The Effects of Auditory Bandwidth and Spatial Congruence on Early Audiovisual Interactions

---

A Thesis

Presented to

The Division of Philosophy, Religion, Psychology, and Linguistics

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Eli Coston

May, 2012

Approved for the Division

(Psychology)

Michael Pitts

Acknowledgments

I would like to thank my parents for their love and support, and for instilling in me a curiosity about the world. I would like to thank my brother for being my best friend. I would like to thank Dana for being the best companion there ever was. I would like to thank my friends for keeping me lively and sane. I would like to thank all my professors who truly cared about the learning process, and Michael Pitts for being such a positive force throughout my senior year. I would like to thank Josie for helping me collect data. Lastly, I would like to thank all of my participants for their time and effort

Table of Contents

List of Figures

Abstract

Research suggests that audiovisual (AV) stimuli can result in cross-modal interactions during early visual processing stages. But only a couple of investigations have begun to uncover the systematic principles underlying these interactions. Therefore, this study attempts to determine whether these interactions are sensitive to the spectral characteristics of auditory stimuli, or the spatial congruence of the (AV) stimulus. And lastly, this study investigated whether these early AV interactions occur when the AV stimulus is not task-relevant. Therefore, participants performed a task on the fixation cross, while auditory, visual, or audiovisual stimuli were presented. Meanwhile Event-related potentials (ERPs) were recorded. The auditory and visual components of the AV stimulus could be spatially congruent or spatially incongruent. Moreover, the auditory component of the AV stimulus was either a white noise, or a pure tone. ERP's induced by auditory and visual stimuli alone were added together, creating a 'sum' waveform. the 'sum' waveform was compared to the ERP induced by AV stimuli, referred to as the 'simultaneous' waveform. AV interactions were quantified by a significant amplitude difference between these two ERP waveforms. The spatially congruent white noise produced a significant cross-modal interaction over occipital scalp between 84 and 104 ms, which coincided with the visual C1 component. Moreover, the spatially congruent pure tone produced a significant cross-modal interaction over parieto-occipital scalp between 98 and 114 ms. This interaction coincided with the early part of the visual P1 component. Spatially incongruent AV stimuli did not produce any early cross modal interactions. The results suggest that pure tones and white noises interact differently with

early visual processing. Moreover, spatially incongruent AV stimuli may not produce interactions within the early stages visual processing. Finally, while task-relevance was not manipulated experimentally, these results suggest that task-irrelevant stimuli may interact within the early stages of visual processing.

Introduction

For many years scientists have been perplexed and intrigued by the interaction, interference, and integration between the auditory and visual systems. The well-known ventriloquist effect, in which a visual cue displaces the perceived location of an accompanying auditory cue, exemplifies the illusory potential of multisensory integration. The more germane aspects of multisensory integration, however, help to create a unified experience of the world. While observing a chirping bird, one does not see the color of the feathers, and then hear the sound, only to switch back to the feathers. Rather, the observer automatically integrates the visual and auditory signals to produce a seamless perception of a chirping-colored-bird. Different avenues of multisensory research collectively suggest that the brain actively integrates auditory and visual information at almost all levels of neural processing.

Classically, the human parietal cortex has been implicated in the integration of audio and visual cues [Molholm et al., 2006; see Calvert, Spence, And Stein, 2004 for a comprehensive review]. In the past, the parietal cortex was even referred to as 'the association cortex' because it receives and integrates sensory input from many different domains (Andersen et al., 1997). The advent of brain imaging verified and extended this role for the parietal cortex. fMRI data suggests that accurate semantic categorization of audiovisual objects correlates with increased Blood Oxygenation Level Dependence (BOLD) signal in the Intraparietal Sulcus (IPS) (Werner and Noppeney, 2010). Placing an object into an appropriate semantic category may require that low-level and high-level brain regions fire in concert. Through feedforward, feedback, and recurrent signaling of

this sort, past experience organizes the many parallel streams of sensory input, while these many streams constantly update larger-scale object representations. Many brain processes occur in this way. In the IPS, for example, visual and auditory streams may mutually enhance the full-fledged object representation. Recall the chirping bird described earlier. Imagine that only the bird's feet are visible. Alone, the feet provide scarcely enough information to recognize the bird as a bird. They might just be perceived as part of the branch on which the bird is perched. But when the bird chirps, the brain may update its semantic evaluation of these "branches," now representing them more accurately as the feet of a bird. The IPS may be the direct source of this semantic relay across modalities. Or, the IPS may be using semantic information from one sensory modality to direct attention in another sensory modality (Anderson et al. 2010). Regardless, this increased BOLD activity reflects the semantic integration of audiovisual cues, indicating that the separate visual and auditory processing streams have already passed through low-level anatomical and perceptual stages.

However, the very same cross-modal mechanisms can sometimes mal-adaptively alter one's representation of the environment. For example, when participants viewed a mouth saying ba, while listening to the syllable ga, they reported hearing the syllable da. (McGurk and MacDonald, 1976). This seminal study of cross-modal perception was extended by a more recent study, in which the syllable 'ba' was perceived as 'va' because of the misleading visual cue (Saint-Amour et al., 2007). On most of the trials the auditory ba was presented simultaneously with a congruent video of lips, which articulated the same syllable, 'ba.' On a few of the trials, however, a video of lips articulating the incongruent syllable, 'va' was presented. When the incongruent video accompanied the

sound, participants reported hearing va despite the physical characteristics of the sound, which would normally produce the percept ‘ba’. Interestingly, although the participants were never presented deviant auditory stimuli, the researchers recorded event-related potentials (ERPs) and observed a prominent Mismatch Negativity (MMN), which is a neural event elicited by an infrequent, or ‘deviant’ auditory stimulus. In this case, the visual cue led to the experience of a deviant auditory stimulus by cross-modally influencing auditory perception, thereby producing the MMN. Importantly, these neural events occurred anywhere from 175ms to 400ms after the onset of the auditory stimulus. At this point in time, feed-forward neural activity most likely had already swept through many specialized visual areas (Foxe and Simpson, 2002), and those areas must have begun sending feedback signals to basic visual areas (V1, V2). Multisensory integration occurring at later time windows gives us important insight into our cognitive manipulations of cross-modal information. But earlier, more unisensory processing stages are equally important, since they determine aspects of these later-stage, cognitive manipulations.

If an auditory beep is presented immediately before and after a single visual flash, the participants often report seeing two flashes. This audio-visual illusion called the Shams illusion differs from the McGurk effect in two ways. Firstly, in the Shams illusion, auditory information influences visual perception, not the other way around. But more importantly, this illusion was correlated with a much earlier modulation of visual cortex compared to the McGurk effect. It occurred between 80 and 100 ms after the presentation of the visual stimulus (Mishra et al., 2007), as opposed to 175 and 400ms. Later processing stages draw on information from earlier stages. Therefore, cognitive

mechanisms, like those manipulated by the McGurk effect, may be influenced by interactions occurring at earlier stages, such as the 80 – 100 ms window that the Shams illusion seems to occur within. As we will learn from animal studies, interactions may occur in even lower brain structures, and earlier processing windows, than the abovementioned ones. Of specific interest are the mesencephalic and diencephalic structures that integrate basic sensory features across different modalities.

Single cell studies of cats and primates suggest that the thalamus and the colliculi (structures within the diencephalon and mesencephalon, respectively) are key sites for multisensory integration (Stein and Meredith, 1993). Subsets of cells were identified that respond to unimodal, bimodal, and trimodal stimulation. Because we are currently unable to measure human thalamic activity with the required temporal precision, we can only postulate an analogous function of the thalamus in humans. Nevertheless, this finding opens the possibility for much earlier and more basic multisensory effects in the human brain. Moreover, cross-modal information at such low levels could be embedded in feedforward projections to higher brain areas.

In the last decade, some research has suggested that integration of auditory and visual signals may occur as early as 50ms post-stimulus. This evidence is only made available by the ERP technique, because it is ideally suited to investigate the unfolding of neural events with high temporal precision. Very generally, a single neuron becomes active when the ions flow either in or out of the cell membrane. This chemical event creates a transient voltage difference orthogonal to the cell membrane. These electric potentials at the cell body are on the order of microvolts. Even with sensitive electrodes covering the entire scalp, the activity of a single neuron cannot be detected. However,

when a group of similarly oriented cells fire simultaneously, they produce a large enough electric dipole to propagate through the layers of brain tissue and the skull. If the same stimulus is presented many times under the same experimental conditions, the pattern of neural activity associated with the stimulus event will be relatively similar each time, while spontaneous neural activity in other brain regions (noise) will vary randomly across trials. Because of this consistency, we can time lock the electrical measurements (waveforms) to the onset of the stimulus, and average many trials together. This averaging process attenuates the random noise, producing a smoother waveform that is directly related to the processing of the stimulus. Because this analog electrical signal is converted into a digital value every 2ms, one can observe extremely transient increases and decreases of electrical activity.

Researchers have used these techniques to identify audiovisual interactions occurring in timeframes that are traditionally reserved for unisensory processing only (Cappe et al. 2010; Giard and Perronet 1999; Fort et al. 2002; Molholm et al. 2002). These researchers all measured ERPs while presenting stimuli of three different types: auditory alone (A), visual alone (V), and simultaneous audio-visual (AV). Commonly, the ERP waveforms from condition A and condition V are summed together, creating an (A + V) waveform. Then this waveform is subtracted from the simultaneous AV waveform, producing the general equation [AV – (A + V)]. The difference between these two waveforms is proposed to reflect the difference in neural activity between simultaneous presentation of an auditory and a visual stimulus versus independent presentation of the exact same auditory and visual stimuli. The relative validity of this assumption will be discussed in more detail below.

Traditionally, sensory information is thought to flow through Primary Visual (V1) and Primary Auditory (A1) cortex before projecting to higher-level areas, such as the Intraparietal Sulcus (IPS), where it could potentially be integrated. The extremely short latency at which audio-visual interactions have recently been found (45 ms) challenges the hypothesis that auditory and visual processing streams are initially segregated. While a potential revision to the traditional view is exciting, it is critical that we are able to reproduce the effect, and address any methodological concerns. If this early multisensory integration effect is replicable and becomes more established, two possibilities arise: low-level, unisensory brain regions may be communicating rapidly and directly. Or, visual and auditory signals may be represented in high-level brain regions at earlier latencies than previously considered.

Voltage values from each electrode can be configured into a topographic map of the scalp (example: figure 6). Using a map of this sort, researchers can make very rough estimations of which brain area is producing the observed voltage pattern. Within this vein of research some scalp topographies have been observed consistently, but others have not. While discriminating between two bimodally defined stimuli in an object recognition task, participants exhibited a significant amplitude difference between AV and (A+V) at electrodes O2 and PO4 within the time window of 45-90ms (Giard and Perronet, 1999). This time window is traditionally reserved for the early visual ERP waveform known as the C1. However, a visual stimulus must be sufficiently illuminated or contrasted in order to produce a C1 component at all. The visual stimulus was not intense enough to produce a noticeable C1 component. Therefore, the authors suggest that although the latency and topography of the AV – (A+V) difference wave mimics that

of the C1, it most likely derives from a different neural generator. If there is no C1 component to begin with, it seems unlikely that the interaction modulated the same regions that produce the C1. In the 90 – 110ms time range, the difference wave migrated leftwards, now spanning the sagittal midline of the occipital cortex. Its peak amplitude was observed at left (01), right (02), and central (Pz) electrodes. The spatiotemporal similarity between this amplitude enhancement and the well-documented C1, was challenged by another study. A detection paradigm, in which the participant responds as quickly as possible to the presence of any stimulus, revealed an early (60 ms) difference in mean amplitude over occipito-parietal scalp (Senkowski et al., 2011) just as the previously reported study did. But this time the 'pair-sum' amplitude difference was significant over the left visual hemisphere, instead of the right. In both studies, however, the visual stimulus was centrally presented. Within this early time window, the voltage topography of the C1 component is highly sensitive to the location of the stimulus in the visual field (Di Russo et al., 2001). Therefore, this topographic difference across studies initially implies different neural generators. But, individuals also vary considerably in the conformation of their primary visual cortex (V1), which determines the voltage topography of the C1. In sum, slight topographic differences between studies could be due to small sample sizes, or due to truly different neural sources between studies.

These studies suggest an entirely new temporal domain in which cross-modal interactions may occur. And while they nod to a dissociable neural circuit for early multisensory interactions, they fail to quantify this proposed circuit. In other words, they do not analytically distinguish between modulations in amplitude versus modulations in response topography. While one group of electrodes exhibits an amplitude increase, an

adjacent group might exhibit a concomitant decrease. If only the first electrode group is reported upon, the results suggest a modulation of amplitude. But a wider array of electrodes would suggest a topographic shift in amplitude. In an audiovisual detection study, a technique called Global Dissimilarity (DISS), which is a calculation of the root mean square difference between two strength-normalized vectors (or two electrode voltages), was used to distinguish between amplitude and topographic modulations (Cappe et al., 2010). DISS is essentially a reflection of the voltage differences between any two electrodes within the same ERP. In other words, they were asking if the relationships between electrode voltages remained constant for 'simultaneous' and 'sum' ERPs. They found that the voltage relationships between electrodes differed significantly between 'simultaneous' and 'sum' condition. They propose that the 60ms difference wave over parieto-occipital scalp was due to a modulation in topography as well as response gain. This suggests that topographic modulations may contribute to, or even account for the amplitude differences between 'simultaneous' and 'sum' ERPs reported in similar studies.

Source analysis techniques estimate the most likely anatomical source from which a particular voltage distribution would arise. While this technique is helpful, it has inherent limitations. Specifically, a set of neural generators will produce the exact same voltage topography every time they are active. In contrast, a given voltage topography could arise from multiple neural generators. This is referred to as the inverse problem. Nevertheless, estimates are often relatively accurate. With source analysis techniques, Cappe et al. (2010) attributed these interactions to a number of areas, including V1, the Superior Temporal Sulcus (STS), and the Superior Temporal Gyrus (STG), which

contains the Primary Auditory Cortex (A1). However, they suggest that the primary generator may be STS, and more generally the occipito-parietal junction.

Thus far, the data suggest that somewhere in the visual cortex the amplitude difference between 'pair' and 'sum' waves begins at 45ms, probably over parieto-occipital and occipital scalp (Giard and Perronet, 1999). Moreover, visual cortex is initially active by approximately 50ms post-stimulus (Di Russo et al., 2003). In theory, this implies that the very first cortical visual activity interacts with auditory signals. The first possibility to consider is that auditory information arrives at V1 by 45 ms. However, the spatial resolution of ERP data prevents us from validating this prediction. With a more spatially precise technique referred to as anatomical tracing, a specific area of cortex is injected with dye, which then stains the length of the axon, revealing the layout and prominence of axonal connections between brain areas. With this technique, axonal projections to V1 were revealed in macaque primates (Rockland and Ojima, 2003; Cappe and Barone, 2005; Clauvingier et al., 2004; Falchier et al., 2002). Some afferents projections were found originating in area A1, suggesting that at least in the macaque, primary visual and primary auditory cortex share direct cortical connections. This may or may not translate to human anatomy. However, the first cortical auditory component likely begins at 20ms in A1 (Hillyard, et al., 1998). If the projections between A1 and V1 do exist in humans, cortical auditory activation may occur early enough to arrive in V1 on time for the first visual activation.

