

Neural Integration of Multimodal Events

by

Jean M. Vettel

B.A., Carnegie Mellon University, 2000

M.Sc., Brown University, 2008

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE
DEPARTMENT OF COGNITIVE AND LINGUISTIC SCIENCES AT BROWN
UNIVERSITY

PROVIDENCE, RHODE ISLAND

MAY 2010

© Copyright 2010 by Jean Marie Vettel

**This dissertation by Jean M. Vettel is accepted in its present form
by the Department of Cognitive and Linguistic Sciences as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.**

Date: 7/27/2009

Michael J. Tarr, Advisor

Recommended to the Graduate Council

Date: 7/27/2009

Sheila Blumstein, Reader

Date: 7/27/2009

Laurie Heller, Reader

Date: 7/27/2009

David Sheinberg, Reader

Approved by the Graduate Council

Date: _____

Sheila Bonde, Dean of Graduate School

Jean Marie Vettel

jvettel@gmail.com

<http://www.vettel.org/jean>

Neuroscientist, Army Research Labs, Aberdeen, MD

Academic History

- **Brown University, Providence, RI**
 - *M.S. Cognitive Science, May 2008
 - *Ph.D. Cognitive Science, August 2009
 - *Committee: Michael Tarr (chair)
Sheila Blumstein, Laurie Heller, & David Sheinberg (readers)
- **California University of Pennsylvania, California, PA**
 - *M.S. Exercise Science, Performance Enhancement & Injury Prevention, May 2008
 - *online degree program
- **Carnegie Mellon University, Pittsburgh, PA**
 - *B.A. in Professional Writing, Design concentration, May 2000
 - *University Honors, Phi Beta Kappa, Phi Kappa Phi,
Dean's List High Honors (all semesters)

Honors, Fellowships, and Awards

- **SMART (Science, Mathematics, And Research for Transformation) Fellowship**
Department of Defense (Sept 2007 – graduation)
- **Graduate Student Fellowship**
NASA-Rhode Island Space Grant Consortium (declined due to SMART Fellowship)
- **Graduate Research Fellowship**
National Science Foundation (Sept 2004 – Aug 2007)
- **Funded Attendee, Mathematics in Brain Imaging**
Institute for Pure & Applied Mathematics, UCLA (July 14-28, 2008)
- **Fifth Year Scholar Research Fellowship**
Carnegie Mellon University (Aug 1999 – May 2001)
- **Carol M. Wilson Scholarship**
Kappa Alpha Theta (Aug 1998 – May 1999)

Professional History

- **Laboratory Manager**
Dynamic Cognition Lab (J. Zacks), Washington University in St. Louis
(Sept 2001 – June 2004)

Publications

- Vettel, J.M. (in press). Receptive Fields. In Kreutzer, J.S., DeLuca, J., & Caplan, B. (Eds.) *Encyclopedia of Clinical Neuropsychology*.
- Righi, G. & Vettel, J.M. (in press). Dorsal Visual Pathway. In Kreutzer, J.S., DeLuca, J., & Caplan, B. (Eds.) *Encyclopedia of Clinical Neuropsychology*.
- Righi, G. & Vettel, J.M. (in press). Ventral Visual Pathway. In Kreutzer, J.S., DeLuca, J., & Caplan, B. (Eds.) *Encyclopedia of Clinical Neuropsychology*.
- Nestor, A., Vettel, J.M., & Tarr, M.J. (2008). Task-Specific Codes for Face Recognition: How they Shape the Neural Representation of Features for Detection and Individuation. *PLoS ONE*, 3(12), e3978 doi:10.1371/journal.pone.0003978
- Zacks, J.M., Speer, N.K., Vettel, J.M., & Jacoby, L.J. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging*, 21, 466-482.
- Zacks, J.M., Swallow, K.M., Vettel, J.M., & McAvoy, M.P. (2006). Visual motion and the neural correlates of event perception. *Brain Research*, 1076, 150-162.
- Michelon, P., Vettel, J.M., & Zacks, J.M. (2006). Lateral somatotopic organization during imagined and prepared movements. *Journal of Neurophysiology*, 95, 811-822.
- Zacks, J.M., Michelon, P., Vettel, J.M., & Ojemann, J.G. (2004). Functional Reorganization of Spatial Transformations After a Parietal Lesion. *Neurology*, 63, 287-292.
- Zacks, J.M., Vettel, J.M., Michelon, P. (2003). Imagined Viewer and Object Rotations Dissociated with Event-Related fMRI. *Journal of Cognitive Neuroscience*, 15(7), 1002-1018.

Manuscripts in Preparation

- Vettel, J.M., Green, J., Heller, L., & Tarr, M.J. (submitted). Temporal and Semantic Effects on Multimodal Integration.
- Righi, G., Vettel, J.M., Lebrecht, S., & Tarr, M.J. (in preparation). Individual differences in face processing: establishing a link between brain and behavior.
- Peissig, J.J., Vuong, Q.C., Vettel, J.M., & Tarr, M.J. (in preparation). Contrast Reversal Effects on Recognition of Common Objects.

Invited Talks

- Unraveling the organizational principles of perceptual cortex. Talk presented at the Human Research and Engineering Directorate, Army Research Labs, Aberdeen, MD. January 2008.

Presentations at National Meetings

- Vettel, J.M., Nestor, A., Bird, C.W., Heller, L.M., Curran T., & Tarr, M. J. (2009). Investigating the Interplay of Time & Semantics during Multimodal Integration. Poster presented at the 10th International Multimodal Research Forum, New York, NY.
- Lebrecht, S., Righi, G., Vettel, J. M., & Tarr, M. J. (2008). Predicting individual differences in cortical activations using behavioral measures. Poster presented at the Society for Neuroscience Meeting, Washington, DC.
- Vettel, J.M., Green, J., Heller, L., & Tarr, M.J. (2008). The Neural Representation of Dynamic Real-World Auditory/Visual Events. Poster presented at the Annual Meeting of the Vision Sciences Society, Naples, FL.
- Nestor, A., Vettel, J.M., & Tarr, M.J. (2008) Task-specific feature codes for face processing. Presentation given by A. Nestor at the Annual Meeting of the Vision Sciences Society, Naples, FL.
- Vettel, J.M., Barenholtz, E., & Tarr, M.J. (2006). A dynamic cue for figure ground assignment: Advancing vs. receding. Poster presented at the Annual Meeting of the Vision Sciences Society, Sarasota, FL.
- Peissig, J.J., Vuong, Q.C., Vettel, J.M., & Tarr, M.J. (2006). Does Contrast Reversal Affect the Recognition of Common Objects? Poster presented at the Annual Meeting of the Vision Sciences Society, Sarasota, FL.
- Zacks, J.M., Swallow, K.M., Vettel, J.M., & McAvoy, M.P. (2004). The Role of Movement in the Neural Processing of Simple Events. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society, San Francisco, CA.
- Zacks, J.M., Michelon, P., Vettel, J.M., & Ojemann, J.G. (2004). Functional Reorganization of Spatial Transformations After a Parietal Lesion. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society, San Francisco, CA.
- Zacks, J.M., Vettel, J.M., & Michelon, P. (2003). Imagined Viewer and Object Rotations Dissociated with Event-Related fMRI. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society, New York, NY.

Presentations at Regional Meetings

- Vettel, J.M., Green, J., Heller, L., & Tarr, M.J. (2008). Temporal Dynamics of Multimodal Integration. Poster presented at Temporal Dynamics of Learning Center NSF Site Visit in April, San Diego, CA.
- Vettel, J.M., & Tarr, M.J. (2006). How Do We Acquire Multimodal Expertise? Talk presented at the Spring Meeting of the Perceptual Expertise Network.
- Vettel, J.M., Tarr, M.J., & Pessoa, L. (2005). Signal detection methods for fMRI data analysis. Talk presented at the Spring Meeting of the Perceptual Expertise Network.

Service

- **Organizer: Vision Journal Club**
Brain Science Program, Brown University (June 2005-May 2007)
Duties: recruited presenters, maintained website & email listserv, coordinated biweekly meetings
- **Organizer: Cognitive Neuroscience Journal Club**
Brain Science Program, Brown University (June 2007-June 2008)
Duties: recruited presenters, maintained website & email listserv, coordinated biweekly meetings
- **Coordinator: Prospective Graduate Student Visit (2 day)**
Cognitive & Linguistic Sciences Department, Brown University (2005-present)
Duties: coordinate housing & transportation for visiting prospective students, organize lunches and dinners with current graduate students, design documents with visit information, etc.
- **Planning Committee: Perceptual Expertise Network Meeting XV (Oct 4-5, 2007)**
Annual meeting of a collaborative research network
(<http://gauthier.psy.vanderbilt.edu/pen>)
Duties: solicited talks from the network members, organized schedule of talks, gave opening presentation to begin the meeting
- **Department Talent Show Videographer**
Cognitive & Linguistic Sciences Department, Brown University (2 shows, 06 & 07)
Duties: filmed each act in the show, edited the acts on a DVD for distribution
- **Co-organizer: Biweekly Departmental Social Events**
Cognitive & Linguistic Sciences Department, Brown University (04-05)
Duties: organize themed snacks and drinks for the socials, collect monies

Teaching Experience

- **Sheridan Center Teaching Certificate**
Brown University (awarded May 2006)
- **Sheridan Center Teaching Consultant**
Brown University (Fall 2007 – present)
- **Teaching Assistant**
Introduction to Children's Thinking, Brown University (Spring 2007 & Spring 2008)
- **Teaching Assistant**
Introduction to Cognitive Science, Brown University (Fall 2005 & Fall 2007)
- **Teaching Assistant**
Introduction to Computer Programming, Carnegie Mellon (Fall 2000 – Summer 2001)

Acknowledgments

My advisor, Mike Tarr, taught me in my first year that good science is about taking risks: when I fretted, “what if this experiment doesn’t work,” he quickly and emphatically rejoined, “but if it does, it is a nature paper!” Throughout my five years as a tarlabite, Mike has always provided a nurturing and supportive environment, full of wild, theoretical speculation (aka “mike’s little theories”) as well as crazy, tangential stories (his co-authored novel about jews with elan, computational models of sports/weather, the infamous iMac fish bowl, etc). Most importantly, Mike has convinced me that all good lab meetings must include a healthy mixture of science and laughter (well that...and lots of cake). I will continue to seek out his advice for years to come.

I could not have asked for a more supportive dissertation committee. Laurie Heller offered expert advice about the chosen stimulus set, environmental events, and fostered a love of waveforms; Sheila Blumstein educated me on the complicated role of the inferior frontal gyrus and its role in domain-general competition; and David Sheinberg instilled in me a desire to embrace theoretical speculation (including that found here in Chapter 5, which David rightfully termed “Jean’s wacky theory”).

The third dissertation experiment was my first adventure in the land of EEG. Tim Curran not only provided an expert road map to guide me but he also enlisted his army of researchers to efficiently collect the entire dataset of 31 participants in 3 weeks flat. Chris Bird was the skillful leader of this Curran lab army, and he spent many weeks in the fall semester battling the Eprime code so it would show my event movies. They had a joke in the lab for a bit, asking Chris “so does Jean have a dissertation yet?” Many thanks to the Curran lab for (1) not sharing this joke with me until AFTER the full dataset was

collected and (2) making my dreams for studying the timecourse of multimodal integration a reality. Their location in a cycling mecca was also appreciated...

Many people in the CLS department supported me throughout this arduous process, sometimes with scientific discussion and other times with relaxing chatter. In particular, my fellow tarrlabites top the list. The tarrlab functions as a cohesive unit, always ready to organize, improvise, or entertain. I reached the finish line because the friendship and support of Jessie Peissig (and Nigel Gomez), Giulia Righi, Adrian Nestor, Sophie Lebrecht, Elan Barenholtz, Julia Green, Ashley Kingon, Ana Van Gulick, and Heida Sigurdardottir. Outside of the lab, special thanks to Emily Myers, Elena Tenenbaum, Reinette Sousa, and Rosa Moniz.

There are so many people outside of the department that helped to keep me sane. As I have since childhood, I heavily relied on the steady, strong support from my family. Some of them generously went to work helping with the actual dissertation (Mom & Aunt Nancy as my camera crew, Dad as FileMaker script writer and Photoshop editor), while most of them provided lovely emails/texts, FaceBook posts, or snail mail cards (packed with \$20s from Aunt Sandy and Uncle Erwin). My best friend, Christine Hoyer, always lend her ear for my long-winded, stress-releasing, tangent-friendly chatter fests – often providing a reassuring and believable “you can do it” when the deadline oppression felt unbearable. My friends at the gym (Healthtrax & Bally) and on the road (both cycling and running) are too numerous to list all, but they each deserve a page of their own for proper expression of my thanks (especially Cathy, Gil, Kevan, Karen, Rose, Dean, Mo, Vivian, Mike, Erin). Finally, my last acknowledgments are to my two beloved companions, Cagney Vettel (92-07) and Adrian Nestor (now & for all my years to come)!

