different scalp topography than a real flash.

We also found that the P3 was larger in amplitude and had a more defined peak in the unimodal task-attention condition compared to the bimodal task. There are at least three possible explanations for this attention effect across task conditions. First, the unimodal task is a simpler task that may require less cognitive resources. As part of this, the bimodal task, requires attending to both sensory channels simultaneously and may deplete attentional resources that could be applied to the stimulus categorization. A second possible explanation involves the additional processing time for completion of the more difficult stimulus evaluation made in the bimodal matching task. This could contribute to the more dispersed peaks found in the bimodal condition. Finally, we can’t rule out the possible impact of order effects. As noted in chapter 2, the unimodal task always preceded the bimodal task in our experiment. It is possible that decrements in the P3 amplitude in the latter task could be due to fatigue or habituation to the stimuli. This confound remains a possibility in all interpretations to follow.

When looking at response-locked ERP components, we first note that small ERNs of equal size were reported for both the valid AV_AV responses and the false alarms to the AV_A illusions. These ERNs were also found to be smaller in amplitude in the bimodal task. ERNs to correct responses have been reported previously (Vidal et al., 2000), and have been interpreted as indicating a ‘response evaluation process’ rather than explicit error detection. It is noteworthy that the findings by Vidal and colleagues (2000) relied on stroop-like stimuli with a fast presentation rate. It is possible that the ERNs associated with these correct responses reflect strong response conflict between automatic responses and the more considered responses required to evaluate incongruent stoop stimuli. In a similar vein, we suggest that the small ERN to both valid and false alarm trials in our tasks reflect a difficulty in determining exactly what counts as an appropriate response. As the AV_A illusion is mistaken for the AV_AV trials, response monitoring indexed by the ERN may signal equal difficulty in choosing a response for these perceptually similar stimuli.

We noted that the ERN was smaller in the bimodal task. We know from our behavioral data that the ability to distinguish valid and illusory stimuli decreased in the bimodal task. If participants interpreted the AV_AV and AV_A stimuli as more alike in the bimodal task, this may have paradoxically reduced the response conflict by encouraging a more consistent response selection. In this case, the response monitoring expressed in the

ERN may have determined that the ‘match’ response was less problematic for all stimuli that appeared to contain two flashes and two beeps.

Our examination of the later-occurring Pe revealed a larger mean amplitude for illusory AV_A stimuli when compared to valid AV_AV stimuli across both attention conditions. We interpret this as evidence that participants had less confidence in their responses to the illusory trials. Finding an effect in Pe amplitude without a corresponding difference in ERN amplitudes is very unusual. Di Gregorio *et al.* (2018) similarly reported a Pe without a preceding ERN difference. In their task, the stimuli to be evaluated were masked in order to make explicit categorization difficult or impossible. Errors had to be inferred from other flanker stimuli that were also present. In accordance with the accumulating evidence hypothesis (Steinhauser & Yeung, 2010), the authors concluded that while there was insufficient information available in the stimuli to conclude that an error had occurred, participants developed a growing sense that their responses were mistaken. This highlights the independence of the ERN from the Pe and suggests that the ERN is more determined by matching stimulus evaluation with a specific response. We speculate that an increasing amplitude in Pe for illusory stimuli similarly reflects an awareness that the responses to the illusion may be incorrect, but that participants have insufficient evidence to detect the error based on the stimuli alone.

Taken together, we find evidence that participants can detect some difference between AV_AV and illusory AV_A stimuli. This is reflected in their decreased P3 potential, showing less confidence in the categorical decisions for illusions and an increased Pe which indicates an accumulation of evidence that responses to the illusion are erroneous. However, an inability to perceptually distinguish AV_AV from AV_A trials denies participants sufficient evidence to conclude than an error has occurred. This is consistent with the findings from Rosenthal *et al.* (2009) who find that people may be able to distinguish the real from illusory trials under increased reward structures, but that they continue to report the illusion once the reward is taken away.

The differences in P3 and ERN found across attention conditions also provides evidence for the effect of endogenous task-directed attention on multisensory perceptual decisions. When attending to both sensory modalities, participants encounter greater difficulty in distinguishing real (AV_AV) and illusory (AV_A) stimuli. We speculate that this results from an increase in multisensory integration during bimodal attention. This may be due to a comparative increase in sensory gain in the auditory stimulus when is it not actively gated as in the focused attention visual task. It may alternatively result from the spread of attention in the more difficult matching task, making fewer resources available for correct categorization of stimuli. More work needs to be done to discern which of these two possibilities has better explanatory value.

Our results, coupled with the results from Mishra *et al*. (2009) reporting effects of spatial attention on ERPs associated with the SFI, provide converging evidence for the impact of selective attention on multisensory integration by selection filters. Modulating endogenous attention through the use of task directions has a measureable effect on both the early sensory components associated with perception of the SFI (Mishra et al., 2009), and later occurring decision and response potentials in trials where the SFI was detected (current experiment). These findings are also in accord with the *directed attention hypothesis* (Andersen *et al*., 2004). This hypothesis argues that crossmodal stimuli are not completely fused in an automatic fashion when presented simultaneously. Our ERP results demonstrate that endogenous attention affects multisensory integration at both the stimulus evaluation (P3) and response evaluation (ERN and Pe) stages of processing. This supports broader theories arguing for attention as a necessary and dynamic component of multisensory integration (Talsma *et al.*, 2010).

While these are interesting results for the sound-induced flash illusion, this also provides an avenue for future research in the evaluation of multisensory stimuli and other perceptual illusions. Teasing apart the stimulus-locked and response-locked components, under differing exogenous and endogenous task conditions, provides an opportunity to better understand the different stages of multisensory processing. It can also inform our understanding of how feedback and situational constraints can impact learning under conditions of perceptual uncertainty.

Chapter 5 is co-authored with Townsend, Jeanne and Westerfield, Marissa. The dissertation author was the primary author of this chapter.