It is important to keep in mind, however, that the earliest cortical visual processing does not necessarily occur in anatomical area V1. Extrastriate areas such as V2 lie just anterior to V1. V2 is also among the earliest visual areas to be activated, and

may contribute considerably to the C1 component. Surprisingly, when a visual stimulus moves at a certain speed, its neural signal may arrive at motion specific areas such as V5 before arriving at primary areas such as V1 (Ffytch et al., 1995). This may occur from a thalamic or midbrain projection that bypasses V1. It most likely travels via the Magnocellular visual pathway, which usually processes motion. And while V5 primarily processes visual motion, it also exhibits a negative response modulation to auditory stimuli (Beauchamp, 2005). Although none of the abovementioned studies include visual motion, this is just one example of the short latency at which higher visual areas can be activated, and the variety of anatomical regions (including visual ones) that intercept information from multiple sensory modalities. Another example is the Inferior Parietal Sulcus (IPS). Amazingly, the IPS is one of two human brain areas, which have been intra-cranially probed for multisensory interactions (Molholm et al., 2006). Single neurons in this area responded to visual stimuli at 75 ms and auditory stimuli at 30 ms. This emphasizes the overall temporal difference between auditory and visual processing. Cortical visual processing does not begin until 45 ms. In contrast, this human single cell study shows that auditory signal processing has already traveled to multisensory areas by 30 ms. So, higher up visual areas may be activated as early as V1, and auditory information extends into dorsal visual areas at very early latencies. Together, these facts begin to provide new explanatory routes, and they alleviate responsibility from the notion that auditory information must be accessing striate cortex (V1) at the earliest visual processing stages.

Other evidence suggests that the C1 component actually derives from multiple anatomical generators, including striate and extrastriate cortex (Foxe and Simpson,

2001). As mentioned earlier, these early AV interactions may arise from "distinct early multisensory circuits" that deviate from the C1 component. However, as the anatomical limits of the C1 component expand, they may engulf these early AV interactions. In order to fit these cross-modal effects into an early time frame, we may need to invoke the possibility that higher anatomical areas integrate multisensory signals at earlier latencies. Moreover, systematically varying the stimulus set and task conditions will guide our conception of these effects, and their evolutionary or functional value.

Notably, some studies have shown no early interaction effects (Gondan et al., 2005). Recent research suggests that this failure to produce early interactions is due to the physical intensity of the stimuli used (Senkowski et al., 2010). By measuring ERP's during a detection task, they showed that physically weaker auditory and visual stimuli integrate to a greater extent than more intense or salient auditory and visual stimuli of the same type (Senkowski et al., 2011). This phenomenon is commonly referred to as "inverse effectiveness." From an information-processing standpoint, the brain wants as much information as possible about a pertinent sensory event. If two related sensory events have a low signal-to-noise ratio, they both benefit from sharing information with one another. If one sensory modality has more physical information about an event than another modality, it may inform the impoverished modality. This may also apply to perceptual strength as opposed to physical stimulus strength. In the first study of early multisensory interactions, subsequent analysis was performed according to the innate auditory or visual tendencies of each subject (Giard and Perronet, 1999). Those subjects who were better at visual object recognition exhibited slightly smaller multisensory amplitude enhancement over early visual scalp. The opposite was true for subjects who

performed better at auditory object recognition than visual object recognition, suggesting that the less capable modality benefited more from multisensory integration.

Despite the growing numbers of studies that have replicated this inverse effect (lower stimulus strength = greater multisensory integration), none have investigated the effects of stimulus *type* on early multisensory integration. The current investigation seeks to shed light on the underlying structures of early multisensory interactions by varying the spectral characteristics of the auditory stimulus (i.e. bandwidth). The cortical and subcortical representations of pure tones and white noises differ in many ways. The cochlea is arranged according to frequency (tonotopically) (Wolfe, 2010). Therefore, at the first step of neural processing, the sensory epithelia represent white noise with a greater overall distribution. This relationship is also reflected in A1 (Upadhyay, 2007), as well as higher areas of human cortex. An fMRI study in humans suggests that a single frequency tone induced less activation in primary (A1) and non-primary auditory fields than a complex harmonic tone (Hall et al., 1991). Similarly, in the antero-, middle-, and caudal-belt areas of the rhesus monkey, neurons fire heavily to a band passed noise. But in response to a single frequency pure tone they fire at a fraction of that rate (Rauschecker and Tian, 2000). The neural discrepancies between pure tones and white noise predict a reduced 'effectiveness' for pure tones relative to white noise. Moreover, this

Recent research suggests a possible difference in the latency of neural activation between high bandwidth, and low bandwidth sounds. For instance, in the cat auditory cortex, a complex acoustical noise burst produced earlier activation in many auditory regions (A1, AAF, PAF, A2) than a pure tone of the same dB level (65dB) (Carrasco and

Lomber, 2011). This may be due to the tonotopic organization of A1, and the lateral fibers connecting frequency selective bands (Upadhyay et al., 2007). If different tonotopic regions activate one another, it is possible that the relationship between bandwidth (or activated regions) and total activation in A1 is non-monotonic. As in, if intensity remains constant, we may see more total activation in A1 in response to high bandwidth as opposed to narrow bandwidth sounds (Rohl et al., 2011). If we consider these findings in conjunction with the discovery of anatomical connections between auditory cortex and V1 in non-human primates, (Rockland and Ojima, 2003; Cappe and Barrone, 2005; Falchier et al., 2002; Clavagnier et al., 2004) it seems likely that a high bandwidth sound may produce a greater amplitude integration effect, or simply an earlier integration effect, over visual scalp. So far, none of the studies investigating early multisensory interaction have manipulated bandwidth. Moreover, all of the early multisensory interaction studies to date have used pure tones instead of complex tones or noise. One of the primary goals of this study was to investigate differences in the latency, amplitude, or topography of early multisensory interactions due to variations in auditory spectral bandwidth.

Bandwidth also plays a well-documented role in our ability to localize sounds (Brungart and Simpson, 2009). In one study, the localization of low-pass filtered noise (0.2 – 3 kHz) was significantly worse than that of high-pass filtered noise (3 – 15 kHz), which in turn was significantly worse than broadband noise (0.2 – 15kHz). (Brungart, 1999). Another study showed a similar effect but for both directions. In other words, localization got worse as the researchers lowered the cutoff of a low pass-filter, and as they raised the cutoff of a high-pass filter, suggesting that in general, a greater bandwidth

allows for more accurate sound source localization (King and Oldefield, 1997). This difference in localization has considerable implications for cross-modal interactions. Across many levels of processing, spatial congruence plays a major role, if not *the* major role, in the interactive potential of a multimodal stimulus pair. The tracing studies mentioned earlier reported that some form of spatial congruence exists between the auditory regions (A1 and the Superior Temporal Plane) and area V1 (Rockland and Ojima, 2003). Bimodal neurons in the superior colliculus of anaesthetized cats respond more vigorously to a bimodal event when both stimulus components are spatially congruent, than when they are spatially incongruent. (Meredith and Stein, 1996). This held true for 88% of the bimodal neurons targeted in the superior colliculus. Moreover, when the two stimulus components were presented to disparate spatial locations, the same neurons showed either no multisensory enhancement, or a decrease in response compared to unimodal conditions. This study provides reasonable evidence that in the cat brain, coding of multisensory spatial congruence occurs before signals reach the cortex. This cannot be directly applied to humans, but it begins to provoke the hypothesis that spatial congruence may be a prerequisite for multisensory integration. After all, spatial attention seems to control the processing of other visual dimensions. After all, an organism generally must direct its focus to the location of an object in order to process any of its other features.

Importantly though, neuroscientific principles such as the abovementioned one are often overturned, or at least challenged by outliers. This is exemplified in a study that shows how color attention may precede spatial attention in special instances (Zhang and Luck, 2009). Just as in visual attention, spatiality may not be the limiting factor of

multisensory integration. Even in the midbrain, multisensory receptive fields coordinate flexibly across modalities. Shifts in ocular orientation induce corresponding shifts in the multisensory receptive fields of neurons in the superior colliculus (SC) of cats. This makes sense, as one of the primary functions of multisensory SC neurons may be visually orienting to cross-modal objects. A similar flexibility may exist in higher up areas. For example, the Intraparietal Sulcus (IPS) likely contains multi-modal spatial maps (Andersen et al. 1997). And these maps can adjust their receptive field properties in response to shifts in body orientation, allowing the organism to spatially coordinate motor responses. It seems that receptive field size and flexibility depends on the specific function of the area in question. If this is the case, knowledge of the receptive field properties of early multisensory interactions may be crucial to understanding their specific role in the integration of the senses. Therefore, the other primary goal of this study was to investigate whether the congruence of cross-modal stimulus location would influence the latency, amplitude, or topography of early multisensory interactions. A spatially congruent audiovisual event should result in greater multisensory integration than a spatially incongruent audiovisual event. Moreover, if we see response suppression for spatially incongruent stimulus pairs, we may be able to draw functional parallels to other anatomical areas. And most ambitiously, our differential ability to localize pure tones and white noises may interact with the flexibility of the multisensory receptive fields.

The manipulation of attention and task seems to be critically important for the latency, intensity, and directionality of many multisensory interactions (Talsma et al., 2010). In a passive viewing study, participants were asked to pay attention to visual,

auditory, and audiovisual stimuli, but were not asked to respond to the stimuli in any way. Yet, the researchers still observed significant increases in the mean amplitude of 'simultaneous' ERPs relative to 'sum' ERPs (Vidal et al., 2008). This difference occurred over the frontal and central scalp between 55 and 95ms. Investigations of early audio-visual interactions have primarily required participants to attend and respond only to the stimuli in question, (stimuli in question refers to the auditory alone, visual alone, or audiovisual stimuli). One study, however, attempted to determine whether attention to the AV stimulus is required for interactions to occur. They asked participants to pay attention to the visual domain, the auditory domain, or both domains, on different trials (Talsma et al., 2007). They only found audiovisual interactions over the central scalp when both domains were attended simultaneously. In contrast, participants in the present study were asked to perform a distracting task for the duration of the experiment. Directing their attention away from the primary stimuli allowed for an investigation of whether task irrelevant stimuli would interact across modalities during early processing stages. Talsma et al. (2006) showed that attending and responding to the auditory modality allowed greater Steady State Visual Evoked Potentials (SSVEP's) to a random letter stream than attending to a separate visual stimulus, or attending to the visual and auditory modalities simultaneously. This has important implications for the distribution of top-down attention across modalities, and will come to the fore later on. But on its own, it does not indicate whether task relevance, or voluntary attention necessarily results in greater or more cross-modal interactions.

However, it is critically important to determine whether early cross-modal interactions change according to task-relevance. Multisensory interactions and attention

share many qualities. For instance, some stimuli grab attention because of their salience. Correspondingly, some multisensory pairs integrate because of their temporal or spatial congruence, regardless of top-down attention. From a top-down perspective, humans can attend to a weak stimulus in an array of salient ones. Correspondingly, we can integrate signals from different modalities even if they originate from different sources. These similarities can obfuscate the distinction between the two mechanisms. Effects of attention are easily misattributed to multisensory interactions, and vice-versa. For instance, the temporal or spatial congruence of two stimuli may enhance their neural representation, an example of stimulus-driven response gain. This cross-modal stimulus configuration might result in an involuntary attentional shift towards that stimulus (McDonald et al., 2001). The stimulus congruence, and the resulting attention shift, may both enhance the neural representation of a stimulus. However, just because congruence and attention similarly enhance the representation of the same object, this does not permit the assimilation of these separate mechanisms. This line of reasoning should highlight the importance of distinguishing attention from multisensory interactions. Therefore, in this study, attention was directed toward a distracting visual task. Without experimentally manipulating attention, this study aims to further determine whether attention must be focused directly towards the audiovisual stimuli in order for early cross-modal interactions to arise.

In addition to using a limited stimulus set, the tasks used by many researchers in this field may introduce noise into the actual recording of the EEG. In simple detection tasks, the participant responds approximately every second and a half. This can lead to fatigue, boredom, and decreased alertness. The task used in the current study was

designed to engage the participants without drawing attention to one aspect of the stimulus over another, thus preventing unwanted top-down attention effects.

The goal of the present research was to determine the differential effects of bandwidth and spatial congruence on the early multisensory integration of audiovisual stimuli. In addition, this study aims to determine whether task-irrelevant audiovisual stimuli will produce early multisensory interactions. Using Electroencephalogaphic recordings (EEG) we measured the neural responses to unimodal (auditory or visual) and bimodal (simultaneous audio-visual) stimuli. The visual stimulus was a high contrast sinusoidal grating, while the auditory stimuli varied in bandwidth and location. Specifically, a high bandwidth auditory stimulus compared to a pure tone stimulus (when paired with an identical visual stimulus) was expected to produce a greater mean amplitude and/or earlier latency within the C1 component. Moreover, we predict that spatially congruent, compared to incongruent, audiovisual events will enhance the mean amplitude of simultaneous ERPs compared to sum ERPs. The subject performed the loudness matching task after the electrode cap had been secured to the head, in order to reduce acoustic differences between the loudness matching task and the primary ERP experiment.

Methods

Participants

Participants were students of Reed College, who responded to advertisements posted around the College campus in Portland, Oregon. Fifteen students (ages 18-25) with normal or corrected to normal vision, and normal hearing participated in the study. All subjects reported no history of neurological damage, or neurological disorders. Each subject received lottery tickets, which could result in cash prizes of various amounts. The Human Subject Research Committee established within the institution approved all of the experimental procedures. All subjects consented to the procedures listed below.

Stimuli and Procedure

Participants were presented with nine different types of stimuli. Each stimulus was either unimodal (the single visual stimulus or one of four auditory stimuli) or bimodal (the visual stimulus *paired with* one of the four auditory stimuli) (figure 1).

Visual Stimulus

The visual alone stimulus (V_a) was a Gabor patch (a sinusoidal grating filtered with a Gaussian envelope). The stripes on the Gabor patch were either oriented vertically at 90 degrees or horizontally at 180 degrees. The centers of these patches were located approximately four degrees below visual fixation, and the patch itself was a circle with a

diameter of approximately 5 cm. The patch had a visual contrast level of 100%, meaning that the alternating stripes were completely black and completely white at their centers. Each visual stimulus was presented for 500ms, followed by a variable inter-stimulus interval ranging from 500 to 700ms. The Gabor orientation varied randomly to eliminate the possible confound of sensory adaptation to a particular orientation. Because orientation was not a variable of interest within this study, we collapsed the data across the differently oriented Gabor patches for all subsequent analyses.