Table of Contents

1. Chapter 1: Introduction	1
1.1 Multimodal Interactions: Behavioral	2
1.2 Multimodal Interactions: Neural	6
1.3 Neural Architecture	12
1.4 Multimodal Interactions: Open Questions	15
1.5 Three Multimodal Studies	18
2. Chapter 2: Temporal and Semantic Effects on Multimodal Integration	20
2.1 Introduction	20
2.2 Methods & Materials	22
2.3 Results	27
2.4 Discussion	40
2.5 Conclusion	46
3. Chapter 3: Interplay of Temporal & Semantic Congruency in Multimodal Events	47
3.1 Introduction	47
3.2 Methods & Materials	51
3.3 Results	59
3.4 Discussion	72
3.5 Conclusion	79
4. Chapter 4: Investigating the Timecourse of Multimodal Integration	80
4.1 Introduction	80
4.2 Methods & Materials	86
4.3 Results	93

4.4 Discussion	102
4.5 Conclusion	108
5. Chapter 5: Discussion	109
5.1 Experiment 1	109
5.2 Experiment 2	112
5.3 Experiment 3	115
5.4 Conclusion	118
6. References	120

List of Tables

Table 1 ROI Location, Size, and Peak Coordinates for all AV Contrasts (Exp. 1)	30
Table 2 Whole-brain Analysis: Anatomical Location, Size, and Peak Coordinates (Exp. 2)	64

List of Figures

Figure 1 <i>Proposed Neural Architectures</i>	13
Figure 2 <i>Experimental Design (Exp. 1)</i>	23
Figure 3 <i>Functionally-defined Integration ROI and Overlap of AV Contrast Regions</i> ...	28
Figure 4 <i>AV Contrast: Temporal Processing</i>	32
Figure 5 <i>AV Contrast: Semantic Processing</i>	34
Figure 6 <i>AV Contrast: Congruency Processing</i>	37
Figure 7 <i>Stimulus Design (Exp. 2)</i>	53
Figure 8 <i>Region-of-Interest Analysis Method</i>	59
Figure 9 <i>Region-of-Interest Analysis: Frontal Cortex</i>	61
Figure 10 <i>Region-of-Interest Analysis: Posterior Cortex</i>	63
Figure 11 <i>Nine Conditions and Target Examples (Exp. 3)</i>	87
Figure 12 <i>Early Sensory Evoked Potentials</i>	95
Figure 13 <i>The N400 response for Impacts 1 & 2</i>	98
Figure 14 <i>Spatio-temporal Analysis of Semantic and Temporal Effects</i>	101

1. Chapter 1: Introduction

Our actions produce physical events that give rise to perceptual information across multiple modalities. A trip to the beach can be identified through all five of our senses: seeing the waves roll in from the horizon, hearing the waves break on the sand, smelling the ocean life, feeling the cool sensation of the water on our skin, and tasting the salty water. Yet information from one perceptual modality is not experienced separately from information in another modality; instead, our perception of the ocean blends the modality-specific information from our senses into an integrated, coherent percept. In a series of three experiments, we address how such percepts come to be; that is, what cues in the signal are used to combine different sensory information and, concurrently, what are the neural processing mechanisms sensitive to these integration cues.

Multimodal interactions have been found in behavioral and neural research. In both cases, multimodal stimuli generate responses that are not the simple summation of their unimodal counterparts. This finding has led researchers to examine what stimulus properties enable the efficient integration of information across modalities, and this research has identified three integration cues – time, space, and semantics – that influence the process of integration. In addition, research with both animals and humans reveal a consistent network of brain regions in posterior, temporal, and frontal cortices recruited during multimodal processing.

However, the functional role of the regions within the multimodal network is relatively unknown. Consequently, our studies independently manipulate temporal and semantic congruency using a novel stimulus set of environmental events. The studies ask

whether the two integration cues share the same neural resources or recruit separable neural substrates. We also explore how these two integration cues interact with one another, and in particular, whether the time course of one cue differs from the other. That is, does a low-level cue like temporal synchrony that is carried in the perceptual signals themselves influence multimodal processing earlier than a higher-order cue like semantic congruency that relies on stored, experiential knowledge. In short, the series of three studies aim to unravel the neural basis of multimodal integration.

1.1 Multimodal Interactions: Behavioral

Behavioral research has highlighted three major characteristics of multimodal interactions. First, the response times for detection or identification of a stimulus are faster for multimodal stimuli than for unimodal stimuli. Second, information from one modality can bias another modality, resulting in perceptual illusions that highlight the role of multimodal integration during perception. Third, amodal stimulus properties, such as spatial location and temporal structure, are defined without reference to a particular modality, and they have been proposed as the stimulus parameters that guide whether two or more modality-specific signals are perceived as a single, coherent event.

Faster Reaction Times

Decades of behavioral research have identified several advantages for a multimodal stimulus compared to a unimodal stimulus. One advantage is the ability to identify a stimulus more accurately and quickly in a multimodal situation (Hershenson, 1962; Ulrich et al., 2007). Intuitively, having redundant information about an event in

multiple modalities should improve perceptual accuracy since there is more information available to detect and/or identify the event, and this intuition has been formalized in the “redundancy signals effect” that states that multiple stimuli with redundant information will decrease the amount of time needed to make a response in comparison to a single stimulus (Miller, 1982).

Multimodal research has explored whether convergent information from two modalities leads to a faster response time than what would be expected by the response to either of the modalities in isolation. This is formalized in the race model. According to this model, each modality competes independently from one another to serve as the only source for the behavioral response. If the response time is faster than what is predicted by the race model predictions, researchers have argued that this provides evidence for interactions of the two modalities at the neural processing level since the reaction time is faster than the processing time needed for the unimodal response. Violations of the race model predictions have been found in numerous behavioral paradigms (for a review, see Cinel 1998), indicating that multimodal processing is not just the sum of its unimodal counterparts.

Intersensory Bias

Research exploring how information is integrated across modalities has often shown conflicting information between two modalities to investigate the effect on the final percept. The information in each modality could be weighted equally in the final percept or one modality may dominate and supercede the information in the other modality. Typically, one modality dominates, resulting in intersensory bias, and what

modality will dominate is predicted by the modality specificity hypothesis (Welch and Warren, 1980). That is, the dominant modality provides the more accurate information about the behavioral judgment of interest. Visual domain provides more precise information about spatial location and dominates spatial judgments, while the auditory domain has better temporal precision and dominates in temporal judgments. In fact, the amount of weight given to each modality has been shown to be statistically optimal, weighing each modality according to its reliability (Ernst and Banks, 2002). Thus, multimodal processing combines modalities in order to optimize the reliability of perceptual judgments.

Occasionally, intersensory bias leads to perceptual illusions. The two most famous crossmodal illusions are the ventriloquist effect and the McGurk effect. The ventriloquist effect (Howard and Templeton, 1966) reveals a bias to perceive spatial congruence, as a location mismatch between an auditory stimulus and a visual stimulus is incorrectly perceived as both emanating from the location of the visual stimulus. The McGurk effect (McGurk and MacDonald, 1976; van Wassenhove et al., 2005) reveals the strong perceptual influence of temporal synchrony, as lips speaking one phoneme, i.e., “ga,” in temporal synchrony with the sound of another phoneme, i.e., “da,” results in an illusory percept of a third phoneme, “ba.” Interestingly, in both cases, knowledge of the illusion is unable to disrupt its perception. The inability of higher-order knowledge to disrupt these illusions suggests that spatial and temporal congruence may be fundamental to the way in which the low-level information is combined across modalities in the brain. Subsequent research has sought to identify cues, such as the spatial and temporal constraints driving these illusions, that can influence the integration process.

Amodal Stimulus Properties

Behavioral research has identified stimulus properties that can be defined without reference to a particular sensory modality, called “amodal” properties, and these properties have been shown to influence whether multimodal integration occurs in adults (Welch, 1999). As suggested by the ventriloquist and McGurk illusions, these properties include spatial location and temporal structure, but behavioral evidence in adults shows that size, shape, orientation, intensity, motion, and texture also influence the likelihood of multimodal integration (for review, see (Welch and Warren, 1980)). The more amodal properties shared between modalities, the more likely the modality-specific information arose from the same event.

Interestingly, research on infants has shown that they are also able to detect amodal properties of auditory-visual stimuli, and thus, these amodal properties may be fundamental to the way that our perceptual system develops multimodal capabilities. In one experiment with 4-month old infants (Spelke, 1976), movies were generated of simple percussion sequences using wooden blocks, tambourines, and batons. During the experiment, two videos were shown side-by-side and an audio track played from a centralized location that was temporally congruent with only one of the videos. Infants looked longer at the video that was temporally synchronous with the audio track, suggesting that 4-month olds are able to detect amodal stimulus properties that indicate a common cause for multimodal information. These results and others have been formalized in the Intersensory Redundancy Hypothesis, a set of principles that predict what stimulus properties are salient in different sensory situations, including the

prediction that infants are most sensitive to amodal properties in multimodal situations (Bahrick and Lickliter, 2002). This infant research suggests that amodal properties may provide the basis for the development of multimodal integration as well as subserve the process of integration throughout the lifespan (Lewkowicz, 1999).

Behavioral Interactions: Summary

Multimodal stimuli confer a behavioral advantage over unimodal stimuli in regard to both identification accuracy and faster reaction times. This advantage is greater than what is expected by merely summing the unimodal information, and these multimodal interactions are confirmed by showing violations of the race model predictions. Furthermore, modalities can bias one another in perceptual judgments, including biases that generate robust illusions (e.g., ventriloquist effect, McGurk effect) that persist despite higher order knowledge that they are, in fact, illusions. This persistence suggests that some stimulus properties may be fundamental to how unimodal information is integrated neurally to form coherent percepts, and research on both infants and adults have identified several amodal properties that are thought to guide multimodal interactions, both during development and across the lifespan. Given this strong behavioral evidence for multimodal interactions, we turn to research on the brain that aims to elucidate the neural substrates that subserve multimodal integration.

1.2 Multimodal Interactions: Neural

The initial evidence for multimodal interactions at the neural level was found in the superior colliculus of cats (Stein and Meredith, 1993). These neural interactions

share the same, fundamental characteristic identified by the race model for behavioral interactions; that is, the neural response to the multimodal stimulus is greater than that predicted by the linear sum of the neuron's response to the unimodal stimuli, showing a "superadditivity" response.

Following this discovery of neural multimodal interactions, subsequent research investigated what stimulus properties influenced these enhanced neural responses. Interactions were consistently identified when the modalities shared spatial congruence, temporal synchrony, and semantic congruence. The robust influence of these three cues to integration have been shown in both animals and humans using a variety of methods (neurophysiology, fMRI, and/or EEG).

Multimodal Interactions

The seminal work by Stein & Meredith (1993) in the superior colliculus of cats identified neurons that demonstrated a multimodal interaction where the neural response for a multimodal stimulus could not be predicted by the response to either unimodal stimuli. In some cases, the response was "superadditive" where the neuron fired more for the multimodal stimulus than the linear sum of the two unimodal responses; in others, the response was "subadditive" where the neuron fired less than the weakest of the two unimodal firing rates. These multimodal interactions were evidenced by every measure they investigated: reliability that the response was greater for the multimodal stimulus than either unimodal stimuli, the neural firing rate was higher, the number of peak responses within the response window was greater, and the duration of the response period itself was longer (Stein and Meredith, 1993). Next, they began investigating the

conditions under which this multimodal interaction occurred to identify what factors served as cues to integrate information between modalities.

Spatial Congruence & Temporal Synchrony

Multimodal interactions depended on the spatial and temporal relationship of the unimodal stimuli. Stimuli that were spatially coincident typically resulted in a superadditive response, while stimuli in different spatial locations often produced a subadditive response or no interaction at all. Likewise, stimuli that were temporally coincident led to superadditive responses, whereas temporally asynchronous events typically caused a subadditive response or no effect at all (Stein and Meredith, 1993). A multimodal interaction that is dependent on temporal and/or spatial relationships reflects the dynamics of real-world events; that is, modality-specific information that arises from the same event will originate at the same time and from the same location. Thus, the superadditive and subadditive responses provide a neural mechanism to bind information across modalities that are likely to share a common cause.

Subsequent neurophysiology research has investigated multimodal interactions in brain regions other than the superior colliculus in several animal models. Multimodal interactions dependent on temporal and spatial relationships between modalities have been heavily studied in the primary auditory cortex of both ferrets and monkeys (Bizley et al., 2007; Brosch et al., 2005; Ghazanfar et al., 2005; Kayser et al., 2005; Kayser et al., 2007; Lakatos et al., 2007; Schroeder and Foxe, 2002), but research has also found interactions in primary visual cortex in both ferrets and monkeys (Allman et al., 2008; Kayser and Logothetis, 2007; Schroeder and Foxe, 2005), temporal cortex of

monkeys (Barraclough et al., 2005), somatosensory cortex of monkeys (Zhou and Fuster, 2000), and frontal cortex of monkeys (Fuster et al., 2000; Sugihara et al., 2006). Thus, the neural mechanisms underlying multimodal interactions are distributed throughout the cortex.