Chapter 6: General Discussion

> In physical science a first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you are scarcely, in your thoughts, advanced to the stage of *science*, whatever the matter may be.
>
> - William Thomson, 1st Baron Kelvin (1883)

1. Introduction

As noted by Lord Kelvin (1883), measurement is essential to the scientific enterprise. Quantifying behavior in a consistent and repeatable fashion allows for statistical comparisons which help us gain a better understanding of observable phenomena. In the current set of experiments, we have attempted to measure the effects of endogenous and exogenous attention on multisensory integration. We have relied on reports of crossmodal illusions, primarily the sound-induced flash illusion (Shams *et al.*, 2000, 2002), to help quantify the effects of attention on multisensory integration. The underlying logic has been that factors which increase multisensory integration will tend to increase reports of the SFI, while those that reduce multisensory integration will reduce the frequency of such reports. We have also relied on measures from Signal Detection Theory (Macmillian & Creelman, 2005) to help interpret the meaning of those frequency of illusion changes. Finally, we have attempted to integrate results from electroencephalographic (EEG) data to help isolate the processing stages in the brain during which these changes may occur.

In this chapter, we will begin by briefly summarizing the principal findings from our four experiments. We then examine a few of the critical measures we have used to quantify multisensory integration and further explore how they may be interpreted. We focus primarily on psychophysical measures from Signal Detection Theory (SDT). While these measures and their use are very well-established, we will specifically look at how they have been variously employed in studies of multisensory illusions, with emphasis on the sound-induced flash illusion (SFI). Recent critiques by Witt and colleagues (2015, 2016) have found fault with how SDT measures of sensitivity (d') and bias (c) have been used to explain the perceptual nature of the SFI. We will review their argument, along with a rejoinder by Knotts and Shams (2016). Our twin goals will be to highlight differences in theoretical interpretation where they exist and promote some best practices going forward.

Following a discussion of how signal detection measures may be applied to the SFI and similar multisensory effects, we finish by suggesting future work that may help illuminate the methodological and theoretical issues raised here.

2. Summary of current findings

In experiment 1 (chapter 2), we examined the effect of endogenous modal attention on reports of the sound-induced flash illusion. Modal attention was manipulated through the use of task instructions in a go/no-go paradigm. In the first task (*focused visual attention*), participants were asked to respond to the presence of two flashes, ignoring any auditory stimuli that might be presented. In the second task (*conditional bimodal attention*), the same participants were asked to respond anytime the number of visual flashes matched the number of auditory beeps. We found that participants reported more SFIs in the novel bimodal task than in the unimodal task. This change was accompanied by a significant decrease in visual perceptual sensitivity (*d′*), with correct AV_AV trials labeled as 'hits' and the illusory AV_A conditions labeled as 'false alarms.' Total bias (*c*) remained unchanged across attention conditions.

In our analysis of experiment 1, we noted that participants reported the SFI in higher proportions than seen in earlier experiments with the illusion (e.g. Mishra *et al*., 2007; Shams *et al*, 2001; Watkins *et al*., 2006). We suggested that our experimental parameters may have resulted in a ceiling effect, such that the impact of endogenous attention would not have been large enough to strongly modulate reports of the illusion. Therefore, experiment 2 replicated the design of experiment 1, but used a near-threshold auditory stimulus in lieu of the 60 dB auditory stimulus used in experiment 1. We again found a decrease in visual perceptual sensitivity in the bimodal task associated with the SFI. Unlike the first experiment, this was accompanied by a change in bias, with responses becoming more liberal for both AV_AV and SFI stimuli in the bimodal condition. We also found a significant increase in the effect of attention on reports of the SFI. This was due to a relative *decrease* in reports of the illusion in the focused visual attention task. We speculated that the decrease was due to participants gating the near-threshold auditory stimulus, thereby reducing its effect on multisensory integration. Taken together, experiments 1 and 2 demonstrate that endogenous task-directed attention can affect multisensory integration in our experimental paradigm.

In our third experiment, we sought to isolate the effect of auditory intensity on the focused attention task. Exogenous attention was modulated through the use of two auditory intensities: 60 dB (A) SPL and near-threshold, titrated for each participant. The task was to determine the number of flashes present, ignoring the auditory stimuli. We included an AV_V stimulus in order to balance the mismatching stimuli and test for the presence of the flash fusion (Andersen *et al.*, 2004).

The first noteworthy finding was a reduction in the ability to discriminate one flash (V) from two flashes (V_V) in the louder 60 dB condition when compared to the near-threshold condition. In this case, d' and c were calculated with correct V_V responses as hits and incorrect V responses as false alarms for each intensity condition. This result was surprising as auditory stimuli were not physically present during these visual trials. No corresponding change in bias (c) was detected. We suggest that the expectation of louder sounds in the 60 dB condition caused participants to allocate additional cognitive resources to ignore the auditory modality during this visual task. This had the effect of reducing the attentional resources available when processing unimodal visual stimuli. This result is particularly noteworthy as it demonstrates a crossmodal effect on visual discriminability in response to broad task and environmental conditions in an irrelevant modality.

In experiment 3 we also found that the SFI and flash fusion illusion were reported at a higher rate in the 60 dB auditory condition, compared to the near-threshold condition. This increase was accompanied by a decrease in d' for the flash fusion and a change in response bias (c) for both the SFI and flash fusions. While the lack of a change in d' across intensity conditions for the SFI is a bit puzzling, we note that increasing auditory intensity changed visual sensitivity in a global fashion that affected visual and audiovisual stimuli similarly. This demonstrates an effect of exogenous auditory intensity on both unimodal visual sensitivity and multisensory integration during a unimodal visual task. Such intensity effects, however, may be reduced in the case of the SFI.