Auditory Stimuli

The auditory stimulus varied in terms of spectral characteristics and location. It was either a broadband white noise (A_{bb}), or a pure 1000hz tone (A_{pt}). In addition, it was presented from either a single centrally located speaker (A_c), or from two bilateral speakers (A_l). The central speaker was located directly below the computer screen, in an attempt to make it spatially congruent with the gabor patch. The bilateral speakers deviated a considerable angle from fixation, along the horizontal azimuth. There were four possible combinations of these two dimensions: central broadband noise (A_{cbb}), bilateral broadband noise (A_{lbb}), centrally presented pure tone (A_{cpt}), and bilaterally presented pure tone (A_{lpt}). Each sound was presented for 500ms, followed by a variable inter-stimulus interval ranging from 500 to 700ms. The sounds had the following Sound Pressure Level: central broadband noise = 44 dB, lateral broadband noise = 43 dB, central pure tone = 43 dB, lateral pure tone = 40 dB. The decibel levels of these four sounds varied slightly for each participant, because of a loudness matching pre-test administered prior to EEG recording.

All together there were nine possible stimuli: visual alone (V), bilaterally presented broadband noise alone (A_{lbb}), bilaterally presented pure tone alone (A_{lpt}), centrally presented broadband noise alone (A_{cbb}), centrally presented pure tone alone (A_{cpt}), visual + bilaterally presented broadband noise (VA_{lbb}), visual + bilaterally presented pure tone (VA_{lpt}), visual + centrally presented auditory broadband (VA_{cbb}), visual + centrally presented auditory pure tone (VA_{cpt}).

Loudness Matching

Prior to the experiment, the participant performed a loudness matching task, in which every auditory stimulus was matched to every other auditory stimulus. On each trial, two sounds were presented for 500 ms, separated by a 500 ms inter-stimulus interval (ISI). We chose a relatively short ISI because a longer inter-stimulus interval introduces a confounding memory component which decreases accuracy (Yost, Popper, and Fay, 1993). Each stimulus was attenuated by a value of .05 within the stimulus presentation program (Presentation, Neurobehavioral Systems, Albany, CA). If the participant reported that the second stimulus was louder than the first, we increased the attenuation value of the stimulus by .01. Conversely, if the participant reported that the second stimulus was softer than the first, we decreased the attenuation value of the stimulus by .01. We did this until the participant reported that the stimuli were either equally loud or that one had overtaken the other in loudness or softness. Every stimulus was matched to every other stimulus. This produced three attenuation values for each stimulus. The average of these three attenuation values was used in the primary ERP experiment. The subject performed the loudness matching task after the electrode cap had been secured to

the head, in order to reduce acoustic differences between the loudness matching task and the primary ERP experiment.

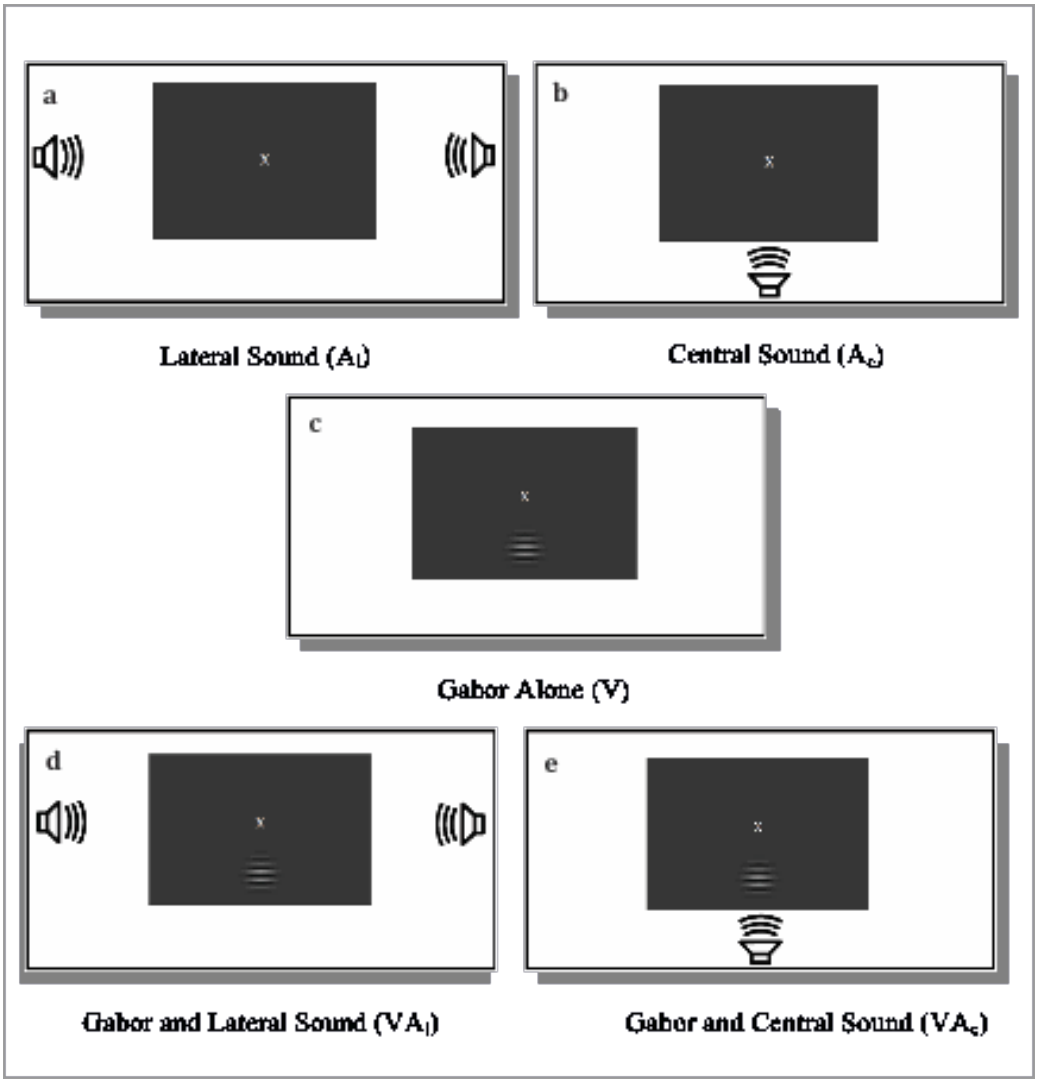

Figure 1., The five possible configurations of A, V, and AV stimuli.
a) Bilateral noise/tone alone (A_{lbb}, A_{lpt}). **b**) central noise/tone alone ($A_{cbb,}$ A_{cpt}). **c**) Visual Gabor alone (V_a). **d**) bilateral audiovisual (AV_{lbb}, AV_{lpt}). **e**) central audiovisual (AV_{cbb}, AV_{cpt}).

Procedure

Participants sat in front of a computer screen, in a sound attenuated, electrically shielded room. They were asked to keep their heads and bodies as still as possible during the trials, and minimize blinks as much as possible, without causing discomfort. They were told to direct their gaze at the white fixation cross in the center of the screen. In one out of approximately ten trials, the fixation cross became slightly grayer (dimmer). They were instructed to press a button as quickly as possible when the fixation cross grew dim momentarily. On these dim trials, any of the nine stimuli were equally likely to show up. These infrequent dim-target trials were excluded from the analysis. Moreover, to further control for response-related neural activity, if a response occurred within -1000 or 1000ms of stimulus onset, the trial was discarded. Between each block, participants were allowed to take a break for as much or as little time as they wished.

Excluding the dim trials, each of the nine different stimulus types was presented fourteen times per block, and each block consisted of 126 relevant trials. Each participant completed a minimum of thirteen blocks, which corresponds to a minimum of 182 trials per condition. Most participants, however, completed closer to 200 trials per condition.

EEG acquisition and analysis

Continuous EEG was acquired from 96 scalp electrodes. Analog to digital conversion occurred at a 500 Hz sampling rate. While recording, the EEG data was referenced to the electrode just posterior to the vertex, which overlaps with the electrode

CPz in the 10-10 international system of electrode placement.[1] In the off-line analysis stage, the data was referenced to the average of all electrodes. The data were epoched at -100ms pre-stimulus and 500ms post-stimulus. The electrical data was band-pass filtered from 0.05 to 100 Hz. Moreover, a notch filter was applied in order to remove 60 Hz electrical noise. Semi-automatic artifact rejection was performed separately for each participant. The 100ms window prior to stimulus onset was designated as baseline. For each participant, the data underwent an algebraic manipulation before any statistical testing. The Gabor alone ERP was added to each auditory alone ERP, producing: (A_{cbb} + V), (A_{cpt} + V), (A_{lbb} + V), and (A_{lpt} + V). Some participants had alpha waves (approximately 10 hz waves) embedded in one or more of their conditions, which are generally an indication of fatigue, boredom, or mind-wandering. If the presence of alpha was prominent in all the ERPs, or if it was more prominent in one ERP than in another, the participant's data were excluded from further analysis.

Again, a 'pair – sum' difference wave refers to the difference in amplitude between the sum of auditory alone and visual alone ERPs (A+V), and the audiovisual, or 'pair' ERP (AV). The ERP components and regions of interest (ROI's) to be probed for 'pair – sum' differences were chosen according to previous studies, which have reported similar phenomena (Giard and Perronet, 1999; Fort et al., 2002; Molholm et al., 2002; Vidal et al., 2008; Cappe et al., 2010; Senkowski et al., 2011) and according to the ERP components that could be visually identified in our data. The components of interest were the visual C1, and P1. In order to temporally and topographically delimit each of these components, the peak amplitude was located in time and topographic space. For instance,

[1]http://faculty.ksu.edu.sa/MFALREZ/EBooks%20Library/EEG/Fundamental%20of%20Electroencephalogram.pdf

the C1 component was greatest at 94 ms, and at the most posterior electrode. Therefore, the five electrodes surrounding the most posterior electrode were chosen for analysis, and the twenty ms time window surrounding the 94 ms mark was chosen for analysis (figure 4) (Guthrie and Buchenwald, 1991). In the case of the C1 component, the 'pair-sum' difference wave coincided almost exactly with the C1 produced by the visual alone condition. For this reason, the exact same time window and electrode array for the visual C1 were used to compare 'pair' and 'sum' voltages. However, during the P1 spatiotemporal window, the 'pair – sum' difference wave peaked slightly earlier, and slightly more dorsally than the actual P1 component produced by the visual alone condition. Therefore, analysis was performed on a spatio-temporal window surrounding the greatest amplitude *difference* between 'pair' and 'sum' voltages. Specifically, the 'pair – sum' difference wave was greatest at approximately 106 ms, and between four, right hemisphere electrodes, just dorsal to the P1 component (figure 7). Therefore, these four electrodes were subjected to analysis for the time period of 98 – 114 ms. Although the two time windows overlap, the 'pair' wave is more positive than the 'sum' wave during the first time window, and the 'pair' wave is more negative than the 'sum' wave during the second time window. Moreover, different electrodes are subjected to each statistical comparison. For these reasons, if both comparisons produce significant results, it is unlikely that the voltage difference during the overlap (98 – 104 ms) is responsible for both significances.

Once these components were identified, mean amplitudes within the designated time-windows and electrode sites were subjected to an omnibus, 5-way, repeated measures ANOVA with factors stimulus type (unimodal sum vs. bimodal), auditory

spectral characteristic (broadband vs. pure tone), auditory location (central vs. bilateral), hemisphere (left vs. right), and electrode. If interactions resulted from this ANOVA, a follow up 4-way ANOVA was performed to derive the underlying factors driving each interaction (figure 3). This follow up method was repeated in some cases, to uncover specific effects of spectral characteristic, and location. However, these follow up comparisons were only performed according to prior predictions. A very similar approach was also used by (Senkowski et al., 2011) and is representative of a number of previous findings.

Results

Behavioral Data

On average, participants correctly identified the dimming of the fixation cross 89% of the time.

Event-related Potentials to Unimodal and Multimodal Stimuli

The gabor patch was presented approximately 2 degrees below fixation. According to Clark et al. (1995), the polarity of the C1 inverts slightly below fixation. Therefore, even though the visual stimulus was presented to the lower visual field, it was not presented below the line of C1 inversion, and thus was negative in polarity. The visual stimulus was presented below fixation in order to make it more congruent with the central auditory stimulus, which was presented directly below the monitor. In response to V_a trials, the C1 reached a maximum around 100 ms, which was centered over the most posterior aspect of the scalp (electrode 63, corresponding to Oz). At the center of its scalp distribution C1 returned to baseline at ~120 ms.

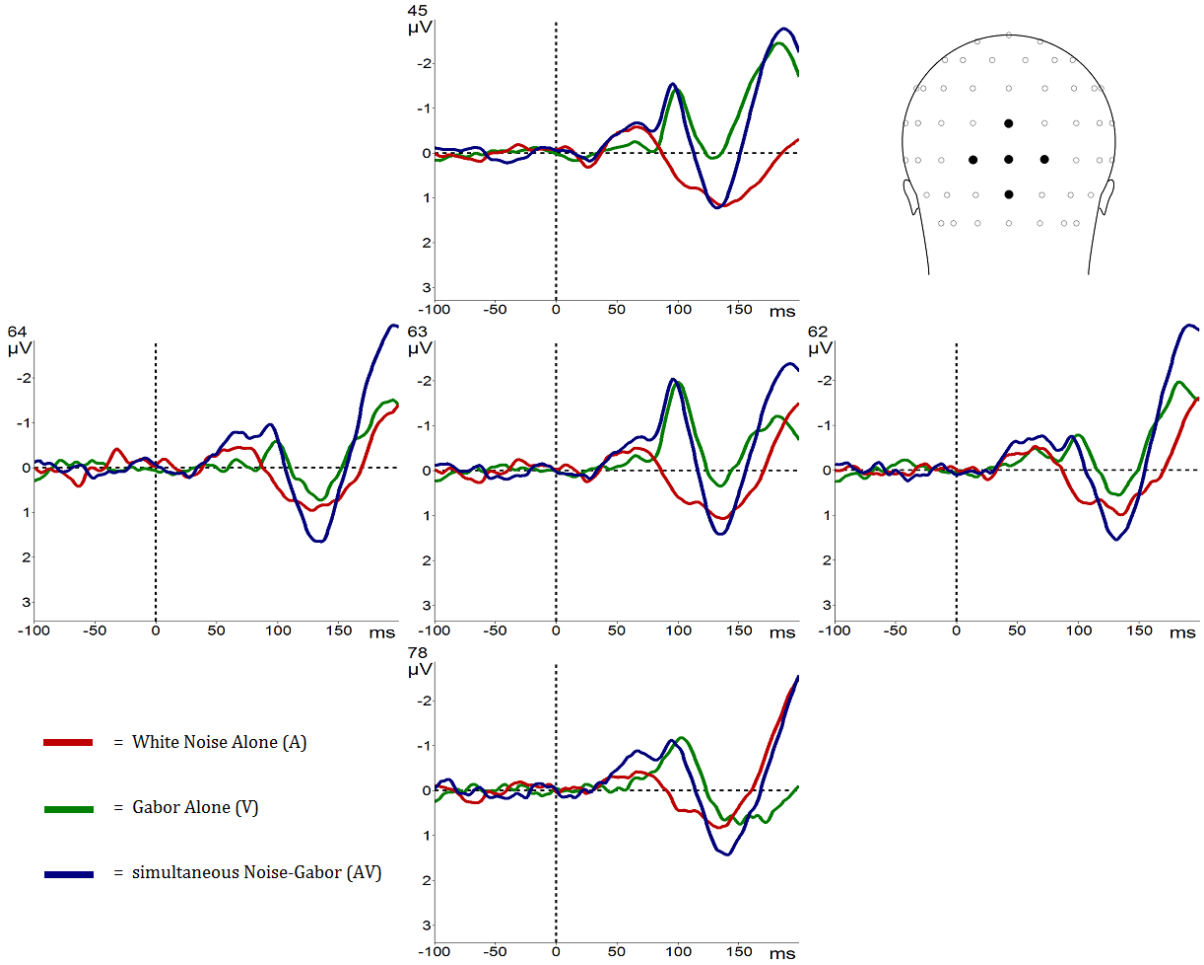

Figure 2. Grand averaged ERP waveforms for unimodal and audiovisual stimuli
The five electrodes that were subjected to the C1 analysis are plotted here. The emboldened black dots on the head in the upper right corner correspond to these five electrodes. The C1 clearly peaks at the central-most electrode. Visually subtracting the red trace (white noise) from the green trace (gabor) would produce the (A+V) trace, whose peak would be lower than that of the blue trace (AV).