Research in humans has also investigated evidence for multimodal interactions. Using functional neuroimaging, studies have found regions dependent on the temporal and spatial relationships of the stimuli across modalities in the same areas as the neurophysiological research on animals, including the primary visual and auditory areas, temporal cortex, frontal cortex, and parietal cortex (Bischoff et al., 2007; Bushara et al., 2001; Bushara et al., 2003; Dhamala et al., 2007; Miller and D'Esposito, 2005; Noesselt et al., 2007; Sestieri et al., 2006). In fact, recent reviews of the animal and human findings have found convergent results for the areas involved across species and methods (Driver and Noesselt, 2008; Kayser and Logothetis, 2007).

These findings have been supplemented by research on humans using electroencephalography (EEG) where multimodal effects using auditory-visual stimuli are seen on the earliest event related potentials related to perceiving an auditory stimulus (Besle et al., 2004; Klucharev et al., 2003; Lebib et al., 2003; Lebib et al., 2004; Stekelenburg and Vroomen, 2007; van Wassenhove et al., 2005) as well as the earliest event related potentials related to perceiving an visual stimulus (Fort et al., 2002; Giard and Peronnet, 1999; Molholm et al., 2004; Yuval-Greenberg and Deouell, 2007). Since these event related potentials are the earliest cortical response to a unimodal stimulus, occurring within 100 milliseconds of stimulus onset, the modulation of these

potentials by a multimodal stimulus suggests a very early interaction of modality-specific processing.

Converging evidence from both animals and humans using multiple methods indicates that the temporal and spatial relationships are strong cues for integration of modality specific information. Both integrative cues are carried in the signal when it arrives at the senses. Thus, temporal and spatial relationships can be thought of as low-level cues that are available early based on how the primary sensory pathways have been organized to process incoming sensory input. However, research has shown that time and space are not the only two cues relevant for multimodal integration.

Semantic Congruence

Among neurophysiological, functional imaging, and EEG studies, the semantic relationship between modality-specific information has also been shown to modulate the multimodal interaction effect in the neural response (for a review, see (Doehrmann and Naumer, 2008)). The multimodal literature uses the term semantic when investigating how stimulus content influences processing, typically manipulating whether the identity of a picture matches the identity of an accompanying sound. A semantically congruent condition may show a picture of a dog with the sound of a barking dog, whereas a semantically incongruent condition pairs a picture of a dog with the sound of a meowing cat. Semantic associations are learned through our experience with the world, allowing these associations to also influence whether information from different modalities is inferred to share a common cause, and consequently, whether they should be bound into a unified percept.

Only a few neurophysiological studies in monkeys have manipulated semantic congruency between modalities. Multimodal interaction effects have been found in monkey temporal cortex (Barraclough et al., 2005; Morgan et al., 2008) and monkey frontal cortex (Sugihara et al., 2006) where semantically congruent stimuli typically lead to a superadditive response. Human functional imaging studies have also identified regions modulated by semantic congruency in frontal and temporal cortices (Belardinelli et al., 2004; Hein et al., 2007; Hocking and Price, 2008; Laurienti et al., 2003; Taylor et al., 2006). Here, the common finding is greater activity for semantically incongruent conditions in frontal cortex, while the neural response in temporal cortex is equivalent for semantically congruent and incongruent conditions (though see (Taylor et al., 2006) for sensitivity to semantics in perihinal cortex). In addition, semantic congruency has also been studied with EEG. These studies have identified an event related potential that occurs about 400 msec after the semantically incongruent stimulus pairing (for a review, see (Van Petten and Luka, 2006)), revealing a marker of the semantic relationship between modalities during the processing of events (Cummings et al., 2006; Orgs et al., 2008; Sitnikova et al., 2008).

Again, converging evidence from both animals and humans using multiple methods indicates that the semantic relationship between modalities is also a robust cue for integration of modality specific information. Unlike the temporal and spatial cues, the semantic relationship is a learned association, and thus, semantic congruency information may not be immediately available when the perceptual signals arrive at the senses. Thus, the semantic relationship can be thought of as higher-level cue that may be available only after some processing of the incoming sensory input. The strongest evidence for this

claim comes from the EEG studies that find effects on the event related potentials 100 milliseconds after stimulus onset for temporally synchronous and spatial congruent stimuli versus the semantic effect that occurs about 400 milliseconds after stimulus onset; however, though there is also converging evidence from one fMRI study that uses a mutual information analysis technique to reveal that primary sensory regions carry information about a multimodal stimulus earlier than frontal regions thought to be involved with semantic processing (Alpert et al., 2008).

Neural Interactions: Summary

The neural evidence for multimodal interactions has been identified throughout a number of regions in the brain, including posterior cortices, temporal cortex, and frontal cortex. These interactions are dependent on factors that reflect a common cause: multimodal information originating from the same source arises from the same point in space, has temporal synchrony, and shares learned semantic associations. Thus, these three cues to integration – space, time, and semantics – may underlie the efficient neural binding of information across modalities that confers the behavioral advantages found for multimodal stimuli.

1.3 Neural Architecture

Across species, multimodal regions are found throughout the cortex, including the primary sensory areas, and research has yet to identify how these regions work in concert with one another. In a recent review by Driver & Noesselt (2008), they discuss three potential neural architectures that may mediate the interactions among modalities,

focusing on audition, vision, and touch. Figure 1 is their schematic representation of three proposed architectures that may account for the multimodal interactions seen across research studies.

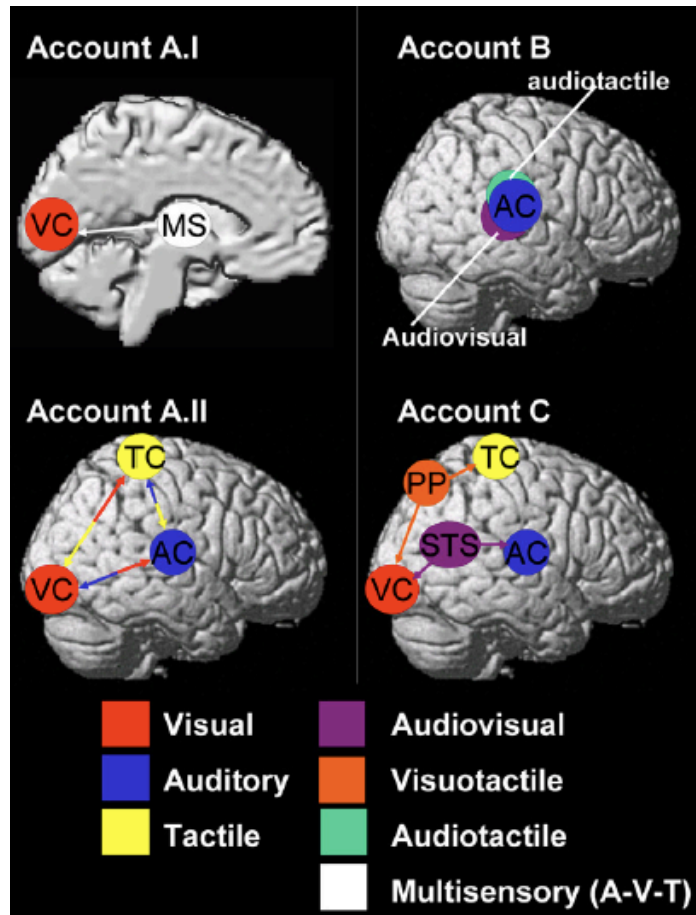


Figure 1
Proposed Neural Architectures
(from Driver & Noesselt 2008)

A.I: multisensory interactions occur in subcortical regions, causing primary sensory regions to only receive multisensory inputs

A.II: all regions are multisensory regions (eliminating the idea of primary sensory areas)

B: each primary sensory area is adjacent to multisensory regions

C: multimodal interactions occur in convergence zones (STS and PP), and these convergence zones use feedback to modulate the primary sensory cortices

As shown in Figure 1, the first proposed architecture (A) has two variations, but the central idea is that all areas are inherently multisensory (Ghazanfar and Schroeder, 2006). This provocative proposal reflects the mounting evidence that reveals multimodal integration in the primary sensory areas that were originally thought to only process one modality. This evidence is based on the neurophysiological and fMRI data that multimodal interactions occur in primary sensory cortices as well as the EEG evidence

for multimodal interactions on the earliest evoked sensory potentials, indicating interactions within 100ms of stimulus onset. In Account A.I, these interactions occur as early as the initial processing in the subcortical regions, while A.II proposes that these interactions occur at the cortical level with direct connections among the primary sensory areas.

The second proposed architecture (B) is a less extreme version of the first account. Here, the proposal is that primary sensory areas maintain their specialization for a particular modality (e.g., audition), but multisensory regions directly abut these primary regions (e.g., auditory-visual just inferior and auditory-tactile just superior). Similar organizations are proposed for the other modalities, though these are not pictured in Figure 1. The data for this proposal is the same as discussed for the first account, allowing for multimodal interactions earlier than proposed in the traditional sequential model where multimodal interactions occur only after primary sensory processing is completed; however, account B preserves the idea that brain regions are predominantly specialized for a given modality.

The third proposed architecture (C) emphasizes the role of feedback circuitry from multimodal convergence zones (also referred to as heteromodal regions, (Beauchamp et al., 2004; Calvert et al., 2000)). Originally, convergence zones were viewed as the second stage of processing, integrating the output from the primary sensory areas. In the current proposal, however, the convergence zones can now influence the primary sensory processing based on feedback circuitry. The posterior superior temporal sulcus (STS) is shown as the convergence zone for auditory and visual processing, while the posterior parietal region (PP) is identified for tactile and visual processing. Driver and

Noesselt cite neurophysiological and EEG data that find delayed multimodal interactions that may reflect the time needed for feedback circuitry, as well as fMRI analyses that use functional connectivity to investigate feedback processing among regions.

Importantly, these three architectures are not proposed as exclusive. In fact, the authors argue that each one may be able to account for particular types of phenomena, but it is the task of future research to relate experimental findings to an underlying neural architecture to better understand how the multimodal regions identified throughout the cortex work in concert with one another.

1.4 Multimodal Interactions: Open Questions

Prior research has established the foundation for investigations of multimodal integration, identifying three cues to integration – time, space, and semantics – as well as a network of brain regions that are consistently recruited during multimodal processing. This evidence is even stable across species. However, this foundational research sparks many new questions. The series of studies in this dissertation were primarily focused on three questions: (1) how do the integration cues interact with one another? (2) what functional roles do the regions recruited in the multimodal network serve in multimodal processing, and are the well-studied convergence zones the primary sites of neural integration? (3) what is the timecourse of the interplay between the integration cues during multimodal processing?

Most research has investigated only one integration cue at a time, so it is unknown if these cues interact with one another. These integration cues influence whether multimodal interactions occur, both behaviorally and neurally; however, it is

relatively unknown whether multimodal regions process every type of integration cue or if regions are specialized for particular cues. One possibility is that a region processes all of the cues, modulating the intensity of its response based on the amount of congruency across all of the integration cues. That is, the intensity of the response reflects the amount of evidence for a common origin for the modalities since a common cause would require congruence across all of the relevant integration cues. Another possibility is that each integration cue recruits unique brain regions that are specialized for its particular type of multimodal relationship, giving rise to regions sensitive only to temporal synchrony, regions sensitive only to semantic congruence, etc. This second possibility would then predict additional regions that show an interaction between the cues (thus, resembling the first possibility) since some regions still need to resolve the overall congruency across the cues to influence whether the modality specific information is bound into a coherent percept. In essence, the question is whether the integration cues are interacting with one another in all regions during multimodal processing or if they recruit separable regions in addition to those regions that show an interaction.

Both monkey and human research identify a network of brain regions recruited by multimodal processing in the primary sensory cortices, posterior temporal cortex, inferior/middle frontal cortex, and often the parietal cortex. What functional roles do the brain regions in the multimodal network serve in multimodal processing? This question essentially returns to the question of the underlying neural architecture discussed in the previous section. Some of these regions may be recruited for sensory-specific processing for each of the modalities, while others may function as convergence zones that integrate the multimodal information. Still other regions may be involved in sending feedback to

regions that are recruited in the earlier stages of processing for reanalysis of the perceptual signals. In short, the goal is to differentiate the functional roles of the regions consistently identified across multimodal studies.

Finally, each of these cues may have a unique timecourse during multimodal processing, influencing the integrative process at different time windows. The neural evidence has suggested that the space and time integration cues may be early, low-level interactions, while the semantic integration cue may be a later, higher-order influence. Thus, the three cues to integration – time, space, and semantics – may influence the multimodal binding process at different time windows during the processing of sensory integration. That is, multimodal information that is temporally synchronous and spatially coincident may be efficiently bound in to a coherent percept during the early stages of processing since these stimulus properties are carried in the perceptual signal and available early during the processing stream. However, as the multimodal information continues to be processed throughout the multimodal network, higher-order information about the modality specific signals can be retrieved and then influence the later stages of multimodal integration. If this higher-order knowledge indicates that the temporally synchronous modalities actually do not belong to the same object, this later perception may be able to correct the perception, unbinding the information. Conversely, the modality specific information may be temporally asynchronous, as is the case in a poorly dubbed foreign film, but yet knowledge about the character's voice and face indicate that the modality specific information (person speaking and the speech heard) did arise from a common source (the same individual), despite the temporal misalignment. Thus, the cues to integration themselves not only interact with one another, but they may have separate

windows of time during the binding process to influence the final percept, i.e., whether the modality-specific information is bound together.