In our fourth experiment, we repeated the auditory intensity manipulation of experiment 3, but used our novel bimodal conditional attention task. As in experiment 3, we found that increasing auditory intensity had the effect of increasing reports of the sound-induced flash illusion in AV_A trials and flash fusions in AV_V trials. Interestingly, participants who received the near-threshold condition first reported *fewer* SFI and flash fusions overall. When these participants were later exposed to the same stimuli in the 60 dB condition, they reported more SFI and flash fusions, but this was accompanied by a change in the bias criterion c and not by a change in sensitivity (d'). We take this to suggest that participants established a perceptual set in the first intensity conditions encountered. This perceptual set established the threshold for classifying trials as 'matching' pairs of audiovisual stimuli in the experiment. When auditory intensity changed, participants retained their discriminability (i.e. d' remained the same), but their response tendency (c) changed to become more weighted toward the auditory stimuli.

In chapter 5, we reported the results of ERP data collected during experiment 1. Although we focused on later (>300 ms post-stimulus) effects, we found early sensory components broadly consistent with a study by Mishra and colleagues (2007), helping to establish the reproducibility of multisensory ERP results with the SFI paradigm. Our first novel finding was a difference in P3 amplitude between AV_AV and illusory AV_A stimuli, with the veridical AV_AV stimuli having the more positive P3 potential. We associate this with participants having greater confidence in the AV_AV stimuli as a valid target, compared to the SFI stimuli. We also found that the P3 was reduced to both AV_AV and SFI stimuli in the bimodal task attention condition compared to the unimodal visual task. This difference could be due to a reduction in confidence in target evaluation during the bimodal task and/or a reflection of greater task difficulty in the bimodal matching task. We also note that the P3 change was concurrent with a change in *d′* across conditions, while bias (*c*) remained the same.

An additional ERP finding was the lack of a difference in the error-related negativity (ERN) for the AV_AV and SFI stimuli. This response-locked potential reflects a monitoring system that is sensitive to the accuracy of a given response compared to the response required by a task (Falkenstein *et al.*, 1990; Gehring *et al*., 1990, 1993). We interpret the lack of a difference as an inability of this response conflict monitoring system to detect a difference between the appropriateness of the hit and illusory false alarm (AV_AV and AV_A, respectively) responses for this task. The ERNs in the bimodal matching task were also found to be significantly smaller that the ERNs in the unimodal visual task. We speculate that as the SFI rate increased and *d′* decreased in the bimodal matching task, this made affirmative responses to the AV_AV and AV_A both seem more appropriate given the task requirements.

While the ERN was unable to differentiate between the AV_AV and SFI responses, the error positivity (Pe; Falekenstein et al, 1991) was sensitive to differences in these responses. This potential is believed to reflect an independent error monitoring system (Falkenstein *et al*., 2000) that is more related to subjective awareness that a response error has occurred (Overbeek *et al*. 2005; Steinhauser & Yeung, 2010). Responses to the SFI were found to have a more positive amplitude than responses to valid AV_AV stimuli in our experiment. We interpret this larger Pe amplitude to mean that participants were aware that some difference existed between valid targets and false alarms in SFI stimuli; however, such differences did not rise to the level of certainty that an error had occurred. If it had, it should have been additionally reflected in the ERN. This is consistent with the *accumulating* *evidence* account of the Pe (Steinhauser & Yeung, 2010), which suggests that the potential reflects the growing certainty that an error has been committed.

Taken together, we find that our experimental results support a growing body of evidence (Talsma *et a*l, 2010) that multisensory integration can be modulated by endogenous and exogenous attentional factors. Specific to the sound-induced flash illusion, we are in agreement with Andersen *et al*. (2004, 2005) that the *information reliability hypothesis* and *directed attention hypothesis* both offer some explanatory power in accounting for the SFI. According to information reliability, the stimulus that has a history of utility and current salience should modulate perception of the SFI and flash fusion. This is consistent with both the order effects in experiment 4 and the effects of auditory stimulus intensity in experiments 3 and 4. The direct attention hypothesis suggests that task instructions can establish an endogenous attentional set and influence the degree to which crossmodal stimuli can be integrated. Experiments 1 and 2 provide strong support for such a possibility, in accord with earlier work with the SFI (Andersen *et al*., 2004; Mishra *et al*., 2009) and a related finding with visuo-tactile stimuli (Werkhoven *et al*., 2009).

In all of our experiments, we have found consistent evidence that endogenous task-directed attention and exogenous stimulus intensity can affect the rate at which the SFI, flash fusion, and related multisensory effects are reported. We have also identified differences in perceptual sensitivity (d') and bias (c) that have been associated with these reporting differences. We have stated that d' represents a more useful measure than c for expressing perceptual changes in experiencing our illusory stimuli in our experimental paradigm. However, as this interpretation of signal detection measures has come into question recently (Witt *et al.*, 2015; 2016), it is useful to revisit these measures in greater detail.

3. Signal detection measures of the sound-induced flash illusion

In a recent critique of how signal detection measures have been interpreted, Witt and colleagues (2015) point out what they consider to be a popular misconception in how measures of bias (e.g. c and β) are construed. They argue that bias is often portrayed as solely a measure of decision strategy or response preference on the part of the observer. Pointing to descriptions of bias as a response tendency (Macmillian & Creelman, 2008) or a bias towards responding a certain way (Stanislaw & Todorov), Witt *et al.* (2015) argue that this often implies that bias *necessarily* reflects an internal response strategy. They contend that this leads some users of SDT to only look for true *perceptual* effects in changes in d' (sensitivity). While Witt *et al.* (2015) agree that measures such as c can reflect a strategic *response-based* bias, they state that it may also reflect a *perceptual* bias. That is, perceptual effects may also be found in bias measures such as c and not just in d' alone. Furthermore, because d' and c are statistically independent, some perceptual differences may *only* be measured as a change in bias (c) and not cause a change in sensitivity (d').