In the AV_{cbb} condition, a distinct amplitude increase preceded the normal C1 component by approximately 20 ms. However, this early increase covered most of the occipital scalp and therefore likely does not reflect a specific visual component, but probably reflects an artifact of using an average reference. (e.g. the opposite end of the dipole that generates the auditory P1 component over the fronto-central scalp). This technical issue of average referencing will be explained further below. The C1

component exhibited comparable amplitude and scalp distributions in AV_{cbb} and V_a conditions.

The P1 component had a typical bilateral occipital scalp topography. AV_{cbb} showed a two- to three-fold amplitude increase, whereas the AV_{cpt} showed a three- to five-fold amplitude increase, depending upon the electrode. This can be observed for the AV_{cpt} condition in figure 3.

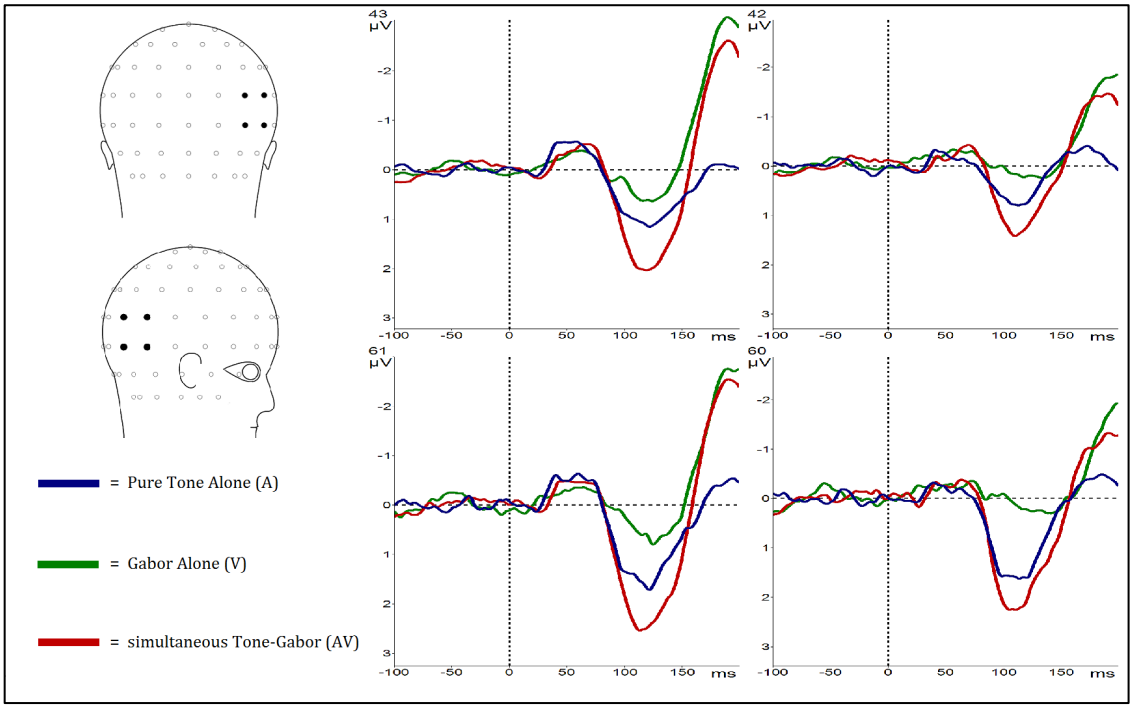

Figure 3. ERPs to A, V, and AV stimuli for centrally presented pure tone.
The emboldened black dots represent the four electrodes that were submitted to analysis.
These dots correspond to the ERP traces on the right, in the same spatial configuration.
Right) ERP traces to A, V, and AV stimuli. The blue arrow points to the P1 component.

The ERP waveforms of all four auditory alone conditions showed normal auditory components over the fronto-central scalp (P1, N1, P2) (See appendix A). However, no AV vs. (A+V) differences were apparent, thus, the present analysis focused on audio-visual interactions measured over the posterior scalp only.

Event-related Potentials of Multisensory interactions

In order to analyze early multisensory interactions, the primary comparisons were between amplitudes of the posterior C1 and P1 components for the AV vs. the (A + V) conditions. For the centrally presented noise, the visual C1 component showed a greater negative amplitude in the simultaneous compared to the sum conditions. This voltage difference overlapped spatially and temporally with the AV C1 component. In response to the centrally presented tone, the visual P1 component showed a greater positive amplitude in the simultaneous compared to the sum conditions. In contrast to the C1 difference wave, this P1 voltage difference seemed spatially and temporally distinct from both the AV and the (A+V) waveforms (figure 9).

C1 analysis

When subjected to a 4-way omnibus ANOVA with factors electrode (5), stimulus type (AV vs. A+V), location (central vs. bilateral), and spectral characteristic (broadband vs. pure tone) the mean amplitudes between 84 and 104 ms did not reveal a main effect of stimulus type, $F(1,11) = 1.82$, $p=0.203$. However, two separate three-way ANOVA's revealed a trending interaction between stimulus type and spectral characteristics for centrally presented but not for laterally presented stimuli, $F(1,11) = 3.79$, $p=0.077$. The mean AV – (A+V) difference was greater for centrally presented noise than centrally presented tones. Accordingly, when centrally presented broadband white noise was subjected to a follow-up two-way ANOVA with factors stimulus type and electrode, this revealed a significant main effect of stimulus type ($F(1,11) = 6.022$, $p=0.032$)). However,

when performed on centrally presented pure tones a similar two-way ANOVA revealed no main effect of stimulus type, $F(1,11) = 0.326$, $p=0.579$. The mean values of AV and (A+V) for all eight combinations of spectral characteristic and location are plotted in figure 5. Moreover, in figure 4, an ERP plot displays the ERP traces corresponding to AV, (A+V), and the difference between the two (AV – (A+V)) for the central noise comparison.

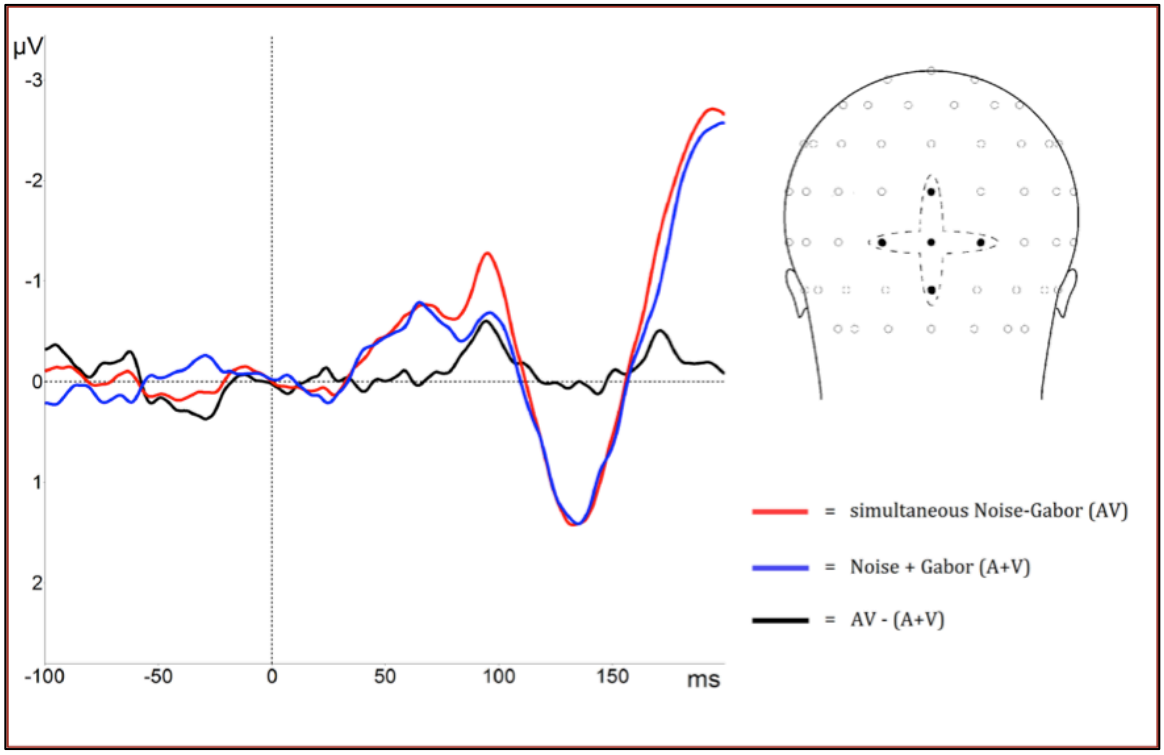

Figure 4. grand averaged 'simultaneous' (AV), 'sum' (A+V), and difference (AV – (A+V)) ERPs for centrally presented noise.
Left) The five C1 electrodes (on right) were averaged together, resulting in this single pooled electrode. When the blue ERP is subtracted from the red ERP (AV), the result is the black ERP (difference). One can clearly observe the amplitude increase for AV compared to (A+V) ERPs. The blue arrow points to the peak of the C1 component.

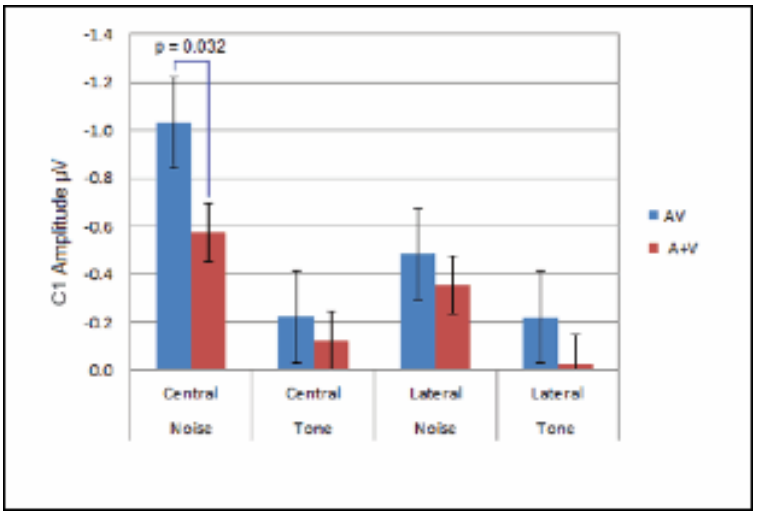

Figure 5. Average mean amplitude values across the the five C1 electrodes between 84 and 104 ms.

AV and (A+V) compared for four different conditions. Blue is 'simultaneous' and red is 'sum.' Only central noise was significantly different between AV and (A+V)

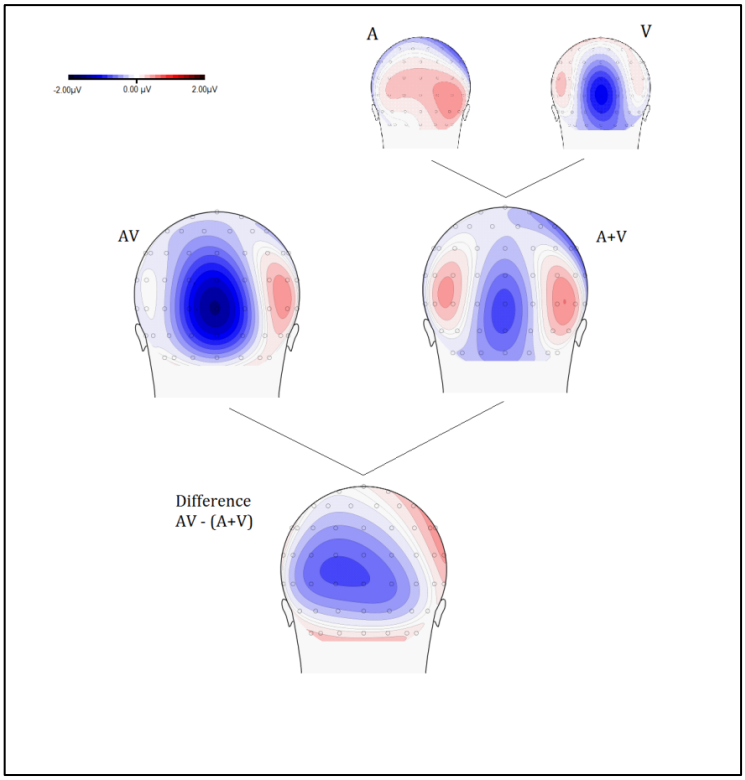

Figure 6. Voltage topographies at 94 ms for the white noise condition.

Views of the back of the head. Blue represents negative voltages and red positive. Upper Left) Voltage scale which ranges from -2.00 microvolts to 2.00 microvolts. The distribution of the difference wave is largely overlapping with the C1 itself.

P1 Analysis

The later AV – (A+V) difference was shorter lived, and less widespread across the scalp than the C1. Therefore, the time window selected for analysis was only 16 ms wide (98 – 114 ms), and included four electrodes on the left and right parietal-occipital scalp (see figure 7).