1.5 Three Multimodal Studies

A series of three studies were conducted to investigate each of these open questions by exploring the interplay of two of the three integration cues, temporal synchrony and semantic congruency. In order to disentangle the contributions from each of the cues, two novel sets of stimuli were generated of audio-visual movies of environmental events, such as tapping a pencil or cutting paper. These novel stimulus sets allowed temporal synchrony and semantic congruency to be manipulated independent of one another. For example, movies were edited such that both the action in the video and the sound heard had the same semantic label, i.e., shuffling cards, but the timing of the action was asynchronous. Thus, each of the three studies aimed to investigate the interplay of temporal and semantic congruency when information in both auditory and visual modalities gradually unfolds over time.

Experiment 1 employed fMRI and had three aims. First, we verified that this novel set of stimuli recruited the common set of multimodal regions in posterior, temporal, and frontal cortices. Second, we explored how the two integration cues, time and semantics, modulated the activity of functionally-defined convergence zones (or heteromodal regions) to explore the underlying neural architecture. And third, we investigated whether the regions identified in the whole-brain analysis were sensitive to one cue or both cues, examining the functional role of these regions.

Experiment 2 was also an fMRI study, but this time, a 2x2 factorial design was used to increase our power to detect regions that were differentially sensitive to each of the two cues, as well as regions showing an interaction between the two cues. In addition, the second set of stimuli created for this study refined the nature of the temporal incongruence, allowing us to better understand the underlying computations of the regions sensitive to the temporal relationship between modalities.

Experiment 3 employed EEG, and the primary aim was to investigate the processing timecourse for each cue, identifying the time windows when the integration cue influenced multimodal processing. In particular, this study examined whether the low-level temporal cue had an earlier influence than the higher-level semantic cue.

This series of studies establish a framework for identifying components of the multimodal network sensitive to integration cues in order to explore the underlying neural architecture, study the functional role of particular regions, and examine the interplay of the cue to integration as the temporal dynamics of events gradually unfold over time. More specifically, the studies provide the first exploration of how time and semantics interact with one another during the perception of dynamic, environmental events. Together, these three studies refine our understanding about how signal-based and knowledge-based information help form our integrated experience of events.

2. Chapter 2:

Temporal and Semantic Effects on Multimodal Integration

2.1 Introduction

Real-world events produce both auditory and visual signals that, together, comprise our perceptual experience of that event. At issue is how the brain integrates information from these independent perceptual channels to form coherent percepts. Low-level factors, such as common temporal and spatial structure (Meredith and Stein, 1993), and high-level factors, such as semantic labeling (Doehrmann and Naumer, 2008), both appear to influence integration. Here we examined the effects of both temporal synchrony and semantic congruency on the neural integration of audiovisual (AV) events.

Early research on multimodal integration revealed localized “heteromodal regions” (Beauchamp et al., 2004; Calvert et al., 2000), while more recent research has identified a distributed network of multimodal regions that includes primary sensory areas, posterior temporal cortex, posterior parietal areas, and inferior frontal areas (Driver and Noesselt, 2008; Kayser and Logothetis, 2007). Given such a wide array of potential integration sites, it is of interest to explore how different integration cues modulate and recruit this network.

Many studies have manipulated only a single factor. Studies of temporal congruency have manipulated the onset synchrony of simple stimuli (e.g., circles and tones) to identify brain regions sensitive to temporal coincidence between auditory and visual signals (Bischoff et al., 2007; Bushara et al., 2001). Studies of semantic

congruency have paired static objects with their characteristic sounds (e.g., dog with barking, hammer with pounding) to identify brain regions sensitive to high-level associations (Belardinelli et al., 2004; Taylor et al., 2006). In contrast, investigating the interplay between temporal and semantic factors necessitates using stimuli in which information in both perceptual modalities unfolds over time.

Examples of dynamic stimuli can be found in studies of AV speech: a video of an articulating mouth paired with a spoken phoneme. Yet these studies often manipulate only a single factor – temporal synchrony (Jones and Callan, 2003; Miller and D'Esposito, 2005; Olson et al., 2002) or phonetic congruence (Ojanen et al., 2005; Skipper et al., 2007). A few studies have manipulated two cues simultaneously: temporal versus spatial lip-reading (Macaluso et al., 2004), semantic versus spatial object-sound matching (Sestieri et al., 2006), and semantic versus stimulus familiarity of animals and objects (Hein et al., 2007). Only one study manipulated semantic and temporal relationships, as in our present study, but the stimuli were printed letters and spoken phonemes, a pairing that may recruit neural substrates specialized for reading (van Atteveldt et al., 2007).

Our fMRI study examines the interplay of two cues to integration – temporal synchrony and semantic congruence – using real-world physical events. We filmed AV events, such as tapping a pencil, and edited them to create three multimodal conditions: *congruent* where both semantic and temporal information match across modalities (TCSC), *temporally incongruent* where semantics is congruent but timing is asynchronous (TISC), and *semantically incongruent* where both semantics *and* timing are incongruent (TISI). Comparisons of the neural activity for these conditions implicate a

network of functionally-distinct brain regions involved in multimodal integration. More broadly, our results indicate that both low-level temporal properties in the signal and high-level semantic knowledge help form our integrated experience of events.

2.2 Methods & Materials

Participants. Nineteen right-handed subjects (10 female/9 male; mean age: 23; range: 18-35) participated in the study, but data from four subjects (2 female/2 male) were discarded due to audio equipment failure during scanning. Subjects were paid for their participation and all provided written, informed consent consistent with and approved by Brown University IRB. All had normal or corrected-normal vision and hearing, and they reported that they were neurologically healthy with no contraindications for MRI scanning. The study was conducted at the Magnetic Resonance Facility at Brown University (Providence, RI).

Stimuli. A total of 40 AV movies were digitally filmed at standard definition for this study, including two exemplars of 20 unique real-world, physical events (e.g., typing on a keyboard, bouncing ball, jingling keys, pouring water, etc). In each movie, only the arm and hand of the actress were seen performing the event (Figure 2). The movies were edited using Final Cut Pro (Apple Co., Cupertino, CA) to be 2300 msec in duration and exported as 720x480 Quicktime™ (Apple Co., Cupertino, CA) movies with 16 bit, 44.1kHz stereo audio.

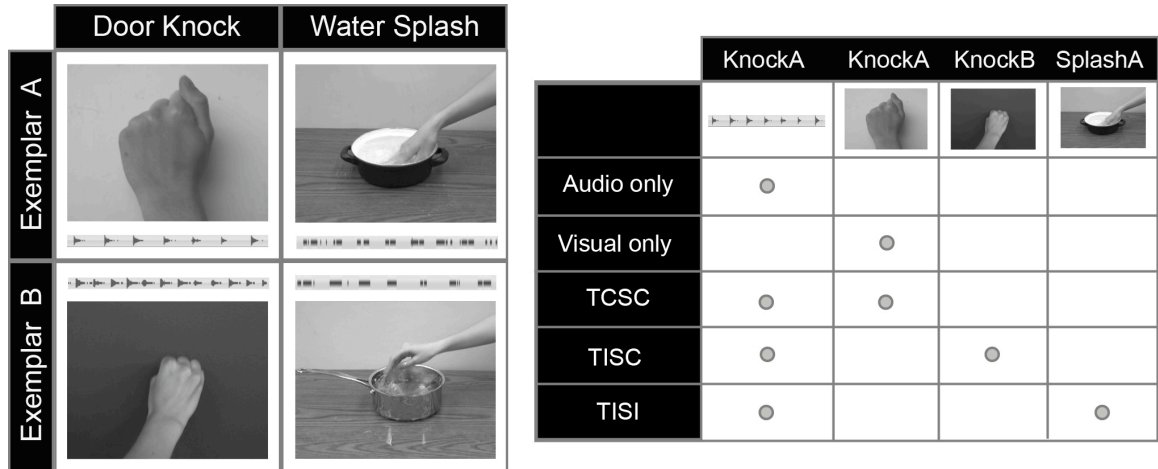


Figure 2 *Experimental Design*

On the left, two exemplars are shown for two events, door knocking and water splashing. The audio track for each exemplar reveals the different temporal characteristics between the two exemplars. The right table provides an example of how the five different experimental conditions are constructed: the Audio only condition presents just the audio track, the Visual only condition presents just the video track, the TCSC (congruent) condition presents the audio and video from same exemplar, the TISC (temporally incongruent) condition presents the audio from exemplar A and the video from exemplar B, and the TISI (semantically incongruent) condition presents the video from water splashing and the audio from door knocking.

Experimental Conditions. Each of the 40 events was presented in five conditions, two unimodal, auditory only (A) and visual only (V), and three multimodal conditions – a *congruent* condition (TCSC) and two incongruent conditions, *temporally incongruent* (TISC) and *semantically incongruent* (TISI). The congruent AV movies always showed audio and video from the same exemplar of a given event type. The TISI movies showed the video from one event type (e.g., door knocking) simultaneously with the sound from a different event type (e.g., splashing water). The TISC movies showed audio and video from the same event type (e.g., door knocking) but the movie was taken from one of the two door knocking exemplars (door knocking B) and the sound was taken from the other exemplar of the same event type (door knocking A). Thus, both modalities should elicit

the same semantic label (“door knocking”) but the timing between the two modalities was asynchronous (Figure 2). The TISC condition was always generated by pairing the video from one exemplar with the audio from the other exemplar of the same event type. For the TISI condition, each event was randomly paired with a different physical event for each subject.

Experimental Task. Prior research has suggested that making an explicit congruency judgment can effect the obtained pattern of neural responses (van Atteveldt et al., 2007); instead, our subjects performed a simple stimulus location judgment, indicating whether the video or audio (different scans) was seen or heard on the right or left side. Each video was offset by 10% of the screen size (100 pixels) from the center fixation, and each sound was delayed by 500 microseconds in one ear so that it was lateralized towards the opposite ear (amplitude was held constant between the two ears so as not to introduce spurious differences in the neural signals between brain hemispheres). In all AV conditions, the audio and video were always offset in the same direction so as to be perceived as being in the same location (i.e., no spatial location incongruency). All subjects pressed the right button with their right index finger for right side and left button with left index finger for left side, thereby maintaining response congruency. Although the right/left location of the two modalities was always congruent, subjects only reported either the location of the video or the location of the audio so that the identical task could be used in the unimodal conditions. The modality used for the judgment alternated between scans and the order of the judgment was counterbalanced between subjects. As

expected, accuracy was equally high for all five conditions across both modalities (96% $\pm$.01%).

Experimental Procedure. Before the scan, each subject practiced the location judgment task using movies from a different study (Zacks et al., 2006) that were not seen during the actual experiment. Subjects then studied a list of the 20 unique physical event types to appear in the experiment, performed a written recall task, and then read aloud the name of any event type not recalled. This procedure ensured that all events were familiar and identifiable during the scanning session. During the scanning session, high resolution anatomical MP-RAGE images were collected first, followed by four functional EPI scans using a fast event-related design. Each scan consisted of 20 unimodal trials (visual-only for scans judging video location and auditory-only for scans judging sound location), 10 TCSC, 10 TISI, and 10 TISC multimodal trials. Thus, each scan had 50 experimental trials intermixed with 40 baseline fixation trials in stimulus orders that were optimized for GLT parameter estimation (AFNI's RSFgen, <http://afni.nimh.nih.gov/afni/doc/howto/3>). Each EPI scan lasted four and a half minutes.

fMRI Details. MRI data were acquired using a 3T Siemens TIM Trio scanner (Erlangen, Germany). Visual stimuli were presented through a rear projection system with a mirror mounted to the head coil, and audio stimuli presented through an Avotec SS-3100 Silent Scan audio system (Stuart, FL). Participants made behavioral responses with their left and right index fingers using the two outside buttons on a 4 button response pad (Mag Design & Engineering, Sunnyvale, CA). The high resolution MPRAGE scan included

160 interleaved slices with 1mm isotropic voxels in a 256x256 matrix with a TR=1900msec, TE=2.98msec, and flip angle=9degrees, and the T1-weighted echoplanar images (EPI) covered the whole brain with 48 interleaved slices, 3mm isotropic voxels, TR=3000msec, TE=28msec, and flip angle=90degrees. Each EPI scan consisted of 90 volumes. Prior to analysis in AFNI (Cox, 1996), the functional data were pre-processed in AFNI to correct for slice timing differences and 3-D head movement, smoothed with a FWHM kernel of 1 voxel (3mm), and normalized to allow the beta coefficients to reflect percent signal change by dividing each time series by its mean.

fMRI Data Analysis

Prior to warping the data to the standard atlas (Talairach and Tournoux, 1988), the data for each individual was fit with a general linear model (GLM) to estimate the activity at each voxel using a regressor for each of the five experimental conditions (two unimodal and three multimodal conditions) as well as regressors of no interest to account for second-order polynomial trends. A gamma hemodynamic response function was assumed. The output of each individual participant's GLM was then warped to the standard atlas and analyzed with a mixed-effects group ANOVA that included the three pairwise statistical contrasts of the parameter estimates for the multimodal conditions: TCSC minus TISC (temporal processing), TISC minus TISI (semantic processing), and TCSC minus TISI (congruency processing). In addition, a conjunction analysis identified integration regions with voxels more active in the TCSC condition than either unimodal condition, with the additional requirement that unimodal condition must also be

significantly greater than baseline, that is, $0 < A_{\text{only}} < \text{TCSC} < V_{\text{only}} > 0$ (Beauchamp, 2005).