What types of perceptual biases should cause a change in c and not in d'? Witt and colleagues (2015) argue that the effects of perceptual illusions should be broadly quantified by changes in c. Though they address a number of perceptual illusions, they specifically point to the sound-induced flash illusion as a prime example. They state, "Theoretically, the sound-induced flash illusion is an example of a perceptual bias, and therefore should present itself in the measure of c (or β), and not in d'. The number of beeps is theorized to bias perception to detect the same number of flashes, not to make perception more sensitive per se," (Witt *et al*, 2015, pg. 291).

Before going further, it is worth noting that the SFI was not explicitly theorized to bias perception. This is only one possible interpretation. The SFI was originally seen as an unexpected multisensory illusion whose perceptual effect is to make AV_A stimuli appear as AV_AV stimuli (Shams *et al.*, 2000, 2001). While an illusion may be construed as a "perceptual bias" in the common language usage, it might not fit the formal definition of a bias in signal detection terms. In SDT, a bias is something that moves the hit rate of a real target (H) and false alarm rate of a mistaken target (F) in the same direction (i.e. both should increase or decrease together when perception is biased). By contrast, sensitivity (*d′*) tells us the distance between the means of the distribution of H and F. The more difficult two stimuli are to distinguish, and thus more perceptually similar, the smaller the value of *d′*. Therefore, from the initial characterization of the SFI as mistaking AV_A for AV_AV, it is perhaps more intuitive that authors would try to quantify this in measures of *d′* rather than *c*.

Witt and colleagues (2015) further state that previous authors have used the wrong stimulus comparisons when calculating changes in *d′* and *c*. When calculating the sensitivity and bias measures for the illusion, one method is to first compare the hit rate of AV_AV and the false alarm rate of AV_A, as we have done during this project. This seems straight-forward as the SFI is mistaken for the AV_AV stimuli. Witt and colleagues take issue with the comparison of this value with unimodal visual stimuli. In this case, *d′* and *c* are calculated using the hit rate of V_V and false alarm rate of V. The comparison is based on the logic that two beeps (A_A) are added as noise in the multimodal trials, which causes the illusion to occur in AV_A trials. The unimodal visual *d′* and *c* thus provide the baseline for comparison. This is the comparison used by many authors (e.g. Rosenthal *et al.*, 2009; Watkins *et al.*, 2006) and in our current project. Witt and colleagues (2015) claim that the correct comparison would be with stimuli where only one beep is present (i.e. AV as a hit and AV_V as a false alarm). They note that when using this comparison, changes in SFI are seen in *c* and not in *d′*.

In a response to the above argument, Knotts and Shams (2016) defend the choice to compare *d′* across the 'two-beep' (AV_AV vs. AV_A) and 'no-beep' (V_V vs. V) conditions. They argue that the original purpose for using this comparison was to highlight the degree to which the sound-induced flash illusion is due to multisensory integration. Comparing the two-beep case with the one-beep (AV vs. AV_V) would not serve that specific goal. They further suggest that Witt *et al*. (2015) specifically chose the comparison with the one-beep conditions as it produces their desired change in criterion *c*, but no change in *d′*. Knotts and Shams make the additional point that *d′* is a perfectly valid measure for measuring perceptual effects in this paradigm, regardless of the potential value of *c*. Therefore, while they concede the potential value of measuring *c* in SFI experiments, they conclude that the criticisms on Witt *et al.* (2015) do not diminish their earlier analyses.

In evaluating these competing analyses, we will first point out that both approaches are valid within the terms of Signal Detection Theory. The difference appears to lie in the choice of SDT measure which each group believes best captures the effects of the SFI. While Witt *et al*. (2015) argue at length for the superiority of *c* as a measure of perceptual bias in the SFI paradigm, they themselves note on page 298, "For discrimination experiments, *d′* can be interpreted as a perceptual effect related to changes in sensitivity, but *c* can be interpreted only as a bias without the ability to distinguish between perceptual bias and response-based bias." We note that our experiments and those critiqued earlier (Rosenthal *et al*., 2009; Watkins *et al*., 2006) are all discrimination experiments. For this reason, we believe that *c* lacks the specificity required to be a reliable measure of perceptual change in these experiments. Given that *d′* reflects a more targeted measure of perceptual sensitivity, this is one piece of support for its use.

Regarding the choice of which trials to use as a comparison with the AV_AV and AV_A trials, we suggest that the decision depends on what question you wish to answer. As Knotts and Shams (2016) argue, the comparison with V_V and V trials is valid and highlights the specific role of multisensory integration in producing the illusion. Witt and colleagues (2015, 2016) prefer the comparison with AV and AV_V, which seems to emphasize the role of the second beep in producing the illusion. As this second case compares a multisensory stimulus with another multisensory stimulus, this has some theoretical appeal. However, a critically important empirical concern is that the AV and AV_V trials are, themselves, part of a *different* illusory effect. In our experiments 3 and 4 and previous work by Andersen *et al*. (2004), the AV_V has been shown to produce flash fusions. This means there is a multisensory effect that works to make AV_V trials appear as AV stimuli. This will have the effect of reducing *d′* for these stimuli. If this is used as the comparison for the SFI, it is actually comparing the strength of two opposing illusory effects. And since the effects produce opposite responses, it is to be expected that *c* will differ strongly, while *d′* may remain similar. While this may uncover an interesting result, this approach is not an effective way of evaluating the strength of the SFI alone.

While we agree with Knotts and Shams (2016) that measuring *d′* in our discrimination experiments is the more useful way to characterize perceptual changes accompanying reports of the SFI, the discussion did serve to point out some methodological concerns in the SFI literature. Therefore, when making comparisons of SDT measures in experiments constructed to manipulate or produce the sound-induced flash illusion, we suggest the following practices to avoid confusion. First, each comparison of hits and false alarm rates should be explicitly stated in terms of the experimental trials used. In our experiments, we have tried to consistently note the specific trials used to calculate *d′* and *c* in the SFI (e.g. AV_AV and AV_A) and how they were labeled for use in calculating our SDT measures (e.g. hits and false alarms, respectively). In our studies, these were compared with the unimodal trials of V_V (hits) and V (false alarms). The discussion between Witt *et al* (2015, 2016) and Knotts and Shams (2016) highlights the fact not all experiments report their comparisons clearly. Referring to comparisons as containing 'one-beep' or 'two-beeps' is somewhat vague and invites misunderstanding.