When subjected to a five-way omnibus ANOVA with factors stimulus type (AV vs. (A+V)), spectral characteristics (Noise vs. Tone), location (Central vs. Bilateral) hemisphere (left vs. right), and electrode (4), no main effects of stimulus type were observed, but a significant interaction between hemisphere and stimulus type was apparent, $F(1,11) = 7.972$, $p=0.040$. The mean differences between AV and (A+V) was approximately three times greater for the right compared to the left hemisphere, therefore a follow-up four-way ANOVA was performed on the right hemisphere only, revealing a trend for stimulus type, $F(1,11) = 2.652$, $p=0.131$. Again, observation of mean amplitudes revealed that the mean AV – (A+V) difference was much greater for central tones and lateral noise than for lateral tones and central noise. This pattern of results accounts for the lack of significant interactions between stimulus type, location, and/or spectral characteristic. A final two-way ANOVA for centrally presented tones with factors electrode and stimulus type revealed a significant main effect of stimulus type ($F(1,11) = 5.375$, $p = 0.0406$)). In contrast, when submitted to the same ANOVA, lateral noise did not exhibit a significant main effect of stimulus type ($F(1,11) = 1.671$, $p = 0.2226$)).

The reader may be curious about the greater negativity observed in (A+V) compared to AV waveforms between 30 and 90 ms. This negativity was not statistically analyzed because it is likely not a true component. As mentioned earlier, using an

average reference has serious limitations and caveats (Luck, 2005). This is probably a good example of an ERP peak that is an artifact created by using an average reference. Specifically, this early negativity is almost perfectly reflected by the P50 wave over the central scalp. The latency and amplitude differences between AV and (A+V) observed in figure 7 are almost exactly mimicked over fronto-central scalp, but in the positive as opposed to the negative direction. The positivity of the P50 raises the average value that all electrodes are compared to. This is especially relevant to the P50. Since it precedes most other brain components, its positivity does not get cancelled out by a variety of positive and negative voltage values across the scalp. To be sure, this problem probably does not contribute significantly to the C1 and P1 effects. Firstly, upon visual inspection of the voltages across the scalp, there are no obvious or widespread differences between AV and (A+V) waveforms within the selected time windows. Moreover, the many positivities and negativities across the scalp during these time windows probably produce a more consistent average voltage against which these components are compared.

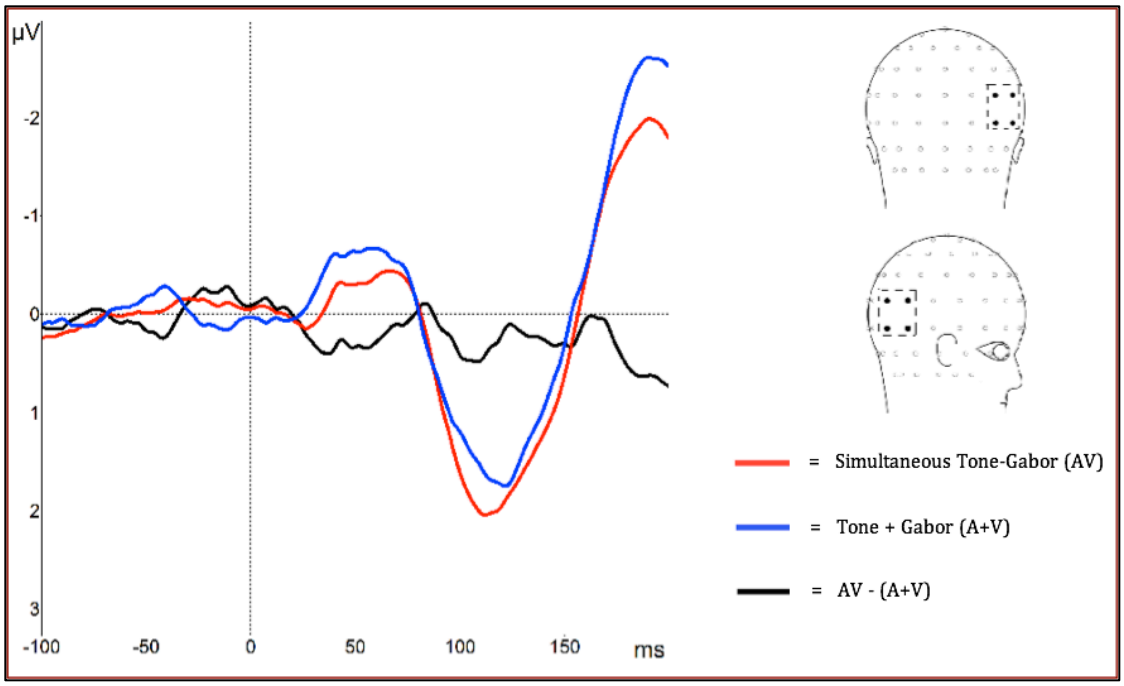

Figure 7. Grand-averaged ERPs of simultaneous (AV), Sum (A+V), and difference (AV – (A+V)) waves for central tone.

The single plot shown is an average of the four P1 electrodes. Right) The four electrodes submitted to analysis, and averaged to create the plot on the left.

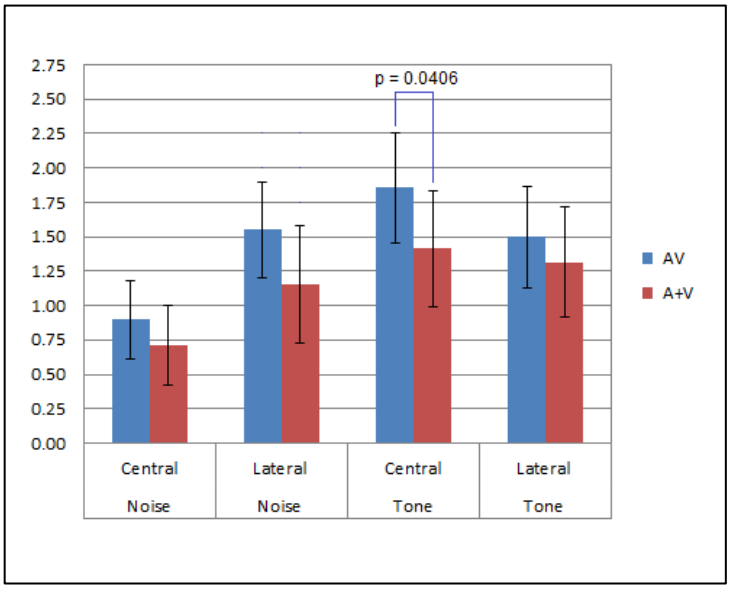

Figure 8. Average mean amplitude values across the the five P1 electrodes between 98 and 114 ms.

AV and (A+V) compared for four different conditions. Blue is 'simultaneous' and red is 'sum.' Only lateral tone was significantly different between AV and (A+V)

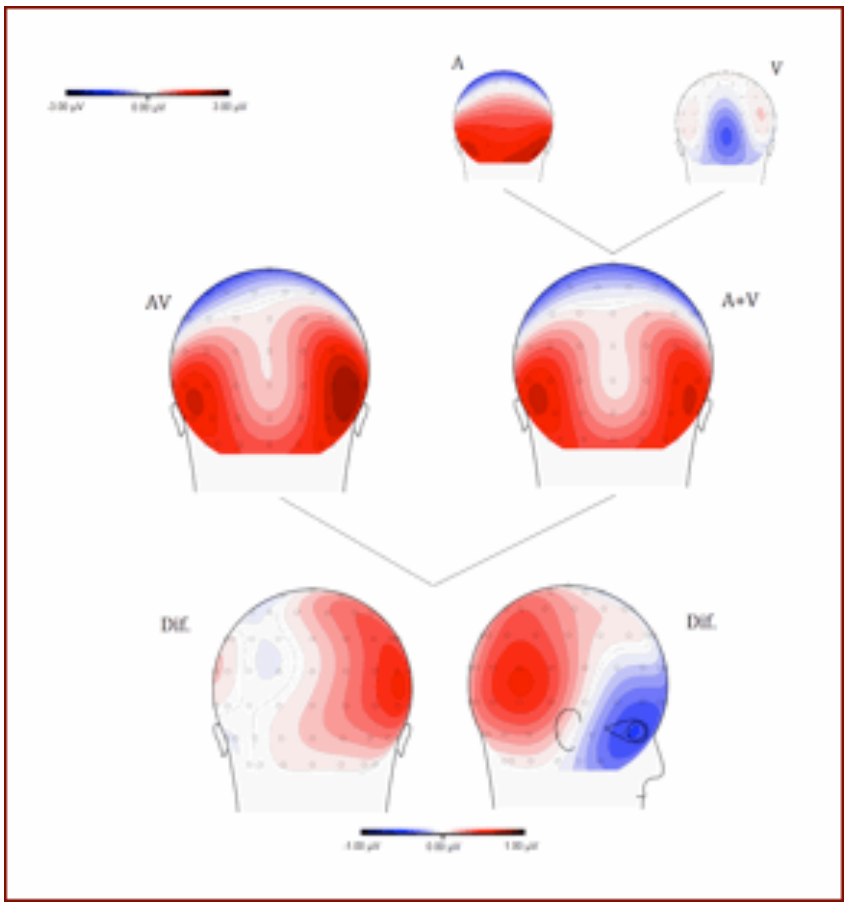

Figure 9. Voltage topographies at 106 ms for the white noise condition.

Views of the back of the head, except for the lower right head, which is viewed from the right side. Blue represents negative voltages and red positive. Upper Left) Voltage scale which ranges from -3.00 microvolts to 3.00 microvolts. The distribution of the difference wave is considerably more dorsal compared to the right side P1 distribution.

Discussion

In the current investigation two separate instances of increased mean amplitude were observed for 'simultaneous' (AV) compared to 'sum' (A + V) conditions. AV stimuli containing centrally presented broadband white noises increased the mean amplitude over occipital scalp within the temporal window of the C1 component compared to A+V stimuli of the same type. In contrast, AV stimuli containing centrally presented, single frequency pure tones increased the mean amplitude over the parieto-occipital scalp in a temporal window partially overlapping with the visual P1 component compared to A+V stimuli of the same type. The spatiotemporal profile of the early modulation for the centrally presented noise stimuli coincided closely with the C1 component itself, whereas, the latter difference for centrally presented pure tones only appeared during the first half of the P1 component, and showed a more anterior/parietal scalp distribution than the P1 component itself. Importantly, both of these amplitude modulations only occurred when the auditory and visual stimulus components were presented in close spatial proximity (in this case centrally). No multisensory interactions during the C1 or P1 intervals were apparent when the auditory stimuli were presented at locations more distant from the visual stimuli (in this case from bilaterally positioned speakers).

Spatial Congruence or Top-down Attention

When an auditory and a visual stimulus are spatially and temporally congruent, they often integrate with one another, rendering a single audiovisual object, or at least a single audiovisual event. The central sounds, but not the bilateral sounds, were spatially congruent with the visual stimulus. This may explain why AV interactions occurred for central sounds only. To a lesser extent, the sounds were spatially congruent with the task-relevant fixation cross, and may have integrated with that visual object as opposed to the gabor patch. While the C1 produced here was induced by the gabor patch and not the fixation cross, other research has shown early occipital interactions in the absence of a robust C1, further suggesting that the effects here could have been unrelated to the gabor patch. However, the AV interactions overlapped almost exactly with the visual C1 component, which was certainly induced by the gabor patch. Also, the fixation never disappeared and reappeared simultaneously with the sound, as the gabor patch did. This might preclude the integration of the fixation cross and the sound into an object. However unlikely, this possibility would significantly rearrange the following discussion of attention and cross-modal integration, and should not be ruled out entirely.

In this study attention was not directed towards the AV stimuli, yet significant AV interactions occurred over the occipital scalp. In contrast, Talsma et al. (2007) observed similar AV interactions only when AV stimuli were directly attended. From where does this discrepancy arise? In both their study and the current one, the visual stimulus was presented at the center of the screen. But, crucially, the sounds came from two speakers, placed slightly behind, and lateral to the computer screen. They claim that this

configuration matched the subjective location of the sound and the visual stimulus. And they justify this claim with the fact that visual attention to one point in space can enhance auditory processing in that location (Talsma, 2010; Eimer and Van Velzen, 2002). But the fact that top-down spatial attention spreads across modalities does not pertain to the spatial congruence of two stimuli. Rather, whether or not two stimuli are spatially congruent depends upon bottom-up receptive field properties. While a thorough exploration of source localization in the midbrain lies outside the scope of this paper, a more analogical argument can be made against this possibility. The Superior Colliculus is essential for the coordination of eye-head movements, and saccadic orienting responses to targets (Calvert et al., 2004). These tasks require extreme accuracy. Therefore, it seems unlikely that at this processing stage, auditory localization is so imprecise that two bilateral stimuli would be localized to a single central source.

Within their methods section, the experimenters admit that without cross-modal attention, their unique configuration of AV stimuli is not spatially congruent. But then, they conclude that attention is necessary for all early cross-modal interactions. Alternatively, their results suggest that cross-modal spatial attention is necessary for integrating spatially incongruent stimuli. To be fair, the authors do mention that the speaker locations could possibly confound their results. But the present results may challenge their main conclusion, suggesting that early AV interactions can occur even for task-irrelevant AV stimuli. Taken together, the work of Talsma et al. (2007) and the present study propose that top-down attention may influence cross-modal receptive field size. Direct attention to the AV stimuli may loosen the criteria for spatial congruence.

In the present paradigm, top-down attention and multisensory integration may be confounded for a different reason. Spatially congruent stimuli showed multisensory enhancement while spatially incongruent stimuli did not. This could be accounted for by receptive field congruence. But, the central auditory stimuli were proximal to the attended fixation cross, while the bilateral auditory stimuli were located many degrees of visual angle from fixation. If attention can spread across modalities, the attention paid to the fixation cross may have carried over to the central but not the lateral auditory stimuli. (Talsma, 2010; Eimer and Van Velzen, 2002). At first glance, one might assume that central sounds interacted with the gabor patch because they fell within the locus of visual attention. But the spread of visual attention to the auditory domain does not enhance auditory processing until 220 ms at the earliest (Busse et al., 2005). This is far too late to have any sort of effect on the interactions observed here. The present effects, therefore did not occur because top-down attention to the fixation cross spread to auditory processing.

The central position of congruent AV stimuli also provokes another, entirely different, interpretation of the observed effects. If we think of attention as a limited capacity system (Lang, 2000), processing of the irrelevant stimuli (A, V, and AV) will draw on the same resource pool as the task-relevant fixation cross. At least in terms of bottom-up attention, the resources devoted to the fixation will be inversely proportional to those devoted to the irrelevant stimulus. Specifically, a spatially and temporally congruent pair of stimuli will capture attention more effectively than an incongruent pair. Therefore, the congruent AV stimuli might actually present a more competitive challenge to the fixation task than the incongruent AV stimuli (Pluta et al., 2011). In other words, in

order to enhance the neural representation of the fixation cross top-down attention also must inhibit the representation of irrelevant stimuli. It may successfully inhibit a distracting visual alone stimulus. But, inhibiting an audiovisual stimulus may require much greater neural effort. This would suggest that the increased voltage observed for congruent AV stimuli reflect a difference in inhibitory effort. However, The P1 is usually increased when a stimulus appears at the attended location (Mishra et al., 2012). A more distracting stimulus would probably minimize this P1 increase: a pattern, which does not fit the present data. And the question of whether the C1 component can be modulated by attention at all is still under heavy debate (Martinez, 1999).