All statistical maps were thresholded at $p < 0.05$ with a requirement that all regions of interest (ROIs) have at least 24 voxels (connected by 3mm, i.e., one side of a voxel) to ensure an alpha level of 0.05 (AFNI's AlphaSim program written by B. Douglas Ward, run with 5000 Monte Carlo simulations).

2.3 Results

Unimodal Regions

As expected, a statistical contrast of auditory-only (A_{only}) versus visual-only (V_{only}) revealed extensive and separable neural activation for both unimodal conditions. Large clusters of activation for A_{only} were found in primary auditory regions as well as the surrounding superior temporal gyrus, insula, supramarginal gyrus, and medial frontal gyrus. Widespread activation for V_{only} was observed throughout the posterior areas in the brain, including primary visual areas, fusiform gyrus, lingual gyrus, cuneus/precuneus, and parietal cortex.

AV Contrast: Integration Regions-of-Interest (ROIs)

As discussed in the Methods section, we adhered to a functional criteria for integration defined by prior research (Beauchamp, 2005); that is, integration regions were identified by a conjunction analysis of brain regions that were more active in the TCSC condition than either unimodal condition with the additional constraint that both unimodal conditions were greater than the baseline activity.

Four regions met this criteria: 1) a medial region in the posterior cingulate gyrus (pCing.G); 2) a posterior region in left middle temporal gyrus (L MTG); 3) a region slightly anterior to L MTG in the middle temporal gyrus that also extended upwards to the superior temporal sulcus (L STS/MTG); and 4) a parietal region in the left hemisphere.

These ROIs are shown in Figure 3A with further details presented in Table 1.

Note that the two incongruent conditions were *not* used to identify the integration ROIs; a *t*-test between these two conditions showed no significant difference, indicating that the ROIs associated with functionally-defined integration did not differentiate between the two incongruent conditions (pCing.G $t(14)=0.63$, $p=0.54$; L MTG $t(14)=0.74$, $p=0.47$; L STG/MTG $t(14)=0.42$, $p=0.68$; L Parietal $t(14)=0.15$, $p=0.88$).

Figure 3 *Functionally-defined Integration ROI and Overlap of AV Contrast Regions*

(A) The four functionally-defined Integration ROIs are pictured on the brain from one subject in the study with the coordinates of the crosshairs listed below the image. The ROIs were defined on the group data, but under each region, the mean percent signal change for each ROI in each individual subject is plotted for all five experimental conditions (A_{only} = Auditory only; V_{only} = Visual only; TCSC = congruent; TISI = semantically incongruent; TISC = temporally incongruent). The error bars reflect standard error of the means between subjects. A *t*-test was performed on the means for the two incongruent conditions, and the n.s. denotes each test was not significant (see Results, Integration ROIs).

(B) Group ROIs in the parietal, posterior cingulate, and frontal regions identified in the three pairwise AV contrasts are pictured: regions identified in the Semantic Processing comparison (TISC-TISI) are shown in orange; regions from Congruency Processing (TCSC-TISI) are shown in light blue; and regions from Temporal Processing (TCSC-TISC) in green. This figure highlights the overlap of the regions identified in these contrasts, with pink voxels identifying overlap between the Semantic Processing and Congruency Processing contrasts (suggesting the region is sensitive to semantics) and maroon voxels showing overlap between the Congruency Processing and Temporal Processing Contrasts (suggesting the region is sensitive to timing; see Results for more details). In addition, the functionally-defined integration regions in parietal and posterior cingulate cortices are shown in purple.

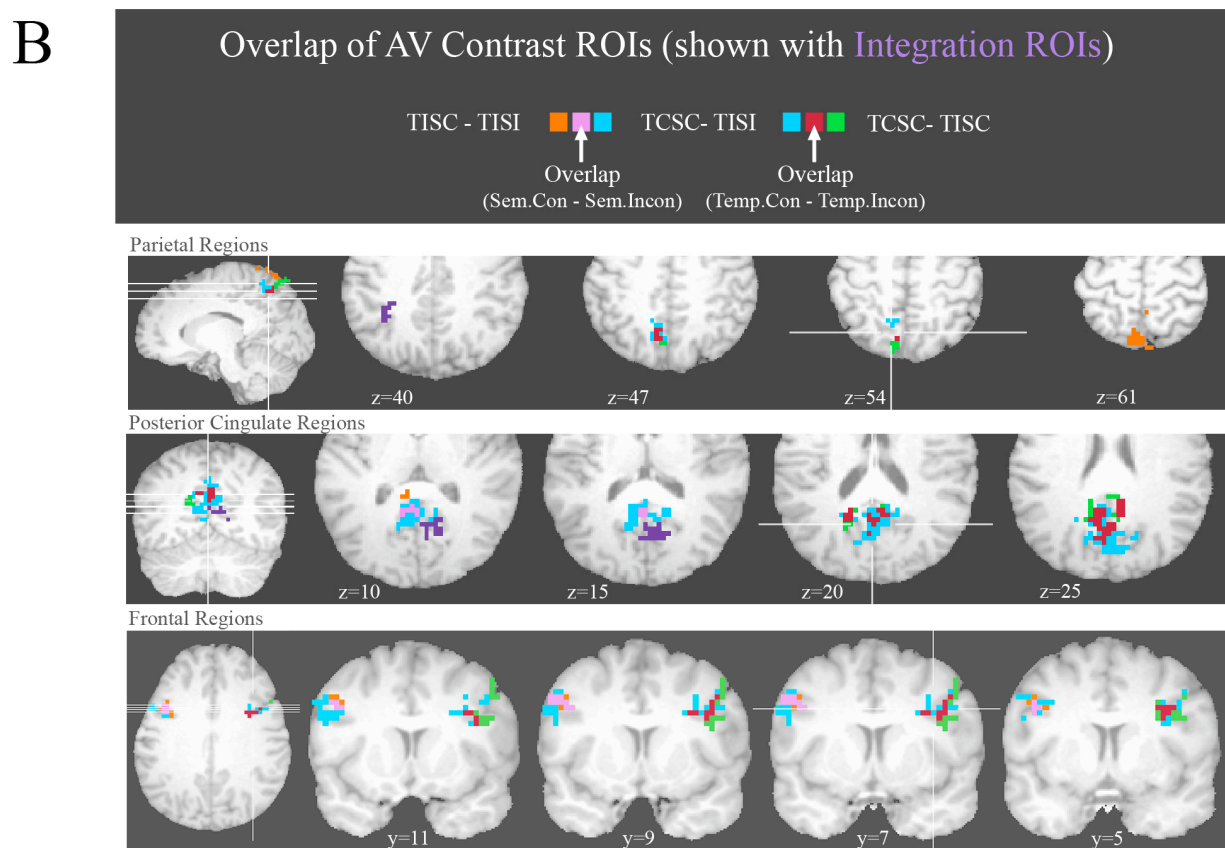
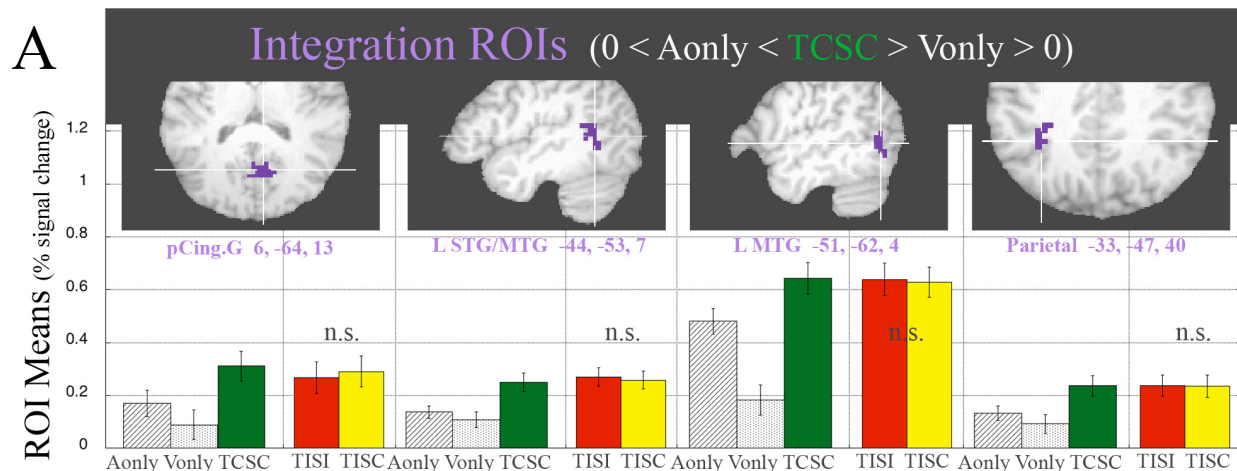


Table 1 ROI Location, Size, and Peak Coordinates for all AV Contrasts

Row Number & ROI Anatomical Location	Hemi.	# of Voxels	Integration ROIs	Con – Temp.I –		Con – Sem.I	Overlap	Peak (Tailarch)		
				Temp.I	Sem.I			x	y	z
1 Middle/Superior Temporal	L	29	*				-	-43.5	-43.5	14.5
2 Middle Temporal (posterior)	L	20	*				-	-52.5	-64.5	-0.5
3 Middle Temporal (posterior)	R	27		□			-	61.5	-46.5	-6.5
4 Middle Temporal (posterior)	R	28			◇		-	55.5	-61.5	17.5
5 Middle Temporal (posterior)	L	24				∧	-	-49.5	-73.5	23.5
6 Middle Temporal (anterior)	R	27		□			-	58.5	7.5	-18.5
7 Cingulate (posterior)	R	35	*				-	1.5	-58.5	11.5
8 Cingulate (posterior)	L/R	117		□			10	-10.5	-49.5	26.5
9 Cingulate (posterior)	L	31			◇		10	-1.5	-46.5	8.5
10 Cingulate (posterior)	L/R	418				∧	8 & 9	-1.5	-31.5	8.5
11 Cingulate (middle)	L	24			◇		-	-16.5	-25.5	29.5
12 Parietal	L	12	*				-	-34.5	-46.5	44.5
13 Parietal	L	32		□			16	-10.5	-73.5	56.5
14 Parietal	L/R	33			◇		-	-1.5	-61.5	62.5
15 Parietal	L/R	25			◇		-	-4.5	-52.5	65.5
16 Parietal	L	25				∧	13	-4.5	-61.5	47.5
17 Lingual	L	51		□			-	4.5	-91.5	-6.5
18 Lingual -9.5	R	35		□			-	16.5	-103.5	
19 Lingual -0.5	R	62			◇		-	13.5	-100.5	
20 Middle Occipital	L	28			◇		-	-25.5	-88.5	2.5
21 Middle Occipital	R	25			◇		-	46.5	-67.5	-3.5
22 Parahippocampus	R	30				∧	-	22.5	-19.5	-6.5
23 Inferior/Middle Frontal	R	68		□			24	46.5	10.5	47.5
24 Inferior/Middle Frontal	R	56				∧	23	43.5	7.5	41.5
25 Inferior/Middle Frontal	L	32			◇		26	-43.5	10.5	35.5
26 Inferior/Middle Frontal	L	70				∧	25	-52.5	10.5	35.5
27 Middle Frontal	L	25				∧	-	-46.5	4.5	47.5
28 Middle/Superior Frontal	L	28			◇		-	-28.5	19.5	50.5
29 Superior Frontal	L/R	51		□			30	-1.5	19.5	50.5
30 Superior Frontal	L/R	31				∧	29	1.5	13.5	50.5
31 Caudate	L	26			◇		32	-4.5	10.5	-0.5
32 Caudate	L	40				∧	31	1.5	7.5	-0.5
33 Cerebellum	L/R	34			◇		-	22.5	-31.5	-33.5
34 Cerebellum	R	27			◇		-	22.5	-55.5	-21.5
35 Cerebellum	L/R	77			◇		36	1.5	-64.5	-33.5
36 Cerebellum	L/R	34				∧	35	4.5	-61.5	-18.5
37 Medial Frontal	L/R	60		□			40	4.5	58.5	11.5
38 Anterior Cingulate	R	29		□			40	-4.5	40.5	-12.5
39 Anterior Cingulate	L	29			◇		40	-1.5	43.5	5.5
40 Medial Frontal/Ant. Cing.	L/R	599				∧	37-39	4.5	67.5	17.5

a. All regions were identified with a $p < .05$, minimum cluster size of 24 voxels, alpha .05

b. Abbreviations: ROI = Region-Of-Interest; Hemi. = Hemisphere; # of Voxels = Number of 3mm voxels in region; Con = Congruent; Temp.I = Temporally Incongruent; Sem.I = Semantically Incongruent

c. Overlap column lists the row number of the ROI that overlaps

AV Contrast: Temporal Processing

The contrast between the TCSC and TISC conditions identifies brain regions that are sensitive to whether the timing relationship between the visual and auditory signals is synchronous for events that are semantically congruent; in other words, both the visual and auditory modalities are labeled as door knocking but the timing of the knock in the visual domain is not in synchrony with the sound of the door knock in the auditory domain.