Second, we suggest that experiments report both the *d′* and *c* measures for each comparison made. Though we believe that sensitivity is the better measure in our discrimination experiments, as noted above, it may help provide a clearer picture if both measures are reported. Having such information available would also assist in evaluating the utility of *c* as a measure in these experiments. Relatedly, the formulae used in calculating each of these values should always be provided, as we have done in chapter 2. The terminology for bias measures is sometimes inconsistent across authors, so providing the calculation method can avoid confusion.

From a design perspective, we offer a few possible suggestions that may assist in avoiding interpretive differences when calculating and interpreting SDT measures. First, as pointed out by Witt and colleagues (2015, 2016), discrimination experiments do not typically allow for isolation of a perceptual bias from response bias. However, in our discrimination experiments, we have compared the SFI across attention conditions. This allows for comparison of the *d′* and *c* using the same trial types (i.e. AV_AV hits and AV_A false alarms) across conditions. This avoids the difficulty of interpreting comparisons with unimodal (V_V and V) or multimodal (AV and AV_V) conditions. Directly comparing d' values for the SFI across conditions allows a simple test of the manipulation on the ability to distinguish the illusion from the real stimulus. As this is the original significance of the illusion, having a direct evaluation of this effect seems paramount. Reporting comparisons of c across conditions are also useful, but as Witt and colleagues (2015, 2016) point out, we are unable to distinguish perceptual from response biases.

An additional methodological design improvement is the possible inclusion of brain behavior measures that are associated with d' or c. Witt and colleagues (2016) note that this may help isolate perceptual changes when evaluating SDT measures. Misha and colleagues (2009) have reported that an ERP potential associated with experiencing the SFI was correlated with d' and not with bias. In chapter 5, we additionally associated changes in P3 brain potentials with changes in d' across attention conditions. The P3 is known to reflect perceptual evaluation processes (Hillyard *et al*., 1971), so joining this measure with SDT measures may help clarify their interpretation. Together, these brain activity and SDT measures may help with characterizations of perceptual change during multisensory integration.

4. Conclusions and future directions

In the current project, we have examined the effects of endogenous and exogenous effects on the sound-induced flash illusion (Shams *et al*., 200; 2002) and related multisensory effects. We introduced a novel experimental task, the bimodal conditional attention task (experiments 1 and 2), which successfully manipulated reports of the illusion and perceptual sensitivity (d'). These effects were found to be larger when using near-threshold stimuli titrated to each participant. These results were consistent with the directed attention hypothesis (Andersen *et al*., 2004) which suggests that focusing attention on different factors (e.g. modality) may alter the degree of multisensory integration. This is also in agreement with findings suggesting that spatial attention (Mishra *et al*., 2009) may affect the processes underlying multisensory integration in the SFI.

We then manipulated auditory intensity in both a unimodal, focused attention task (experiment 3) and our bimodal attention task (experiment 4). We found that reports of the SFI increased with auditory intensity, as did the incidence of a flash fusion illusion. The degree to which perceptual sensitivity changed was found to depend on the order in which the auditory intensity level was presented (experiment 4). These findings support the information reliability hypothesis (Andersen *et al.*, 2004, 2005) which suggests that stimulus salience is another factor which can impact multisensory integration.

We additionally analyzed ERP data collected during experiment 1 and found differences in target evaluation (measured with the P3 potential) for the SFI and real stimuli, suggesting greater confidence in valid targets than the illusory stimuli. P3 amplitude was also found to be greater in the unimodal condition than the bimodal condition. We speculate that this difference could reflect lower confidence in target identification in the bimodal task and possibly index the greater difficulty of that task. The differences in P3 were accompanied by differences in the Pe (response-locked error positivity) for these stimuli, with the Pe larger for the SFI trials. However, the earlier ERNs (response-locked error-related negativities) were indistinguishable in veridical AV_AV and SFI trials, though these decreased in the bimodal condition. We interpret this to mean that participants were less confident in the illusory SFI stimuli as appropriate targets and treated false alarm responses to these stimuli as possibly erroneous. However, the power of the illusion is such that it was not explicitly recognized as an error during the initial response evaluation period indexed by the ERN. These novel findings provide an additional window into how illusory stimuli are treated within different stimulus and response evaluation processing stages.

One of the limitations in our experiments was a lack of balance in the order in which attention and intensity conditions were presented in experiments 1 – 3. As this was found to be potentially important in experiment 4, indicating that experience may establish a perceptual set for evaluating subsequent stimuli, it would be valuable to revisit this issue. It is especially interesting that experience with the illusion can impact perception of the SFI given that it has been shown it to be resistant to feedback (Rosenthal *et al*, 2009). Following the initial success in using decision and response related ERP measures in the present project to uncover difference in target and response evaluation, it also would be useful to add these measures in future projects. Adding other response options (e.g. 'respond to the mismatching stimuli') might further test how directed attention can affect multisensory integration.

Finally, our experiments have focused mainly on the SFI and flash fusion as indices of multisensory integration. While using illusory stimuli to elucidate the complex interactions underlying crossmodal processing is not uncommon (Stein and Meredith, 1993), it does represent something of an edge case. It remains to be seen how the endogenous and exogenous attention effects we uncovered might generalize to other domains. For example, it would be interesting to see if our findings could be applied to other areas of multisensory research such as learning and education. It has been found that perceptual training can be improved through the use of multisensory stimuli, even when the task later becomes unimodal

(Seitz, Kim, & Shams, 2006; Shams and Seitz, 2008). We would be curious to know how manipulations of attention and stimulus intensity might facilitate such training in applied situations.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci, 8*(10), 457-464.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), 257-262.