Another obvious difference between this study and similar ones is that here, AV interactions began at approximately 80 ms, while other studies report their occurrence at approximately 50 ms. Naturally, two major methodological differences stand out as potential causes. The first is that in all other studies, a pure tone elicited early visual interactions, while the present early interactions were elicited by white noise. However, bandwidth obviously cannot account for the longer latency because in this study the pure tone elicited AV effects much later than 50 ms. The other major difference is that here, early multisensory interactions were investigated in response to task-irrelevant stimuli. One might assume, therefore, that in order for auditory and visual stimuli to interact at the commencement of visual processing, those stimuli must be task relevant. And after thoroughly weighing the contributions of spatial congruity and task-relevance, it seems plausible that attention is necessary to produce the earliest AV effects. As Talsma et al. (2007) correctly suggest, some early multisensory interactions may *rely* on attention. Attention might enhance the representation of the sound just enough to project its signal

to other brain regions. If this were the case, a certain level of attention would be necessary to enable the AV interaction, even if the stimuli were spatially and temporally congruent.

On the surface, this explanation is seductively simple. But, there are considerable hurdles when attributing any early visual effects to attention. As mentioned earlier, the literature is rife with contention over whether the C1 component can be manipulated by top-down attention. Kelly et al. (2008) claims that early activity in the striate cortex (V1) varies so much across individuals that a between subjects analysis is too imprecise to locate early attention affects. This research team claims to have revealed an effect of spatial cueing on the amplitude of the earliest visual processing (59 ms). While many have provided converging evidence (Rossi and Pourtois, 2012; Karns and Knight, Kelly et al., 2008), many others challenge these results (Clark and Hillyard, 1996; Fu et al., 2009; Di Russo et al., 2003; Noesselt et al., 2002; Martinez et al., 1999). Some of the latter accounts propose that attention does not modulate processing until 90 ms. However, one of these studies suggests that the earliest aspects of the C1 (50-55 ms) are not affected by attention, while the latter aspects of the early visual processing (70 – 75 ms) may be (Martinez et al., 1999). It seems most prudent to frame our results according to this intermediate model, since more depth on this issue falls outside the scope of this study. According to this intermediate model, AV effects starting at 50 ms post-stimulus operate independent from attention or task-relevance. The longer latency observed here cannot be directly accounted for by attention. Even if participants had paid more attention to the AV objects, this would not have enhanced visual processing and the related AV interactions at 50 ms.

However, top-down attention to a sound can modulate auditory ERP components at the impressive latency of 20 ms (Woldorff et al., 1993; Brockelmann et al., 2011). In the present study, participants were instructed to perform a task on the visual fixation cross. As stated earlier, this may have allowed only slight attentional enhancement of the auditory stimuli. In other paradigms, direct attention to both visual and auditory stimuli, probably enhanced auditory processing to a much greater degree (Giard and Perronet, 1999; Molholm et al., 2002; Cappe et al., 2010; Senkowski et al., 2011). Potentially, this attentional gain could have mobilized early auditory signals to more dispersed areas like the visual cortex. This may partially resolve the contradiction between the present study and that of Talsma et al. (2007). Dedicating attention to the AV object might be a sufficient condition for very early AV interactions (50 ms) over visual scalp. Yet, we cannot equate attention, or task-relevance, with multisensory integration. In other words, increased attention to the AV object allowed the auditory signal to arrive in visual areas. But more importantly, the actual AV interaction will occur regardless of how the auditory signal arrives there. It just so happens that attention performs this preliminary function effectively. Also, while fast-acting auditory attention may account for the lag of early occipital interactions observed here, it does not fully account for the early effects over auditory cortex observed by Talsma et al. (2007). Specifically, visual attention cannot enhance visual processing quickly enough to project information to the auditory cortex at 50 ms, where these interactions occurred.

Finally, the visual nature of the task may allow slightly more attention to the auditory compared to the visual aspects of the AV object. When paying attention to auditory stimuli, the Steady State Visual Evoked Potentials (SSVEP) of a constant letter

stream were greater than when paying attention to another visual stimulus or an audiovisual stimulus (Talsma et al., 2006). Although the roles of each modality are reversed in the present paradigm, this evidence suggests that an intra-modal distractor will receive less attention than a cross-modal distractor. Since participants paid attention to a visual stimulus in this study, the gabor patch may have received less attention than the sounds. This suggests that neural representations of sounds benefitted more from attention than the gabor patch, possibly allowing them to reach the visual cortex, and not vice versa. This potentially explains why interactions occurred relatively early over the occipital scalp (80 ms). But since the auditory stimulus did not receive full attention, these early interactions lagged slightly behind those observed in other studies (50 ms).

Spectral Characteristics

When a broadband sound was presented simultaneously with a gabor patch, the auditory and visual neural signals interacted over occipital scalp between 84 and 104 ms post-stimulus. This was reflected as an amplitude increase. In contrast, when a 1,000 Hz pure tone was presented simultaneously with a gabor patch, the auditory and visual signals interacted over parieto-occipital scalp. This was similarly reflected as an amplitude increase. Before jumping into a full interpretation of these multisensory events, we must compare the two types of sound along their physical, sensory and perceptual dimensions.

A pure tone is made up of a sinusoidal air pressure wave, which corresponds to a single spectral frequency. In contrast, white noise has a frequency-independent spectral profile. Basically, this means that its spectral fluctuations are pseudo-random, producing no rhythm or pitch (Fastl and Zwicker, 2007). How do these physical characteristics interact with the human auditory system? To begin with, the cochlea is tonotopically organized. This means that low bandwidth sounds displace fewer hair cells (and therefore fewer auditory nerve fibers) than high bandwidth sounds. Importantly, the stimuli used in this study fall within the 40 – 60 dB range, within which tonotopic neural organization is most accurately preserved (Fastl, and Zwicker, 2007). This implies that the pure tone did in fact activate a considerably smaller range of hair cells than the broadband stimulus. A similar tonotopy may carry over to a lesser extent in the auditory cortex (Wessinger, 2001). Auditory processing relies largely on the spectral and temporal characteristics of a given sound. A sensory system designed in this way should be able to extract much more

information from a broadband noise than a pure tone. A pure tone shows a high-degree of temporal order. But in sensory terms, this order translates into relatively constant activation of a single frequency band. In contrast, a broadband stimulus activates a greater region of the cochlea, and it does so at temporally jittered intervals. These qualities provide the auditory system with a much greater amount of spectral and temporal information to decode. This may be loosely correlated with fMRI research suggesting that a greater area of the auditory cortex is activated by white noise than pure tones (Wessinger, 2001). The relative complexity of two neural events cannot be expressed by simply comparing the amount of BOLD signal change they induce. But in conjunction with the spectral and temporal differences between tones and noises, it makes intuitive sense that a greater swath of cortex is needed when processing more information. Moreover, the regions that respond most to white noise and not pure tones are generally designated for pattern recognition of different forms (Kaas and Hackett, 1999). At a basic sensory level, the human auditory system can extract less information from a pure tone than a white noise.

If pure tones do in fact carry less information than white noise, the principle of inverse effectiveness should have some utility in our analysis of the present spectral variability. But at first glance, the findings contradict inverse effectiveness. Specifically, if the white noise is more salient, contains more information, or is more effective in any way, it should result in less cross-modal integration. In direct opposition to this prediction, a white noise actually produced an earlier, more statistically reliable interaction than a pure tone. In many examples of inverse effectiveness, the super-additivity of cross-modal interactions increases as stimulus intensity decreases. But often,

researchers will vary the overall effectiveness of the AV stimuli, instead of varying the intensity or effectiveness of each modality separately (Senkowski et al., 2011; Stevenson and James, 2008). These manipulations might gloss over a more nuanced variety of inverse effectiveness. Imagine that the visual component of an AV object is less salient than the auditory component. In all likelihood, the salient auditory component will lend information to visual processing, and not the other way around. The sound might fill in crucial gaps in visual recognition. But since the auditory component pops out so clearly, adding visual information would be redundant. This could manifest as an increase in multisensory interaction within visual areas and a concomitant decrease in interactions within auditory processing areas.

This is a relative, as opposed to an absolute form of cross-modal interaction, which was recently tested in single cells of the dorsal region of the Medial Superior Temporal area (MSTd) of macaque monkeys (Morgan et al., 2008). The monkeys were presented with visual and vestibular cues that indicated a direction of motion. Some cells responded to both visual and vestibular cues. These cells preferred certain directions of motion over others (i.e. fired more in response to a certain direction). Between modalities, these direction preferences overlapped but were not identical. On some trials, vestibular and visual cues were presented simultaneously. Now, each cell had a direction preference value for visual alone, vestibular alone, and visuo-vestibular trials. In this sense, the experimenters could assess which of the two unimodal conditions contributed more to the bimodal direction preference. Lastly, on some of the bimodal trials the visual cue was half as reliable. This reduction should reduce the overall input from visual areas, resulting in a linear shift towards the vestibular preference. In contrast, the bimodal

response actually shifted more towards the vestibular preference than predicted by a constant weight model. So a visual motion decrement resulted in a non-linear weight adjustment towards the vestibular preference. In this study, the relationship between tone and noise may be roughly analogous to the relationship between 50% and 100% cue reliability, respectively. And while single neuron recordings are only indirectly linked to recordings at the scalp, multisensory properties such as inverse effectiveness or spatial congruence have been observed at both levels. Since a pure tone may contain less auditory information than a white noise, it should offer less corroborative information to visual processing than a white noise. Physiologically, this may be reflected by the more occipital distribution of white noise interactions compared to pure tone interactions.

This theory, that modalities undergo increases or decreases in combinatorial weight, has only been applied to stimulus factors. But endogenous or innate modality dominance may also determine the relative combinatorial weight of auditory and visual signals in AV interactions. In the first study of early AV interactions (Giard and Perronet, 1999), Some participants responded to visual stimuli faster than auditory stimuli, while others exhibited the opposite reaction time pattern. With these measures, each participant was roughly designated as auditory dominant or visual dominant. The researchers found that within the first 150 ms, those participants who were visually dominant exhibited small interaction effects over visual cortex, and large effects over auditory cortex. The opposite pattern emerged for auditory dominant individuals. This finding fits intuitively with the theory of relative combinatorial weight. If an individual is predisposed to focus on sound, they should process the auditory component of an AV object more proficiently than the visual component. As a result, the auditory cortex provides information for the

visual cortex. Accordingly, much greater AV interactions were observed over the visual cortex in AUD-dominant subjects. This also clarifies the distinction between AV integration and mechanisms that facilitate it. Specifically, an auditory dominant individual who attends to A and V modalities, and a visually dominant individual who attends to just the Auditory modality, might produce a similar pattern of AV interactions. The method by which A and V signals converge should not determine the resultant pattern of interactions.

The final piece of evidence for relative inverse effectiveness suggests that human behavior may rely less heavily on a modality when the stimulus reliability decreases in that modality. In a psychophysical study of height estimation, participants reported how high a virtual bar was in response to visual and haptic cues (Ernst and Banks, 2002) The virtual cue was a force feedback device that simulated the experience of a touching a horizontal bar, and the visual cue was a binocular random dot array, that simulated the visual experience of a bar. On some trials, the perceived depth of the dots varied randomly, reducing the reliability of the visual cue. On these trials, participants relied more heavily on haptic cues, supporting the theory that relative weights can shift according to reliability.

Until now, the relative stimulus strengths of noise and tone have been weighed against one another. But a similar weighting between the visual and auditory aspects of the AV object may speak to the more occipital distribution of AV interactions. Particularly, if the present visual and auditory stimuli are of relatively equal salience, AV interactions should occur over both sensory cortices, (occipitally and fronto-centrally), or neither. Why, then, were only occipital interactions observed? Harkening back to the

discussion of inter- and intra-modal competition, one is reminded that in the present paradigm, the auditory components of the AV object may have received more attention than their visual counterparts. Moreover, the present AV interactions only modulated visual ERP components, and only occurred over occipital and parieto-occipital scalp. Therefore, greater attention to the auditory modality may have contributed to the occipital distribution in two ways. First, it may have projected the neural representations of sounds to the visual cortex, thereby allowing the interactions to occur. Secondly, it may have enhanced the quality of these representations to a degree where they could inform visual processing, and not the other way around. These two mechanisms of attention may be linked. The endogenous nature of attention may be likened to the endogenous nature of modality-dominance. Either, or both of these factors could simultaneously determine which modality provides aid, and which modality receives aid (Giard and Perronet, 1999).

Thus far, a few things have been moderately well established. In some cases and not others, the properties of spatial receptive fields and the principles of inverse effectiveness are reproduced here. In contrast, an entirely new finding is that noise may modulate earlier, more unisensory visual regions than pure tones. And this may be due to the greater amount of information contained in white noise compared to pure tones. Fourth, the visual nature of the task suggests that auditory processing is more enhanced than visual processing. And finally, in combination with the theory of relative inverse effectiveness, the imbalance of attention may explain why AV interactions modulated visual, as opposed to auditory, ERP components. Presently, the neural and psychophysical differences between tones and noise have provided an interesting

information based theory. But thus far, this theory only applies to the auditory cortex, where none of the present interactions occurred. Are these differences in the auditory cortex preserved in its projections to other areas?

Anatomical Connections

In general, human and monkey auditory cortices share cyto- and chemo-architectonic organization, (Sweet et al., 2005). Specifically, the auditory cortices of both species are grossly divided into three sections. Different names correspond to each species, but they can be generally referred to as the core, lateral belt, and parabelt regions. The monkey auditory core, analogous to A1 in humans, responds well to pure tones, while peripheral areas respond almost exclusively to broader-band stimuli (Rauschecker, 1997). Similarly, central areas of the human Superior Temporal Gyrus (such as A1) respond to both pure tones and broadband noise. But peripheral areas like the parabelt respond to noise and not tones (Wessinger et al., 2001). Clavangier et al. (2004) found many projections from the macaque auditory cortex to V1, but interestingly, 70% of them originated in the peripheral caudal parabelt, while a much smaller percentage originated in the core region. A similar pattern was found for ferrets (The neural basis for multisensory processes), and in another macaque study (Rockland and Ojima, 2003). Thus, the evidence suggests that noises are well represented in unisensory visual areas, while tones are not. This outlines one of many possible pathways that the white noise could have taken to early visual cortex. Moreover, in this case, the greater representation of white noise compared to pure tones in the auditory cortex carries over to visual areas.

Another tracing study that found fibers projecting from peripheral auditory areas to V1, found even more fibers projecting to area V2 (Rockland and Ojima, 2003). These fibers projected to dorsal regions of V2. Therefore, the white noise in this study should

only interact with dorsal regions of V2, whereas the visual alone stimulus should activate a large portion of V2. Interestingly, the voltage distribution of the noise interaction was slightly dorsal and lateral to that of the visual alone condition (figure 6). Moreover, V2 is slightly deeper in the brain than V1. Deeper neural generators often produce wider topographic distributions, (Luck, 2005). Correspondingly, the topographic voltage map of the present AV noise interaction appears wider than that of the visual alone distribution. It is possible that the interaction distribution may just represent activity in V2, while the visual alone voltage distribution represents activity in both V1 and V2. Certainly, if the above pathway in macaques has a human homologue, its origins in parabelt areas, coupled with its terminations in dorsal V2, make it an excellent candidate for the present AV interactions in C1.