Six regions involved in temporal processing are depicted in Figure 4 with further details presented in Table 1: an anterior (maroon) and a posterior (orange) region in the right middle temporal gyrus, bilateral regions (blue and red) in the lingual gyrus, a region that spans from the superior edge of the inferior frontal gyrus up through the medial frontal gyrus (pink) in the right hemisphere, and a medial region in the superior frontal gyrus (purple).

Two additional regions were identified in posterior cingulate gyrus and parietal cortex. These two are depicted in Figure 3B (green) to illustrate their spatial relationship to the integration regions identified in these same anatomical regions but with distinct activation clusters.

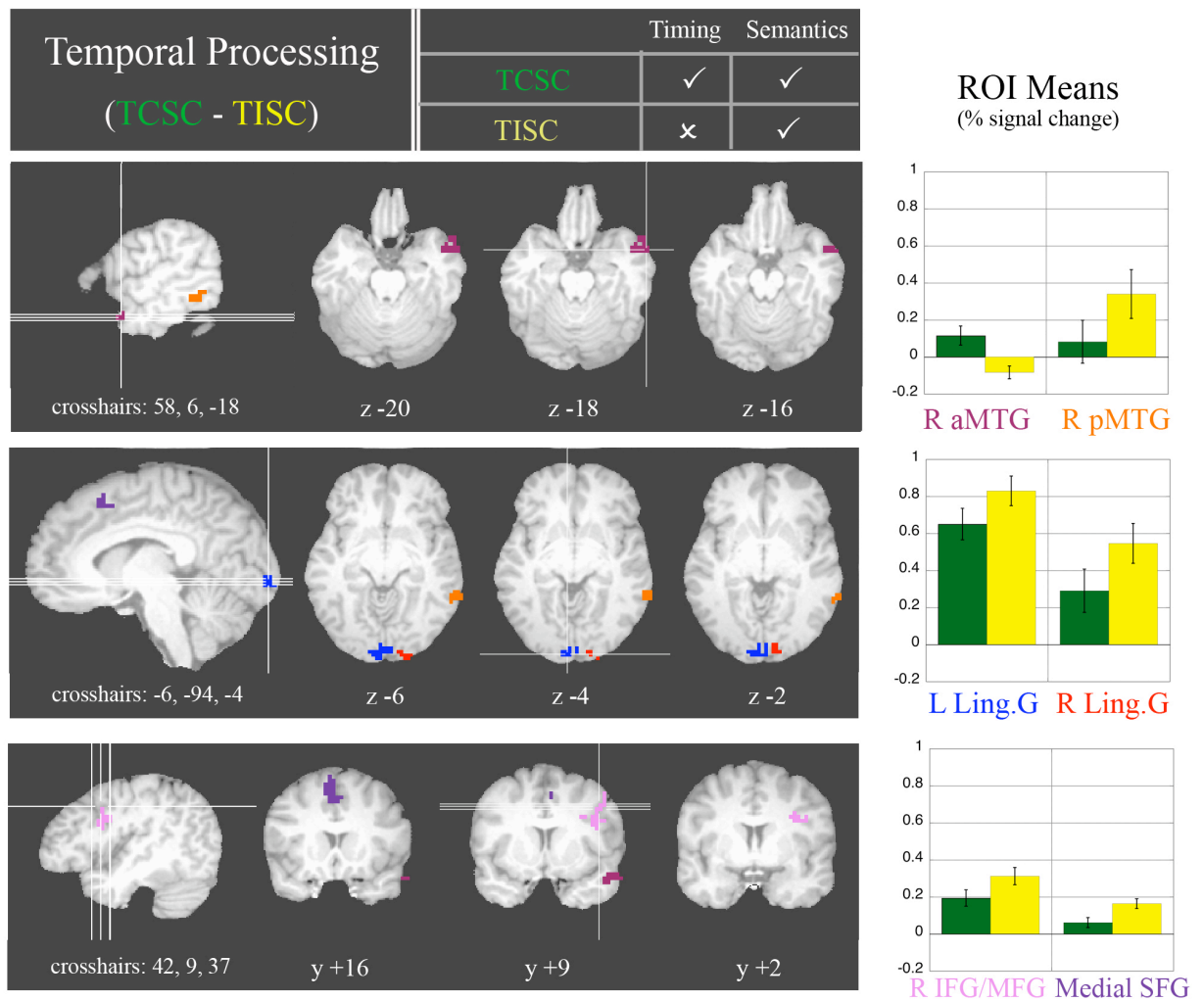


Figure 4 *AV Contrast: Temporal Processing*

The regions identified by the AV contrast of TCSC (congruent) and TISC (temporally incongruent) are plotted in six different colors. The bar graphs on the right show the mean percent signal change in each group-defined ROI from each individual subject (bars are between-subject error): green for TCSC and yellow for TISC. The color of the bar graph label matches the color of the corresponding voxels. Abbreviations: R = right; L = left; a = anterior; p = posterior; MTG = middle temporal gyrus; Ling.G = lingual gyrus; IFG = inferior frontal gyrus; MFG = middle frontal gyrus; SFG = superior frontal gyrus

AV Contrast: Semantic Processing

The contrast of TISC and TISI identifies regions that are sensitive to the semantic labels that arise as a result of event identification between the visual and auditory domains. Critically, for these conditions the temporal structure of the visual and auditory signals is incongruent in both cases. In other words, the visual and auditory modalities are not temporally in synchrony for either condition, but in the TISC condition, both modalities should be labeled the same (e.g., “door knocking”), while in the TISI condition, the auditory signal should prompt one label and the visual signal should prompt a different label (e.g., “door knocking” and “water splashing”).

Eight regions involved in semantic processing are depicted in Figure 5 with further details presented in Table 1: a region in left middle occipital gyrus (dark blue), a region in right lingual gyrus (dark green; lateral and superior to temporal processing Ling.G), a more lateral and anterior region in right middle occipital gyrus (light blue), a region in right middle temporal gyrus (orange; superior and posterior to the temporal processing R MTG region), a region that spans from the superior edge of the inferior frontal gyrus up through the medial frontal gyrus in the left hemisphere (pink), a region slightly superior in the medial frontal gyrus that extends in to superior frontal gyrus (purple; lateral to the temporal processing MFG/SFG region), and a region in the caudate (maroon).

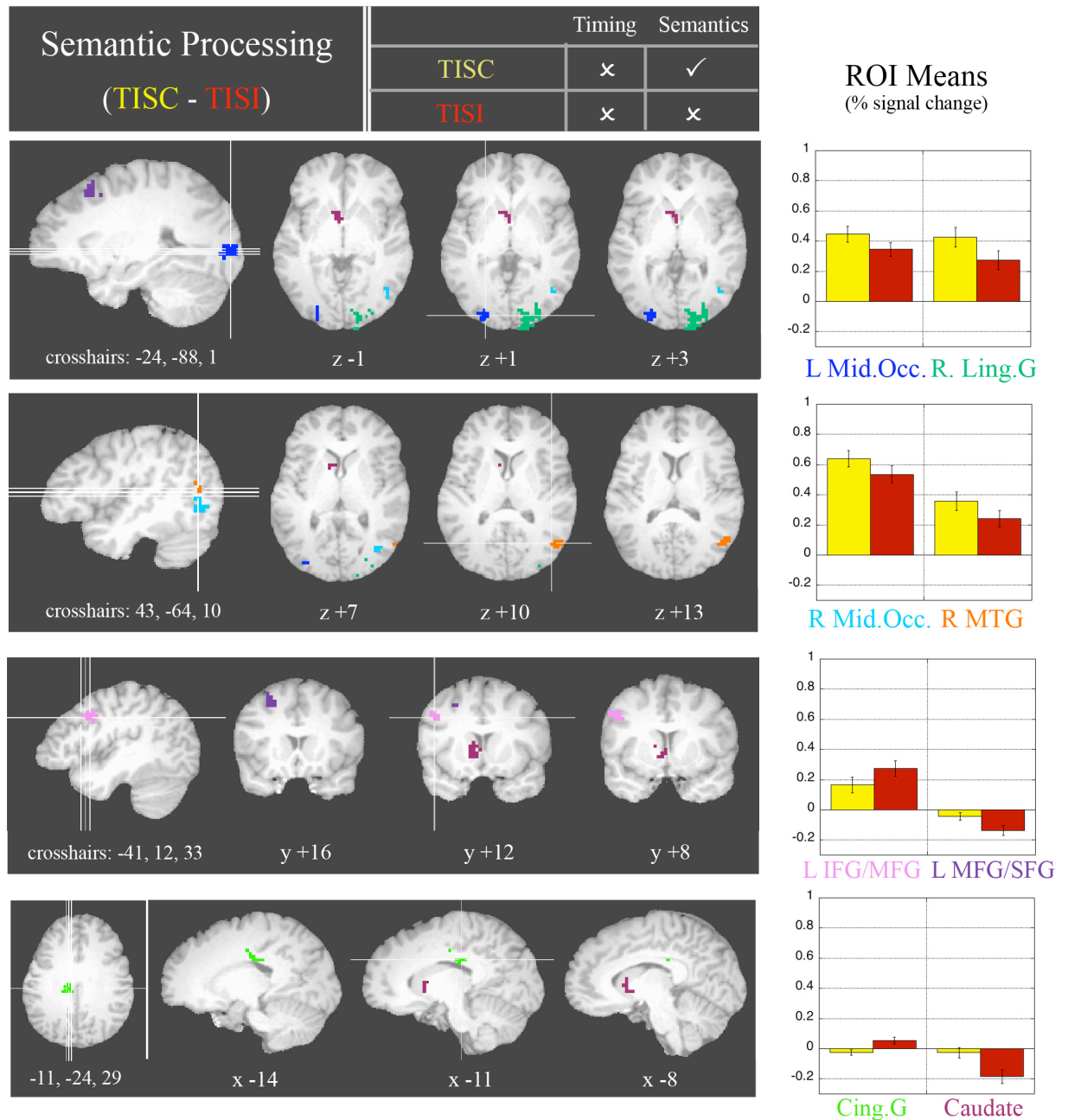


Figure 5 *AV Contrast: Semantic Processing*

The regions identified by the AV contrast of TISC (temporally incongruent) and TISI (semantically incongruent) are plotted in eight different colors. The bar graphs on the right show the mean percent signal change in each group-defined ROI from each individual subject (bars are between-subject error): yellow for TISC and red for TISI. The color of the bar graph label matches the color of the corresponding voxels. Abbreviations: R = right; L = left; p = posterior; Mid. Occ. = middle occipital gyrus; Ling.G = lingual gyrus; MTG = middle temporal gyrus; IFG = inferior frontal gyrus; MFG = middle frontal gyrus; SFG = superior frontal gyrus; Cing.G = cingulate gyrus

Three separate regions in the cerebellum (not pictured), one region in middle cingulate gyrus (not pictured), one region in posterior cingulate gyrus, and two regions in parietal cortex were identified. The posterior cingulate and one of the parietal regions (row 14 in Table 1) are depicted in Figure 3B (orange) to illustrate their spatial relationship to the integration regions as well as the other AV contrasts regions. The not-pictured parietal ROI is just superior to the pictured one, and it also spans both hemispheres.

AV Contrast: Congruency Processing

The contrast of TCSC and TISI identifies brain regions that are sensitive to overall congruence between the visual and auditory modalities; in other words, a movie of water splashing is asynchronous with the sound of door knocking in that both the temporal structure and the semantic labeling provide information that the two domains are incongruent.

Seven regions involved with congruency processing are depicted in Figure 6 with further details presented in Table 1: a region in left middle temporal gyrus (orange; posterior and superior to the integration regions in MTG), a region in left medial frontal gyrus (light blue; posterior to the semantic processing L MFG region), a region in medial superior frontal gyrus (yellow; overlaps with temporal processing medial SFG), bilateral regions that span from the superior edge of the inferior frontal gyrus up through the medial frontal gyrus (purple and pink; the purple overlaps with the semantic processing L IFG/MFG region, and the pink overlaps with the temporally processing R IFG/MFG

region), a region in the parahippocampus (dark blue), and a region in the caudate (maroon; overlaps with the semantic processing caudate region).

One region in the cerebellum (not pictured), one region in posterior cingulate gyrus, and one in parietal cortex were identified. The latter two are depicted in Figure 3B (teal blue) to illustrate their spatial relationship to the integration regions as well as the other AV Contrasts regions.