Alho, K., Woods, D. L., & Algazi, A. (1994). Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology, 31*(5), 469-479.

Alho, K., Woods, D. L., Algazi, A., & Naatanen, R. (1992). Intermodal selective attention. II. Effects of attentional load on processing of auditory and visual stimuli in central space. *Electroencephalogr Clin Neurophysiol*, *82*(5), 356-368.

Alkon, D. L. (1983). Learning in a marine snail. *Sci Am, 249*(1), 70-74, 76-78, 80-74.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839-843.

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research, 183*(3), 399-404.

Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: the McGurk effect revisited. *Multisensory Research, 31*(1-2), 111-144.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Brain Res Cogn Brain Res, 21*(3), 301-308.

Andersen, T. S., Tiippana, K., & Sams, M. (2005). Maximum likelihood integration of rapid flashes and beeps. *Neuroscience letters*, *380*(1-2), 155-160.

Arnold P. & Hill F. (2001) Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *Br J Psychol.* 92 Part 2: 339-355.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept Psychophys, 62*(2), 321-332.

Bhattacharya, J., Shams, L., & Shimojo, S. (2002). Sound-induced illusory flash perception: role of gamma band responses. *Neuroreport, 13*(14), 1727-1730.

Binyon, L. (1922). *Selected Poems of Laurence Binyon*. New York: Macmillan Company.

Bischoff, M., Walter, B., Blecker, C. R., Morgen, K., Vaitl, D., & Sammer, G. (2007). Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia*, *45*(3), 578-586.

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, *35*(8), 3478-3484.

Bolognini N, Frassinetti F, Serino A, & Làdavas E. (2005) "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Exp Brain Res*. 160(3), 273-82.

Brock, J., Brown, C. C., Boucher, J., & Rippon, G. (2002). The temporal binding deficit hypothesis of autism. *Dev Psychopathol, 14*(2), 209-224.

Brown, C., Gruber, T., Boucher, J., Rippon, G., & Brock, J. (2005). Gamma abnormalities during perception of illusory figures in autism. *Cortex, 41*(3), 364-376.

Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., et al. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron, 20*(2), 285-296.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R., & Dale, A. M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *Neuroreport*, *9*(16), 3735-3739.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences*, *102*(51), 18751-18756.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol, 10*(11), 649-657.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J Physiol Paris, 98*(1-3), 191-205.

Carrasco, M. (2011) Visual attention: the past 25 years. *Vision Res,* 51(13):1484-525.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*(5364), 747-749.

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *Journal of cognitive neuroscience*, *16*(6), 1000-1009.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21.

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage, 34*(4), 1443-1449.

Devrim, M., Demiralp, T., Kurt, A., & Yucesir, I. (1999). Slow cortical potential shifts modulate the sensory threshold in human visual system. *Neurosci Lett, 270*(1), 17-20.

Di Gregorio, F., Maier, M. E., & Steinhauser, M. (2018). Errors can elicit an error positivity in the absence of an error negativity: Evidence for independent systems of human error monitoring. *Neuroimage*, *172*, 427-436.

Diederich, A. & Colonius, H. (1987) Intersenosry facilitation in the motor component? *Psych Research*, 49, 23-29.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature, 381*(6577), 66.

Driver, J., & Spence, C. (1998). Crossmodal attention. *Curr Opin Neurobiol, 8*(2), 245-253.

Driver, J., & Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Curr Biol, 10*(20), R731-735.

Eimer, M., & Schroger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology, 35*(3), 313-327.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and clinical neurophysiology*, *78*(6), 447-455.

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., Brunia, C. H. M., Gaillard, A. W. K., & Kok, A. (1990). Psychophysiological brain research. *Tilburg University Press, Tilburg*, 192-195.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological psychology, 51*(2-3), 87-107.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Percept Psychophys, 63*(4), 719-725.

Fiedler, A., O'Sullivan, J. L., Schröter, H., Miller, J., & Ulrich, R. (2011). Illusory double flashes can speed up responses like physical ones: evidence from the sound-induced flash illusion. *Experimental brain research, 214*(1), 113.

Fister, J. K., Stevenson, R. A., Nidiffer, A. R., Barnett, Z. P., & Wallace, M. T. (2016). Stimulus intensity modulates multisensory temporal processing. *Neuropsychologia*.

Forster B, Cavina-Pratesi C, Aglioti SM, & Berlucchi G. (2002) Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp Brain Res.* 3(4):480-487.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Early auditory-visual interactions in human cortex during nonredundant target identification. *Brain Res Cogn Brain Res, 14*(1), 20-30.

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res Cogn Brain Res, 10*(1-2), 77-83.

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport, 16*(5), 419-423.

Frassinetti F., Bolognini N., Bottari D., Bonora A., & Làdavas E. (2005) Audiovisual integration in patients with visual deficit. *J Cogn Neurosci.* 17(9): 1442-52.

Frassinetti F, Bolognini N, & Làdavas E. (2002) Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res*. 147(3), 332-43.

Frens, M.A., Van Opstal, A.J., & Van der Willigen, R.F. (1995) Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys*, *57*(6):802-816.

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *Am J Psychol, 72*, 521-529.

Gehring, W. J., Coles, M. G., Meyer, D. E., & Donchin, E. (1990). The error-related negativity: An event-related brain potential accompanying errors. *Psychophysiology*, *27*, S34. (Abstract)

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological science*, *4*(6), 385-390.

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (2018). The error-related negativity. *Perspectives on Psychological Science*, *13*(2), 200-204.

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.

Giray M, & Ulrich R. (1993) Motor coactivation revealed by response force in divided and focused attention. *J Exp Psychol Hum Percept Perform*.19(6): 1278-91.