Other studies of early AV interactions report vaguely similar scalp distributions (Giard and Perronet, 1999; Molholm et al., 2002; Cappe et al., 2010). Many of these researchers have reported an AV interaction that is shifted dorso-laterally from the typical C1 distribution. However, although it is difficult to compare scalp distributions without properly quantifying them, the distributions in other studies are consistently more dorsal then the one reported here, and shifted further from the C1 distribution than reported here. Moreover, one of these authors attributed the interactions to dorsal visual areas like MT+ and V5 (Molholm et al., 2002) because of their early response profiles, and vaguely multisensory profiles. More detailed source analysis attributed the interactions primarily to the superior temporal sulcus, for similar reasons (Cappe et al., 2010). But the effect has also been attributed to earlier visual areas like V1 and V2 (Cappe et al., 2010; Giard and Perronet, 1999). Obviously, even complex source localization techniques are

inconclusive, and when confronted with the poor spatial resolution of ERP data, these methods alone will not explain the difference between one study and the next, or one condition and the next. The slightly more ventral distribution and the longer latency (84 ms compared to 50 ms) observed for present early interaction might suggest an entirely different pathway. Thus far, the spectral differences between noises and tones have only suggested a greater effectiveness or a greater neural representation for noises.

Even without extensive knowledge of the human brain, one should be intrigued upon learning that a small percentage of V1 projections originate in the macaque primary auditory cortex (Clavagnier, 2004). This is particularly exciting in regards to AV interactions occurring over visual cortex at 50 ms, at which time visual processing should still be focused primarily in V1 (Di Russo et al., 2003). Some of the effects observed in other studies rely on very early communication between the cortices. Importantly, these early AV effects (50 ms) have all been in response to a pure tone. As we now know, pure tones are well represented in A1. It is possible, therefore, that the pure tones in those studies were projected from A1 to V1 via this hypothetical pathway, where they interacted with feed-forward V1 activity. The lack of task-relevance in this paradigm may have blocked the utilization of this pathway.

But this initial excitement is quickly tempered upon further investigation. These A1 → V1 connections are sparse, and terminate in regions corresponding to the peripheral visual field (15 – 20 degrees away from fixation). Most studies of early AV interactions, including this one, have presented visual stimuli relatively close to fixation, or at fixation. For this reason, it is unlikely that these connections alone are responsible for the 50 ms occipital AV interactions. It may be more plausible to attribute those

interactions to more dorsal areas like MT+ and V5, where auditory and visual signals could meet even at 50 ms. As mentioned earlier, this might suggest that the present early interaction illuminates an entirely different pathway. Specifically, the slightly more ventral scalp distribution may indicate a termination in V2, whereas the more dorsal distribution in other studies may indicate a termination in V5 or MT+. Moreover, the present early interaction may occur 30 ms later than others because it utilizes a different pathway. The projection from auditory belt areas to dorsal V2 seems to have some hypothetical potential. What if instead of a simple enhancement, the broad bandwidth of a white noise actually specifies the region of visual cortex, which it gets projected to? Moreover, if the auditory projections responsible for the present interaction originate in a high-level auditory area, which does not become active until a late processing stage, task-relevance may not be able to shorten the latency of the white noise interaction.

Another possibility is that auditory noise information extends up to a cross-modal brain region, only to be projected back down to early visual areas. At 84 ms, the frontal-parietal network may already have intercepted and responded to sensory input (Foxe and Simpson, 2002). Many feedback projections from macaque parietal areas have been found to terminate in V2 and to a lesser extent in V1 (Rockland and Ojima, 2003). Moreover, research on gerbils shows numerous afferent and efferent connections with both occipital and parietal areas. Therefore, it is plausible that auditory information travelled to visual areas via multimodal parietal areas. A similar parietal function has been suggested for visuo-tactile interactions in humans. Some researchers found that compared to the sum of visual and tactile stimulation, simultaneous visuo-tactile stimulation produced a greater BOLD signal in the posterior lingual gyrus (visual area),

primary somatosensory cortex, and the posterior parietal lobule, (Macaluso et al., 2000). Importantly, they suggest that anatomical models do not support a direct link between the lingual gyrus and the primary somatosensory cortex. For this reason they suggest that multisensory integration in these areas may only occur when feedforward signals to the parietal cortex are sent back down the hierarchy to unisensory areas. In some instances, a higher region may send coordinated feedback signals to unisensory regions, which facilitate their communication. This could be a potential mechanism by which the visual cortex resets the phase of auditory oscillations preparing it for incoming sounds (Thorne et al., 2011). Regardless, there seem to be many ways in which the white noise representation could have travelled to the visual cortex, especially at 80 ms.

In contrast to the noise, the pure tone produced AV interactions within the P1 wave. These interactions fell over parieto-occipital scalp, and began around 98 ms. This distribution corresponds closely to multisensory regions such as the Posterior Parietal Cortex (PPC), and more specifically, the Superior Temporal Sulcus (STS). Especially the latter region lies in between auditory and visual cortices, and for is generally more dedicated to multisensory functions. For these reasons, the possible anatomical pathways and neural generators are less mysterious. In comparison to the noise interaction, so many potential pathways and regions could account for the tone interaction, that they will not be discussed in such detail. However, the anatomical basis for this interaction may further describe its relative dearth of information, in comparison to the white noise. STS lies in close proximity to the primary auditory cortex, where this tone is well represented. The pure tone may have lacked the effectiveness, salience, or intensity necessary to innervate

visual cortex. Instead, it may have been integrated with the visual stimulus at a later stage, namely within the P1 component.

Limitations and Future Research

In the superior colliculus of cats, response enhancement of congruent stimuli often goes hand in hand with depression of incongruent stimuli (Meredith and Stein, 1996; Stein and Meredith, 1993). In other words, a cell fires less frequently to a spatially or temporally incongruent AV stimuli than it would to the sum of visual and auditory stimuli separately. If the spatial properties of the present AV interactions mimic those in midbrain regions, bilateral sounds should induce response depression (i.e. a smaller mean amplitude for AV compared to (A+V)). While visual inspection of the first positive auditory component (P50) hints at this pattern, the difference between AV and (A+V) was much smaller than the difference over occipital regions. In comparison, Talsma et al. (2007) found significant audiovisual response depression of the P50 component when participants performed a distracting visual task. They suggest that this reflects the filtering of the irrelevant audiovisual stimulus. However, they did not manipulate spatial congruence, and a more bottom-up account would attribute this to the slight spatial incongruence of the AV stimulus components. The present results shed little light on this issue because attention was not manipulated. Therefore, future studies should manipulate attention and spatial congruity simultaneously.

In contrast to auditory ERP's, visual ERP's did not even hint at a trending response depression. This has tentative implications for the communication between different levels of multisensory integration. Specifically, if AV enhancement or depression is observed in a certain brain area, the immediate assumption is that signals from different modalities converge in that time and place. Alternatively, it is possible that

AV integration occurred at an earlier stage, and subsequently projected to the area in question (Budinger, 2006). Since no response depression occurred in the incongruent condition, it is unlikely that the present AV effects are residual from lower-order multisensory processing. Different receptive field properties, and their electrophysiological consequences need further testing. As it stands, hierarchical models of unisensory processing suggest that receptive field size and flexibility often increase as information ascends from lower to higher brain areas. But, this should be further tested and verified for cross-modal processing as well. Specifically, do multisensory receptive field sizes increase and or gain flexibility in higher order structures, compared to lower order ones? Moreover, how does task-set affect spatial or temporal multisensory congruence?

Multisensory interactions are often so contingent upon attention that properly dissociating the two becomes challenging. In the present paradigm, the spatial proximity of the task relevant fixation prohibited the separation of cross-modal attention and early multisensory integration. In future studies, the paradigm should dissociate the two by requiring a task in a third, irrelevant modality. Alternatively, one might systematically manipulate attention, as Talsma et al. (2007) did. This time, making sure that the auditory and visual components are unequivocally congruent in space.

While it is dangerous to base any conclusions on rough visual estimations of scalp topographies, the available research provides few other options. And the relevant human neural architecture is much less accessible than that of animal models. In the future the temporal precision of ERP's should be combined with the spatial precision of fMRI to investigate the anatomical underpinnings of early multisensory interactions. This might

be a crucial step towards uncovering the underlying components. If white noises actually interact with different regions of visual cortex than pure tones, these anatomical distinctions could be borne out with fMRI. Moreover, anatomical information alone will not provide the complete picture of early multisensory circuits. With ecologically valid AV stimuli, future research could expose much greater complexity in early cross-modal interactions than previously thought.

Generally, spectral characteristics, spatial congruence, and attention collectively determine the latency, power, and anatomical distribution of a given multisensory interaction. These determining properties must all dynamically interact with one another. Therefore, research in the future should aim to show how these properties interact, and how the brain makes sense of these interactions in order to adapt to its environment.

Conclusions

Initially, the interaction between unisensory areas within early time frames seems to challenge the convention that sensory processing is hierarchical. Previously, higher areas were thought to have privileged access to all modalities. The observation that unisensory areas can communicate, robs higher areas of this privilege. But as we gain more information about early multisensory interactions, a new type of hierarchy may come to the fore. Anatomical studies provide good evidence that there are far fewer cross-modal connections between any two unisensory areas than there are cross-modal connections between unisensory areas and higher level areas of the opposite modality. This new form of hierarchy makes intuitive sense. Too many projections between primary sensory areas could have grave consequences on the accurate representation of an environment. To be sure, the amount of recent studies showing connections between primary areas *is* very surprising. One author remarks on this, saying that such great potential exists for unisensory areas to communicate, it is a wonder how the different modalities don't routinely confuse one another (Calvert et al., 2004). The present results may begin to explain how different modalities mutually inform one another without corrupting each other. Specifically, the projections from auditory cortex to early visual areas are so sparse that the representation of a pure tone may only travel this pathway if it receives direct attention. In this sense, top-down mechanisms are still very much in control. These regulatory mechanisms prevent a maladaptive situation, where cross-modal influences go unchecked, potentially prohibiting the accurate representation of the world from any one modality.

Evidence for this type of organization in early multisensory interactions has been building in recent years. But this study reveals an entirely new type of organization. The present results suggest that early multisensory interactions are sensitive to the spectral characteristics of the auditory stimulus. Specifically, a broadband noise elicits an earlier, more occipital interaction, while a pure tone elicits a later, more occipito-parietal interaction. At this point, the major rationale is that white noise contains more information, and therefore influences visual processing at an earlier stage, and in more unisensory visual areas. By contrast, a pure tone contains less information, and therefore influences visual processing at a later stage, and in more multisensory areas. While more studies will be necessary to validate and refine these interpretations, they provide yet another dimension along which early multisensory processes may be systematically organized and regulated. Importantly, though, the pattern of results observed here is probably subject to all sorts of manipulations. If white noise was unattended, and pure tones were directly attended, the latencies could have switched positions, allowing pure tones to interact at 50ms.

As different types of multisensory organization accrue, scientists and theorists will observe interactions between them. Moreover, as multisensory phenomena accrue at different processing levels, the intricate links between low-, medium-, and high-level brain regions should begin to explain the ‘seamless’ integration of different modalities that dominates our experience. This thesis was introduced by the traditional notion that higher-level brain regions like the parietal cortex might be solely responsible for integrating the different sensor modalities. The realization that integration happens at all brain levels may deepen our understanding of just how ‘seamless’ experience really is.

Specifically, cross-modal information is not just cohered at one final perceptual stage, in which polished auditory and visual representations are simply fit together. It is more likely that auditory information is embedded in visual processing from the very beginning, albeit sparsely. While researchers must interpret their results in terms of adaptive behavior, one must also ask how such ubiquitous cross-modal integration colors our conscious experience of and intuitions about our environment? Moreover, it is possible that the unity of conscious experience is not a by-product of adaptive evolution, but is an adaptive trait in its own right. Multisensory integration may cohere the disjointed aspects of one's environment to a degree where they seem tailored to an organism. Therefore, the coherence of the senses provides obvious adaptive mechanisms, but may also provide less obvious adaptive mechanisms that manifest on a global scale. In sum, the multisensory nature of the brain might correlate with some of the foundations of experience. Furthermore, understanding the multisensory nature of experience should go hand in hand with an understanding of the global multisensory circuits of the brain.

Appendix A: Average ERP's from Six Electrodes Over Fronto-Central Scalp

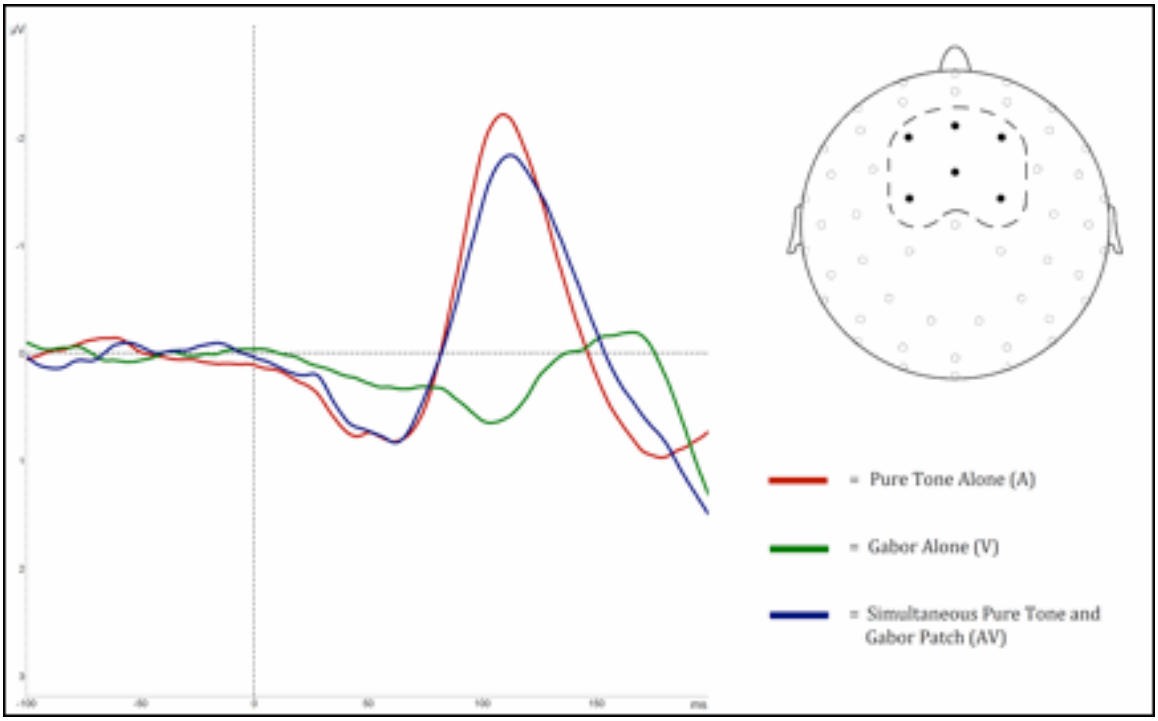

Figure 10. Grand Averaged ERPs to Pure Tone, Gabor Alone, and Simultaneous Tone - Gabor (AV) over fronto-central scalp.