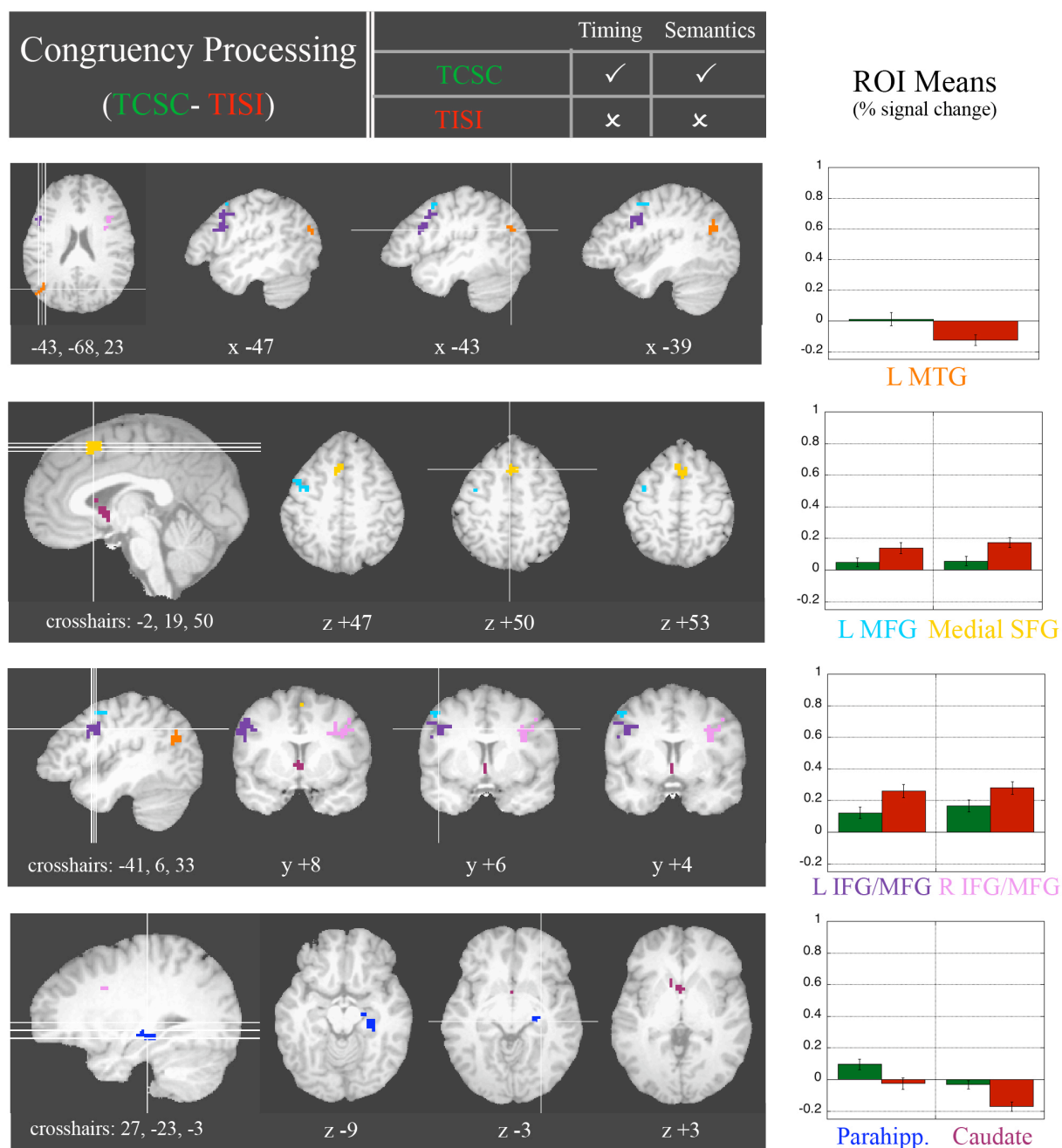


Figure 6 *AV Contrast: Congruency Processing*

The regions identified by the AV contrast of TCSC (congruent) and TISI (semantically incongruent) are plotted in seven different colors. The bar graphs on the right show the mean percent signal change in each group-defined ROI from each individual subject (bars are between-subject error): green for TCSC and red for TISI. The color of the bar graph label matches the color of the corresponding voxels. Abbreviations: R = right; L = left; MTG = middle temporal gyrus; MFG = middle frontal gyrus; SFG = superior frontal gyrus; IFG = inferior frontal gyrus; Parahipp. = parahippocampus

Posterior Cingulate

In each of the four AV contrasts, regions were found in the posterior cingulate, and these are depicted in Figure 3B and listed in Table 1. The most extensive activity arose from the Congruency Processing comparison (light blue). The inferior portion of this region overlapped with the Semantic Processing ROI (orange). This overlap region (pink) may reflect computations concerning semantic relationships since both AV contrasts compare a semantically congruent condition (TCSC and TISC) to the TISI condition. This overlap region seems insensitive to the temporal relationship, as one contrast is temporally synchronous (TCSC) and one asynchronous (TISC). Although this region is adjacent to the Integration ROI, no voxels overlap.

The superior portion of the Congruency Processing ROI (light blue) overlaps with the Temporal Processing ROI (green), and this overlap region (maroon) may reflect sensitivity to the temporal relationships in multimodal stimuli since both of these AV contrasts compare the TCSC condition, which is temporally synchronous between modalities, with two temporally incongruent conditions, that is, where the modalities are temporally asynchronous (TISC and TISI). In contrast, the overlap region appears insensitive to semantic relationships since one contrast is semantically congruent (TISC) and the other is incongruent (TISI). This activity is superior to the Integration ROI.

Parietal Cortex

As in the posterior cingulate, parietal regions were found in each of the three AV contrasts: these are depicted in Figure 3B and listed in Table 1. The integration ROI is

inferior and lateral to the other parietal regions. The integration ROI is located within the left inferior parietal lobule, while the other ROIs are superior and located in the precuneus, very close to the midline, and extending upwards in the superior parietal lobule.

The Congruency Processing ROI (light blue) and Temporal Processing ROI (green) have a high percentage of overlap (maroon), and as described in the previous section on the Posterior Cingulate, the overlap of these two AV Contrast maps may identify a region sensitive to the timing relationship between modalities.

Superior to all of the other AV Contrast ROIs, the Semantic Processing ROI (orange) identifies two regions in the parietal cortex, one directly superior to the other. The more inferior of the two is shown in Figure 3B. Unlike in the posterior cingulate, there is no overlap of this ROI with the Congruency Processing ROI (light blue) to identify a region that may be more sensitive to the semantic relationship between modalities (that is, no pink voxels).

Frontal Cortex

A frontal ROI was identified in all three AV contrasts, and as shown in Figure 3B, there is once again overlap among the regions. In the right hemisphere, the Congruency Processing ROI (light blue) overlaps with the Temporal Processing ROI (green), and this overlap region (maroon) may reflect sensitivity to the temporal relationships. Conversely in the left hemisphere, the Congruency Processing ROI (light blue) overlaps with the Semantic Processing ROI (orange), suggesting a region (pink voxels) sensitive to the semantic relationship between modalities. Thus, the congruent TCSC condition identifies

bilateral regions, but there appears to be some hemispheric specialization for different cue types that is revealed only when the two integration cues are pitted against one another in the TISC and the AVIS conditions.

Resting State Network and the AV Contrasts

This series of AV contrasts also revealed regions within the anterior portion of the medial frontal gyrus and anterior cingulate, regions thought to be part of the resting state network and characterized by deactivations that are largely task independent (Raichle et al., 2001). Consistent with this classification, extracting the ROI means from these anterior regions revealed that each condition in the contrast was deactivating the region. Other studies have found similar patterns of deactivation within this resting state network (Beauchamp, 2005; Laurienti et al., 2002) and excluded them from further analysis. Similarly, in our analysis these regions are excluded from further discussion though details are listed in Table 1.

2.4 Discussion

This study investigated how semantic and temporal congruency interact and modulate the brain regions consistently identified by prior research on multimodal integration, including primary sensory areas, posterior temporal, posterior parietal, and inferior frontal regions (Driver and Noesselt, 2008; Kayser and Logothetis, 2007). Using pairwise comparisons between our AV conditions, we successfully identified ROIs in all of these regions, and our study suggests how this network of brain regions may be differentially

recruited to support specific processing of the semantic and/or temporal relationship between modalities. In particular, one novel and unexpected result is the hemispheric difference in the frontal cortex showing that the right hemisphere may be more sensitive to timing relationships and the left to semantic relationships. Interestingly, none of the four functionally-defined integration regions were modulated by congruency.

Because we explicitly examined the interplay between temporal synchrony and semantic congruency across modalities, we are better able to isolate specific neural subregions that may support a particular type of cue integration, rather than generic integration effects. That is, by identifying the separable components of the multimodal processing network, our study advances the understanding of how modality-specific information is bound into coherent event percepts. Consequently, our discussion will focus on the degree of overlap between AV conditions. However, before we address what these overlap analyses reveal about multimodal integration, we first review the functionally-defined approach to localizing integration regions.

Functionally-defined Integration ROIs

Four integration regions were identified that showed higher activation for the TCSC condition than either of the unimodal conditions, with the constraint that both unimodal conditions were greater than baseline ($0 < A_{\text{only}} < \text{TCSC} < V_{\text{only}} > 0$; (Beauchamp, 2005)): two left posterior temporal regions, one region near the left intraparietal sulcus, and one medial region in the posterior cingulate.

Based on earlier multimodal research (Beauchamp et al., 2004; Calvert, 2001), many studies focused on the posterior STS/MTG as a site of integration, and often as the

source of feedback to other regions in the multimodal network (van Atteveldt et al., 2007); however, several recent reviews have failed to show congruency effects within pSTS/MTG (Hocking and Price, 2008; Doehrmann et al., 2008; Hein et al., 2007) and other studies failed to find any multimodal effects in this brain area (Belardinelli et al., 2004; Bushara et al., 2001) or find congruency effects in anterior temporal regions instead (van Atteveldt et al., 2004). Our two temporal integration regions are within 3mm of the two average locations identified in Hocking & Price (2008; $\pm 50, -52, 8$ & $\pm 50, -56, 4$), and neither temporal nor semantic congruency modulated the region (Figure 3A). Furthermore, Hocking & Price (2008) report that there is no significant difference in activation between intramodal (two auditory or two visual stimuli) and cross-modal (auditory and visual) trials. Thus, the pSTS/MTG region may not process whether information across multiple modalities should be integrated into a common percept. Instead, the pSTS/MTG may access amodal representations, as suggested by Hocking & Price, or serve as a gateway between the frontal and medial temporal lobes whose function varies based on task-dependent connections (Hein and Knight, 2008).

The parietal cortex is consistently activated in multimodal studies (Amedi et al., 2005; Driver and Noesselt, 2008), and our integration ROI matches the commonly activated region near the intraparietal sulcus. However, the locus of our semantic and temporal congruency effects lie superior and posterior in the precuneus. Similar regions have been found in numerous studies using stimuli as diverse as vertical bars and beeps (Bushara et al., 2003), animals and novel objects (Hein and Knight, 2008), and articulating mouths (Miller and D'Esposito, 2005; Ojanen et al., 2005). This area appears to be recruited when the two modalities are incongruent (semantically incongruent in

Hein et al; phonetically incongruent in Ojannen et al; temporally incongruent in Miller & D'Esposito). Bushara et al (2003) found similar regions that were more active when the audio and visual stimuli were not bound into a coherent percept, which is likely to have occurred in our conditions that activate this region. Finally, parietal regions are identified when studies have manipulated the spatial congruency of AV stimuli, finding that this region responds to spatially congruent stimuli (Saito et al., 2005; Sestieri et al., 2006), which is true for every AV condition in our study (and true for the studies cited above). Thus, the precuneus may be differentiating when modalities are spatially congruent but yet still in conflict (be it a semantic, phonetic, or temporal conflict).

To our knowledge, no other study has found an integration region in the posterior cingulate, but this likely reflects our novel stimulus set of environmental events (not the typical tools, animals, or speech). Consistent with this interpretation, Lewis and colleagues (2004) investigated brain regions recruited for recognizing environmental sounds (compared to unrecognizable reversed versions of each sound), and identified several regions including caudate, posterior cingulate, posterior MTG, and IFG that mirror our results. Similar to the parietal cortex, the posterior cingulate integration region did not show congruency effects but adjacent (not overlapping) regions were modulated by temporal and/or semantic congruency. Future research is needed to delineate the role of these differentiated effects, both within the posterior cingulate and the precuneus.

Overlap of Regions Among the 3 AV Comparisons

Looking beyond the functionally-defined integration regions, the pairwise comparisons of the three multimodal conditions in the whole-brain analysis identified

regions that may process a particular integration cue, semantics or timing. In particular, regions identified in two of the three AV comparisons that overlap with one another may indicate a sensitivity to one of the two integration cues that is independent of the other.