Goldring, J. E., Dorris, M. C., Corneil, B. D., Ballantyne, P. A., & Munoz, D. R. (1996). Combined eye-head gaze shifts to visual and auditory targets in humans. *Experimental brain research*, *111*(1), 68-78.

Gordon, B. (1973). Receptive fields in deep layers of cat superior colliculus. *J Neurophysiol*, *36*(2), 157-178.

Helfer K. S. (1997) Auditory and auditory-visual perception of clear and conversational speech. *J Speech Lang Hear Res*.;40(2): 432-43.

Hershenson, M. (1962) Reaction time as a measure of intersensory facilitation. *J Exp Psych*, 63(3), 289-293.

Hewig, J., Coles, M. G., Trippe, R. H., Hecht, H., & Miltner, W. H. (2011). Dissociation of Pe and ERN/Ne in the conscious recognition of an error. *Psychophysiology*, *48*(10), 1390-1396.

Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals: when spatial correspondence matters. *Experimental Brain Research*, *195*(2), 261-272.

Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science*, *172*(3990), 1357-1360.

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1257-1270.

Howard, I. P., & Templeton, W. B. (1966). *Human Spatial Orientation*. London, New York: Wiley.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, *160*, 106-154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. *J Neurophysiol*, *28*, 229-289.

Hughes, H.C., Nelson, M.D., & Aronchick, D.M. (1998). Spatial characteristics of visual-auditory summation in human saccades. *Vision Res*. *38*(24):3955-63.

Jiang, W., Jiang, H., & Stein, B. E. (2006). Neonatal cortical ablation disrupts multisensory development in superior colliculus. *J Neurophysiol*, *95*(3), 1380-1396.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, *14*(8), 1129-1133.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163-178.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745-1758.

Kansaku, K., Carver, B., Johnson, A., Matsuda, K., Sadato, N., & Hallett, M. (2007). The role of the human ventral premotor cortex in counting successive stimuli. *Exp Brain Res, 178*(3), 339-350.

Kansaku, K., Johnson, A., Grillon, M. L., Garraux, G., Sadato, N., & Hallett, M. (2006). Neural correlates of counting of sequential sensory and motor events in the human brain. *Neuroimage, 31*(2), 649-660.

Knotts, J. D., & Shams, L. (2016). Clarifying signal detection theoretic interpretations of the Müller–Lyer and sound-induced flash illusions. *Journal of vision*, *16*(11), 18-18.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta psychologica*, *134*(3), 372-384.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557-577.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792-795.

Làdavas E. (2008) Multisensory-based approach to the recovery of unisensory deficit. *Ann N Y Acad Sci.* 1124: 98-110.

Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp Brain Res*, *166*(3-4), 289-297.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & psychophysics*, *63*(8), 1279-1292.

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.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, *8*, 213.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn Sci, 8*(5), 204-210.

Martınez, A., Di Russo, F., Anllo-Vento, L., & Hillyard, S. A. (2001). Electrophysiological analysis of cortical mechanisms of selective attention to high and low spatial frequencies. *Clinical Neurophysiology, 112*(11), 1980-1998.

McCormick, D., & Mamassian, P. (2008). What does the illusory-flash look like?. *Vision research*, *48*(1), 63-69.

McDonald, J. J., Teder- Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*(6806), 906.

McDonald, J. J., Teder-Sälejärvi, W. A., & Ward, L. M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science*, *292*(5523), 1791-1791.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Meredith, M. A., & Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res, 365*(2), 350-354.

Meredith, M. A., & Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol*, *56*(3), 640-662.

Miller, J. (1982) Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, 14, 247-279.

Miller, J. (1986) Timecourse of coativation in bimodal divided attention. *Perception & Psychophysics*, 40, 331-343.

Miller, J. (1996). The sampling distribution of d′. *Perception & Psychophysics*, *58*(1), 65-72.

Mishra, J., Martinez, A., & Hillyard, S. A. (2008). Cortical processes underlying sound-induced flash fusion. *Brain research, 1242*, 102-115.

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J Neurosci*, *27*(15), 4120-4131.

Mishra, J., Martínez, A., & Hillyard, S. A. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of cognitive neuroscience*, *22*(8), 1714-1729.

Mottonen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Res Cogn Brain Res, 13*(3), 417-425.

Ngo M. K., & Spence C. Crossmodal facilitation of masked visual target identification. *Atten Percept Psychophys*. 72(7): 1938-47.

Odegaard, B., Wozny, D. R., & Shams, L. (2016). The effects of selective and divided attention on sensory precision and integration. *Neuroscience letters*, *614*, 24-28.

Overbeek, T. J., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of Psychophysiology*, *19*(4), 319-329.

Passamonti, C., Bertini, C., & Làdava,s E. (2009) Audio-visual stimulation improves oculomotor patterns in patients with hemianopia. *Neuropsychologia*. 47(2): 546-55.

Pavlov, I. P., & Anrep, G. V. i. i. (1927). *Conditioned reflexes; an investigation of the physiological activity of the cerebral cortex*. London: Oxford University Press: Humphrey Milford.

Polich, J. (1987). Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *68*(4), 311-320.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3-25.

Reisberg, D., McLean, J. & Goldfield, A. (1987) Easy to hear but hard to understand: a lip-reading advantage with intact auditory stimuli. In *Hearing by eye: the psychology of lip-reading* (eds B. Dodd & R. Campbell), pp. 97–113. Hillsdale, NJ: Lawrence Erlbaum Associates.

Remez, R. E. (2005) Three puzzles of multimodal speech perception. In *Audiovisual Speech* (eds E. Vatikiotis-Bateson, G. Bailly & P. Perrier), pp. 12–19. Cambridge, MA: MIT Press.

Rippon, G., Brock, J., Brown, C., & Boucher, J. (2007). Disordered connectivity in the autistic brain: challenges for the "new psychophysiology". *Int J Psychophysiol*, *63*(2), 164-172.

Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain topography*, *21*(3-4), 185-192.

Saldana, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Percept Psychophys*, *54*(3), 406-416.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., et al. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci Lett*, *127*(1), 141-145.

Scheffers, M. K., & Coles, M. G. (2000). Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 141.

Seitz A. R., Kim R., Shams L. (2006) Sound facilitates visual learning. *Curr Biol.* 16(14): 1422-7.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia, 45*(3), 561-571.

Senkowski, D., Talsma, D., Herrmann, C. S., & Woldorff, M. G. (2005). Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp Brain Res, 166*(3-4), 411-426.

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: an MEG study. *Neurosci Lett, 378*(2), 76-81.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature, 408*(6814), 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Res Cogn Brain Res, 14*(1), 147-152.

Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport, 12*(17), 3849-3852.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport, 16*(17), 1923-1927.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences, 12*(11), 411-417.

Shams L., Wozny D. R., Kim R., & Seitz A. (2011) Influences of multisensory experience on subsequent unisensory processing. *Front Psychol.* 2:264.

Shipley, T. (1964). Auditory Flutter-Driving of Visual Flicker. *Science, 145*, 1328-1330.

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition, 92*(3), B13-B23.

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

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *Neuroreport, 11*(9), 2057-2061.

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing research*, *258*(1-2), 134-142.

Sprague, J. M., & Meikle, T. H., Jr. (1965). The Role of the Superior Colliculus in Visually Guided Behavior. *Exp Neurol, 11*, 115-146.

Stanford, T. R., & Stein, B. E. (2007). Superadditivity in multisensory integration: putting the computation in context. *Neuroreport, 18*(8), 787-792.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149.

Stein, B. E., & Clamann, H. P. (1981). Control of pinna movements and sensorimotor register in cat superior colliculus. *Brain Behav Evol, 19*(3-4), 180-192.

Stein, B. E., Magalhaes-Castro, B., & Kruger, L. (1976). Relationship between visual and tactile representations in cat superior colliculus. *J Neurophysiol, 39*(2), 401-419.

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

Stein, B. E., 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.

Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *Journal of Neuroscience*, *30*(46), 15643-15653.

Störmer, V. S., McDonald, J. J., & Hillyard, S. A. (2009). Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. *Proceedings of the National Academy of Sciences*, *106*(52), 22456-22461.

Straschill, M., & Hoffmann, K. P. (1970). Activity of movement sensitive neurons of the cat's tectum opticum during spontaneous eye movements. *Exp Brain Res, 11*(3), 318-326.

Sumby, W. H. & Pollack, I. (1954) Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb Cortex, 17*(3), 679-690.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2006). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration?. *Cerebral cortex*, *17*(3), 679-690.

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.

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J Cogn Neurosci, 17*(7), 1098-1114.

Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience & Biobehavioral Reviews*, *61*, 208-224.

Teder-Salejarvi, 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. *Brain Res Cogn Brain Res*, *14*(1), 106-114.

Thomson, W. (1883). Electrical units of measurement. *Popular lectures and addresses*, *1*(73).

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognit Psychol*, *12*(1), 97-136.

Todd, J.W. (1912) *Reaction to multiple stimuli.* NewYork, New York. The Science Press.

Tunik, E., Rice, N. J., Hamilton, A., & Grafton, S. T. (2007). Beyond grasping: representation of action in human anterior intraparietal sulcus. *Neuroimage, 36 Suppl 2*, T77-86.

Van der Burg, E., Olivers, C. N., 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.

Van der Stoep, N., Van der Stigchel, S., & Nijboer, T. C. W. (2015). Exogenous spatial attention decreases audiovisual integration. *Attention, perception, & psychophysics*, *77*(2), 464-482.

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the ‘error negativity’specific to errors?. *Biological psychology*, *51*(2-3), 109-128.

Violentyev, A., Shimojo, S., & Shams, L. (2005). Touch-induced visual illusion. *Neuroreport*, *16*(10), 1107-1110.

der Malsburg, C. (1995). Binding in models of perception and brain function. *Curr Opin Neurobiol*, *5*(4), 520-526.

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.

Vroomen J. & de Gelder B. (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J Exp Psychol Hum Percept Perform*. 26(5):1583-90.

Wallace, M. T., Carriere, B. N., Perrault, T. J., Jr., Vaughan, J. W., & Stein, B. E. (2006). The development of cortical multisensory integration. *J Neurosci, 26*(46), 11844-11849.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Exp Brain Res, 91*(3), 484-488.

Wallace, M. T., & Stein, B. E. (2000). Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences. *J Neurophysiol, 83*(6), 3578-3582.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent Negative Variation: An Electric Sign of Sensorimotor Association and Expectancy in the Human Brain. *Nature, 203*, 380-384.

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *Neuroimage, 37*(2), 572-578.

Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage, 31*(3), 1247-1256.

Werkhoven, P. J., van Erp, J. B., & Philippi, T. G. (2009). Counting visual and tactile events: The effect of attention on multisensory integration. *Attention, Perception, & Psychophysics*, *71*(8), 1854-1861.

Wessel, J. R. (2012). Error awareness and the error-related negativity: evaluating the first decade of evidence. *Frontiers in human neuroscience*, *6*, 88.

Wijers, A. A., & Boksem, M. A. (2005). Selective attention and error processing in an illusory conjunction task. *Journal of Psychophysiology*, *19*(3), 216-231.

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, *44*(3), 289-300.

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2016). Further clarifying signal detection theoretic interpretations of the Müller–Lyer and sound-induced flash illusions. *Journal of vision*, *16*(11), 19-19.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological review*, *111*(4), 931.

Zampini M, Torresan D, Spence C, & Murray MM.(2007) Auditory-somatosensory multisensory interactions in front and rear space. *Neuropsychologia*. 45(8): 1869-1877.