The scalp shown in the upper right is a view from above. The dashed line surrounds six electrodes, which roughly contain the topographic peaks of early auditory components (P50 - blue arrow, N1 - red arrow). These six electrodes were averaged to create one representative ERP plot shown on the left hand side. Since the blue (AV) and red (A) lines are almost identical in the first 100 ms, one can imagine that adding the green line (V) to the red (A) would produce an (A+V) wave greater than the blue (AV) wave. If this trend, in which the sum ERP is greater than the simultaneous ERP, became more noticeable, there might be grounds for statistical analysis. This could represent very slight sub-additive interactions, or response depression, in which the simultaneous presentation of A and V stimuli actually reduces the P50 amplitude. Alternatively, it could represent a Contingent Negative Variation (CNV) which is a slow building negativity that forms in anticipation of a motor response.

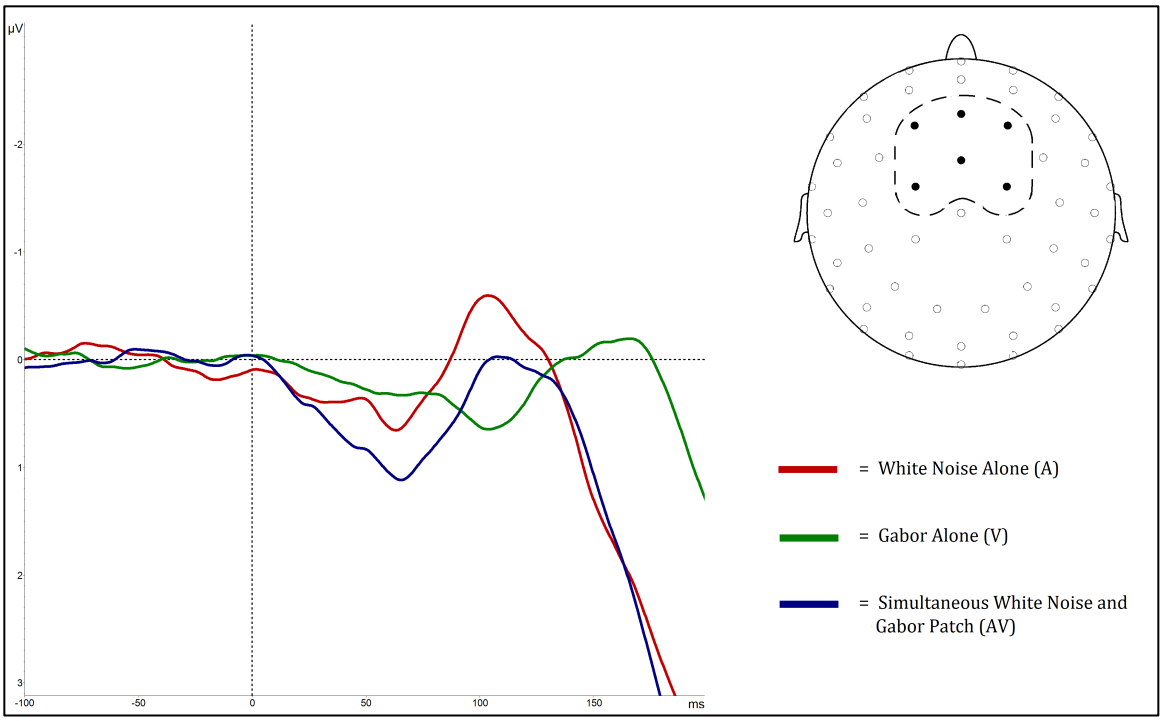

Figure 11. Grand Averaged ERPs to Pure Tone, Gabor Alone, and Simultaneous Tone - Gabor (AV) over fronto-central scalp.

This ERP plot is exactly the same as the previous one, except it represents white noise as opposed to pure tones. It is immediately obvious that the P50 component has a more distinct peak, and the N1 component (100 ms) is much less negative. While this cannot be specifically related to the differing neural representations of white noise and pure tones outlined in the text, it provides general evidence that spectral characteristics greatly influence the pattern of brain activity.

Bibliography

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (January 01, 1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annual Review of Neuroscience, 20, 303-30.

Anderson, J. S., Lopez-Larson, M., Yurgelun-Todd, D., & Ferguson, M. A. (November 16, 2010). Topographic maps of multisensory attention. Proceedings of the National Academy of Sciences of the United States of America, 107, 46, 20110-20114.

Beauchamp, M. S. (January 01, 2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Current Opinion in Neurobiology, 15, 2, 145-53.

Brockelmann, A.-K., Steinberg, C., Elling, L., Pantev, C., Junghofer, M., & Zwanzger, P. (May 25, 2011). Emotion-associated tones attract enhanced attention at early auditory processing: Magnetoencephalographic correlates. Journal of Neuroscience, 31, 21, 7801-7810.

Budinger, E., Heil, P., Hess, A., & Scheich, H. (December 28, 2006). Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. *Neuroscience, 143,* 4, 1065-1083.

Brungart, D. S. (January 01, 1999). Auditory localization of nearby sources. III. Stimulus effects. The Journal of the Acoustical Society of America, 106, 6, 3589-602.

Brungart, D. S., & Simpson, B. D. (January 01, 2009). Effects of bandwidth on auditory localization with a noise masker. The Journal of the Acoustical Society of America, 126, 6, 3199-208.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (January 01, 2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 51, 18751-6.

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

Cappe, C., Murray, M. M., Thut, G., & Romei, V. (September 22, 2010). Auditory-visual multisensory interactions in humans: Timing, topography, directionality, and sources. Journal of Neuroscience, 30, 38, 12572-12580.

Cappe, C., & Barone, P. (December 01, 2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. European Journal of Neuroscience, 22, 11, 2886-2902.

Carrasco, A., & Lomber, S. G. (January 01, 2011). Neuronal activation times to simple, complex, and natural sounds in cat primary and nonprimary auditory cortex. Journal of Neurophysiology, 106, 3, 1166-78.

Carriere, B. N., Royal, D. W., & Wallace, M. T. (May 01, 2008). Spatial heterogeneity of cortical receptive fields and its impact on multisensory interactions. Journal of Neurophysiology, 99, 5, 2357-2368.

Clavagnier, S., Falchier, A., & Kennedy, H. (January 01, 2004). Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. Cognitive, Affective, & Behavioral Neuroscience, 4, 2, 117-126.

Di, Russo. F., Martínez, A., & Hillyard, S. A. (May 01, 2003). Source Analysis of Event-related Cortical Activity during Visuo-spatial Attention. Cerebral Cortex, 13, 5, 486-499.

Di, Russo. F., Pitzalis, S., Spitoni, G., Aprile, T., Patria, F., Spinelli, D., & Hillyard, S. A. (January 01, 2005). Identification of the neural sources of the pattern-reversal VEP. Neuroimage, 24, 3, 874-86.

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

Eimer, M., & Van Velzen, J. (January 01, 2002). Crossmodal links in spatial attention are mediated by supramodal control processes: Evidence from event-related potentials.Psychophysiology, 39, 4, 437-449.

Eimer, M., & Schroger, E. (January 01, 1998). ERP effects of intermodal attention and cross-modal links in spatial attention. Psychophysiology, 35, 3, 313-327.

Ernst, M. O., & Banks, M. S. (January 01, 2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature, 415, 6870, 429.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (January 01, 2002). Anatomical evidence of multimodal integration in primate striate cortex. The Journal of Neuroscience : the Official Journal of the Society for Neuroscience, 22, 13, 5749-59.

Fastl, H., & Zwicker, E. (2007). Psychoacoustics: Facts and models. Berlin: Springer.

Ffytche, D. H., Guy, C. N., & Zeki, S. (January 01, 1995). The parallel visual motion inputs into areas V1 and V5 of human cerebral cortex. Brain : a Journal of Neurology,118, 1375-94.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (January 01, 2002). Early auditory-visual interactions in human cortex during nonredundant target identification. Brain Research. Cognitive Brain Research, 14, 1, 20-30.

Foxe, J. J., & Simpson, G. V. (January 01, 2002). Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale, 142, 1, 139-50.

Giard, M. H., & Peronnet, F. (January 01, 1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. Journal of Cognitive Neuroscience, 11, 5, 473-90.

Gondan, M., Niederhaus, B., Rösler, F., & Röder, B. (January 01, 2005). Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. Perception & Psychophysics, 67, 4, 713-26.

Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (January 01, 2002). Spectral and temporal processing in human auditory cortex. Cerebral Cortex (new York, N.y. : 1991), 12, 2, 140-9.

Hillyard, S. A., Teder-Sälejärvi, W. A., & Münte, T. F. (January 01, 1998). Temporal dynamics of early perceptual processing. Current Opinion in Neurobiology, 8, 2, 202-10.

Inui, K., Okamoto, H., Miki, K., Gunji, A., & Kakigi, R. (January 01, 2006). Serial and Parallel Processing in the Human Auditory Cortex: A Magnetoencephalographic Study. Cerebral Cortex, 16, 1, 18-30.

Kaas, J. H., & Hackett, T. A. (January 01, 1999). 'What' and 'where' processing in auditory cortex. Nature Neuroscience, 2, 12, 1045-7.

Kelly, S. P., Gomez-Ramirez, M., & Foxe, J. J. (January 01, 2008). Spatial Attention Modulates Initial Afferent Activity in Human Primary Visual Cortex. Cerebral Cortex, 18, 11, 2629-2636.

Lang, A. (January 01, 2000). The limited capacity model of mediated message processing. Journal of Communication, 50, 1, 46-70.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Mass: MIT Press.

Macaluso, E., Frith, C. D., & Driver, J. (January 01, 2000). Modulation of human visual cortex by crossmodal spatial attention. Science (new York, N.y.), 289, 5482, 1206-8.

Martínez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., Wong, E. C., ... Hillyard, S. A. (January 01, 1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. Nature Neuroscience, 2, 4, 364-9.

McDonald, J. J., Teder-Salejarvi, W. A., & Ward, L. M. (January 01, 2001). Multisensory integration and crossmodal attention effects in the human brain. Science (new York, N.y.), 292, 5523.)

McGurk, H., & MacDonald, J. (January 01, 1976). Hearing lips and seeing voices. Nature, 264, 5588, 23-30.

Meredith, M. A., & Stein, B. E. (January 01, 1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. Journal of Neurophysiology, 75, 5, 1843-57.

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (January 01, 2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. The Journal of Neuroscience : the Official Journal of the Society for Neuroscience, 27, 15, 4120-31.

Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., Dyke, J. P., ... Foxe, J. J. (January 01, 2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. Journal of Neurophysiology, 96, 2, 721-9.

Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., Dyke, J. P., ... Foxe, J. J. (January 01, 2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. Journal of Neurophysiology, 96, 2, 721-9.

Morgan, M. L., DeAngelis, G. C., & Angelaki, D. E. (August 28, 2008). Multisensory Integration in Macaque Visual Cortex Depends on Cue Reliability. Neuron, 59, 4, 662-673.

Pluta, S. R., Rowland, B. A., Stanford, T. R., & Stein, B. E. (January 01, 2011). Alterations to multisensory and unisensory integration by stimulus competition. Journal of Neurophysiology, 106, 6, 3091-101.

Pooresmaeili, A., Herrero, J. L., Self, M. W., Roelfsema, P. R., & Thiele, A. (January 01, 2010). Suppressive lateral interactions at parafoveal representations in primary visual cortex. The Journal of Neuroscience : the Official Journal of the Society for Neuroscience, 30, 38, 12745-58.

Rauschecker, J. P., & Tian, B. (October 24, 2000). Mechanisms and Streams for Processing of "What" and "Where" in Auditory Cortex. Proceedings of the

National Academy of Sciences of the United States of America, 97, 22, 11800-11806.

Rockland, K. S., & Ojima, H. (January 01, 2003). Multisensory convergence in calcarine visual areas in macaque monkey. International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology, 50, 1-2.

Rohl, M., Kollmeier, B., & Uppenkamp, S. (January 01, 2011). Spectral loudness summation takes place in the primary auditory cortex. Human Brain Mapping, 32, 9, 1483-96.

Rossi, V., & Pourtois, G. (May 01, 2012). State-dependent attention modulation of human primary visual cortex: A high density ERP study. Neuroimage, 60, 4, 2365-2378.

Saint-Amour, D., De, S. P., Molholm, S., Ritter, W., Foxe, J. J., & Advances in Multisensory Processes. (January 01, 2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. Neuropsychologia, 45, 3, 587-597.

Senkowski, D., Saint-Amour, D., Hofle, M., & Foxe, J. J. (June 15, 2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. Neuroimage, 56, 4, 2200-2208.

Smith, D. A., Boutros, N. N., & Schwarzkopf, S. B. (January 01, 1994). Reliability of P50 auditory event-related potential indices of sensory gating. Psychophysiology, 31, 5, 495-502.

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

Stevenson, R. A., & James, T. W. (February 01, 2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. Neuroimage, 44, 3, 1210-1223.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (September 01, 2010). The multifaceted interplay between attention and multisensory integration. Trends in Cognitive Sciences, 14, 9, 400-410.

Talsma, D., Doty, T. J., & Woldorff, M. G. (January 01, 2007). Selective Attention and Audiovisual Integration: Is Attending to Both Modalities a Prerequisite for Early Integration?. Cerebral Cortex, 17, 3, 679-690.

Upadhyay, J., Ducros, M., Knaus, T. A., Lindgren, K. A., Silver, A., Tager-Flusberg, H., & Kim, D. S. (January 01, 2007). Function and connectivity in human primary auditory cortex: a combined fMRI and DTI study at 3 Tesla. Cerebral Cortex (new York, N.y. : 1991), 17, 10, 2420-32.

Vidal, J., Roux, S., Barthelemy, C., Bruneau, N., & Giard, M.-H. (April 01, 2008). Cross-modal processing of auditory-visual stimuli in a no-task paradigm: A topographic event-related potential study. Clinical Neurophysiology, 119, 4, 763-771.

Werner, S., & Noppeney, U. (February 17, 2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. Journal of Neuroscience, 30, 7, 2662-2675.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (September 15, 1993). Modulation of Early Sensory Processing in Human Auditory Cortex During Auditory Selective Attention. Proceedings of the National Academy of Sciences of the United States of America,90, 18, 15.

Yost, W. A., Popper, A. N., & Fay, R. R. (1993). Human psychophysics. New York: Springer-Verlag

Zhang, W., & Luck, S. J. (January 01, 2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience, 12,* 1, 24-25.