The overlap of regions in the Semantic Processing and Congruency Processing comparisons suggest a role in differentiating the semantic relationship between modalities, as both of these contrasts compare conditions with congruent semantics across modalities to the condition in which semantics are incongruent (mixing whether the timing relationship is synchronous or not). Consequently, these overlap regions suggest a sensitivity to semantics that is independent of the temporal relationship between modalities. These contrasts reveal overlap in the left IFG/MFG region. Left lateralized effects are consistent with the well-known language dominance in the left hemisphere, as well as previous multimodal research on lip reading (Paulesu et al., 2003) and events/tools (Lewis et al., 2005; Doehrmann et al., 2008). Similarly, research using amodal semantic priming identified a similar ROI (Buckner et al., 2000). The only other region to overlap in these two contrasts is in the inferior portion of the posterior cingulate, and as discussed in the previous section, this region may be specific to the stimulus class of environmental events (Lewis et al., 2004).

Conversely, the overlap of regions in the Temporal Processing and Congruency Processing comparisons suggests a role in differentiating the temporal relationship between modalities, as both of these contrasts compare the condition that is temporally synchronous across modalities with the conditions in which the timing relationship is asynchronous (ignoring whether the semantic relationship is congruent or not). Thus, the overlap of these regions suggests a preference for timing relations that is independent of

semantics. Interestingly, one of the overlap regions is in the right IFG/MFG, a region that is bilateral to the overlap region in left IFG/MFG, is identified for semantic processing. Thus, our study suggests a hemispheric difference in the frontal cortex, with semantics focused in the left hemisphere and temporal relationships in the right hemisphere. This novel result arises from the addition of the TISC condition in our study, as prior research has typically looked at semantic congruency with conditions equivalent to our Congruency Processing (TCSC-TISI) comparison, finding bilateral activation in frontal cortices (Beauchamp et al., 2004;Naumer et al., 2008;Taylor et al., 2006;van Atteveldt et al., 2007). However, two studies that looked at onset synchrony of simple stimuli (e.g., circles/tones) have reported right hemisphere dominance for temporal synchrony (Bushara et al., 2001;Calvert et al., 2001). Finally, in addition to the left MFG/IFG region, two additional overlap regions sensitive to temporal relationships were found in the superior portion of the posterior cingulate and the medial region of the precuneus, both of which are discussed in the previous section.

2.5 Conclusion

Our study was designed to disentangle some of the factors that contribute to the integration of perceptual signals arising from multimodal events. More specifically, we were interested in exploring the neural substrates recruited by integration cues arising from both low-level information in the signal (e.g., common temporal structure) and high-level knowledge (e.g., the semantic labels applied to perceptual signals). To address this question we developed a unique stimulus set of movies showing real-world physical events that unfold over time. These movies allowed us to manipulate congruency between the temporal structure of the auditory and visual signals independently from the semantic labeling of the same auditory and visual signals. As a result, our study effectively isolated the separable effects of both signal-based and high-level cues and, in doing so, identifies the neural substrates independently and commonly recruited by these two contributors to the process of multimodal integration.

3. Chapter 3:

Interplay of Temporal & Semantic Congruency in Multimodal Events

3.1 Introduction

In Experiment 1, a pairwise comparisons of three multimodal conditions revealed separable networks of brain regions for two integration cues, time and semantics. These three comparisons, a congruent TCSC condition, a temporally incongruent TISC condition, and a semantically incongruent TISI condition, identified regions within the well-established “multimodal network” (Driver and Noesselt, 2008). Additionally, overlap in a subset of the identified regions suggests that some brain areas are sensitive to only one cue, e.g., temporal synchrony independent of the other cue, e.g., semantic congruence.

However, the design of Experiment 1 leaves open a number of interesting questions that are the focus of Experiment 2. First, the two incongruent conditions were not equivalent in Experiment 1 because the two integration cues are in conflict in the TISC condition but not in the TISI condition. Recall that the movies for the TISC condition contained the visual and auditory track from the same event (semantically congruent) but were temporally asynchronous, while the movies for the TISI contained visual and auditory components from different events (semantically incongruent) and the timing of these events was also different (temporally asynchronous). Thus, in the TISC condition the semantic cue signals that the two modalities should be bound together into a

coherent percept while the temporal cue signals the opposite - however, no such conflict existed between the two integration cues in the TISI condition in that both time and semantics were incongruent, signaling that the visual and auditory streams do not form a coherent event. The impact of this asymmetry in intra-cue conflicts is unknown in Experiment 1.

Second, although the overlap of regions across the 3 AV comparisons suggests separable networks for each integration cue, a stronger test of this neural organization would be a complete 2x2 factorial design that included a missing condition in Experiment 1: that is, where the auditory and visual information are temporally synchronous but semantically incongruent. A factorial design can identify brain regions that show a main effect for each of the two cues, providing the best evidence for subregions that show specialized processing for a particular integration cue. A full factorial design also better addresses whether any brain regions show an interaction between the two cues. For example, there may be functional brain areas that resolve conflicts between integration cues, that is, determining if the modalities form a coherent percept and should be bound or not.

Third, in Experiment 1 the nature of the temporal incongruence varied among the twenty events used as stimuli. More specifically, to accomplish the objective of ecological validity, we relied on real-world instances of events with natural variation in the rhythm of the action and overall event timing. Thus, when the audio tracks for the two event exemplars were swapped for playback with the video footage, the temporal structure of the mismatched was incorrect, but this mismatch was not systematically incorrect. This natural variability in the temporal structure of the event exemplars led to

two main types of temporal incongruence among the events. (1) For some events, the tempo/frequency of the action varied between the two exemplars, such as fast door knocking versus slow door knocking. (2) Others varied based on the phase offset between modalities, such that you would see the action before you heard it or hear it before you see it, and the amount of offset varied across the events. In addition, the amount of time between the onset of the movie and the first evidence for temporal asynchrony between the auditory and visual modalities was random, so that the amount of exposure time for each movie for the subject to know that it was temporally incongruent also varied between each pair of event exemplars. Although one can argue that this variation in the amount of exposure time needed to identify temporal incongruence is not critical in our fMRI study (as it takes longer to acquire a full brain volume, 3 seconds, than present the full length of the event, 2.5 seconds), equating exposure time is essential for an EEG design and analysis as reported in Chapter 4 (in that EEG acquires neural data with 4ms resolution). In summary, the benefit of all of this variation in the temporal structure of the events for Experiment 1 ensured that many factors contributed to the overall temporal incongruence across the events and reflected the inherent variability of real-world events; however, the ecological validity is a trade-off with experimental control. Namely, the variability in temporal structure limits how precisely functional roles can be spelled out for the brain regions identified as sensitive to temporal cues (in that, the frequency and/or the phase offset between the auditory and visual modalities varied across stimuli).

In Experiment 2 the experimental design addresses each of these three concerns by generating and using a new set of event stimuli. As discussed below in the Methods section, these 18 new event movies contained actions that involved multiple, discrete

impacts during the 2 second duration of the movie in order to facilitate the movie editing required to complete the full 2x2 factorial design with four conditions:

CONDITION	Temporally Congruent	Semantically Congruent?
1 TCSC (original movie)	Yes	Yes
2 TISC	No	Yes
3 TCSI	Yes	No
4 TISI	No	No

Events were chosen such that the start and end of each impact could be clearly identified; for example, an event like breaking glass would be difficult to edit for the temporally asynchronous condition in that each break follows in quick succession and the timing of each break would be hard to visibly discern. Consequently, the event would be a continuous event of successive glass breaks rather than containing discrete impacts where the sight of a given piece of glass breaking could be clearly offset in time from the sound of the break. In contrast, an event such as dropping ice cubes in a clear glass was filmed such that several ice cubes fell, one at a time, into a glass, with a pause between successive ice cubes to provide temporal windows where no movement is seen or heard. These temporal windows allow the visual impact to be offset from the auditory impact for the temporally incongruent condition. Given the precision such stimuli afford, in Experiment 2 the temporally incongruent condition was generated by a consistent phase offset of 500 milliseconds. Events were edited so that either the visual impact was seen 500 milliseconds before the corresponding auditory impact was heard or the reverse – the visual impact was seen 500 milliseconds *after* the auditory impact was heard. This

constrains any brain regions identified as sensitive to the temporal relationship in Experiment 2 as specifically processing the phase relationship between the auditory and visual modalities.

Discrete impacts within each event are also essential for generating the two semantically incongruent conditions. More specifically, since all stimulus events have discrete impacts, the auditory sound for one impact can be extracted from the waveform of one event and then edited to occur at the precise time of a visual impact for another event, creating the condition omitted from Experiment 1: where the event is temporally synchronous but semantically incongruent (Condition 3 shown above). It is also simple to edit the auditory impact from one event to occur at a time that is offset from the visual impact of the other event, thereby creating a semantically and temporally incongruent event that has the same phase offset (i.e., 500 milliseconds).

In short, the discrete impacts in each stimulus event in Experiment 2 allow temporal and semantic relationships to be better disentangled from one another. More generally, Experiment 2 exploits the power of a factorial design to validate and better elucidate the separable functional networks for two cues to multimodal integration, time and semantics, as well as more effectively identify any regions that are sensitive to interactions between the two cues.

3.2 Methods & Materials

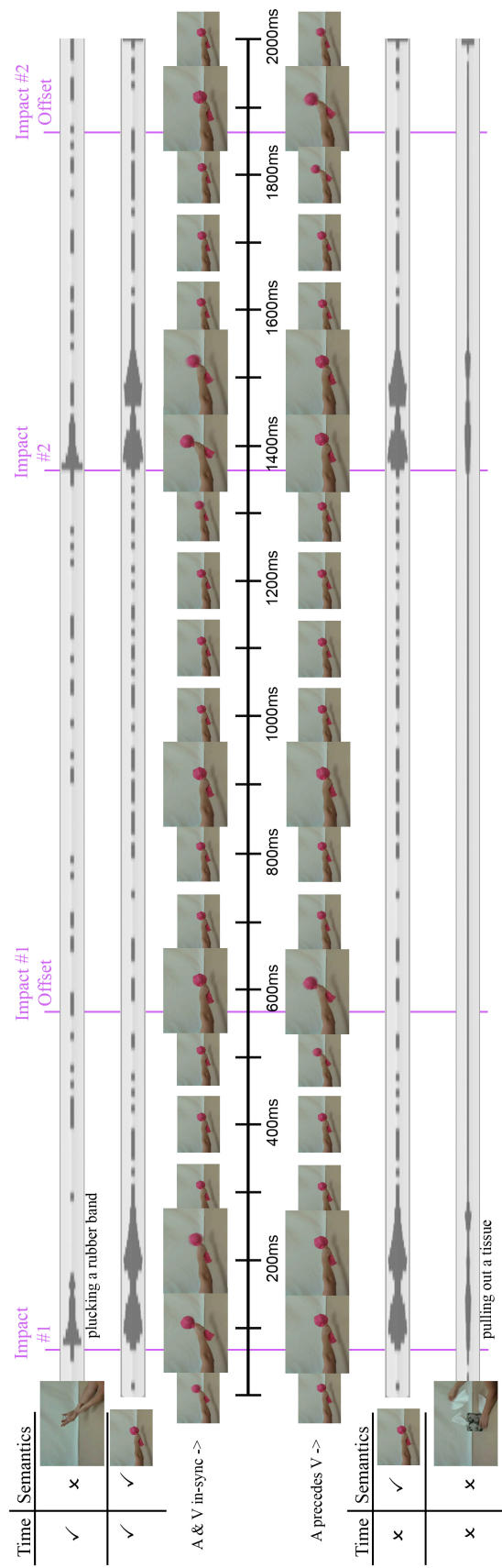
Participants. Sixteen right-handed subjects (9 female/7 male; mean age: 23.6; range: 19-33) participated in the study, but data from two subjects (both female) were discarded due

to excessive head movement during the scanning session (one subject coughing, the other had total movement of 16mm during the session). Subjects were paid for their participation and all provided written, informed consent consistent with and approved by Brown University IRB. All had normal or corrected-normal vision and hearing, and they reported that they were neurologically healthy with no contraindications for MRI scanning. The study was conducted at the Magnetic Resonance Facility at Brown University (Providence, RI).

Stimuli. A total of 36 AV movies were digitally filmed at standard definition for this study, including two exemplars of each of 18 unique real-world, physical events that had multiple, discrete impacts (e.g., tapping a pencil, shaking a maraca, squeezing a squeeze toy, etc). In each movie, only the arm and hand of the actress were seen performing the impact events (Figure 7). The movies were edited using Final Cut Pro (Apple Co., Cupertino, CA) to be 2000 msec in duration and exported as 720x480 MP4 movies with 16 bit, 44.1kHz stereo audio.

Figure 7 *Stimulus Design*

One event, shaking a maraca, is depicted with a timeline that highlights the timing of the visual and auditory impacts in the four multimodal conditions. The top two rows show the two temporally congruent conditions, where the visual frame shows the impact occurring at the same time as the waveform. In the first row, the impact was extracted from another event, plucking a rubberband, to generate a TCSI condition. In the second row, the waveform is the one originally recorded and used in the TCSC condition. The bottom two rows show the temporally incongruent condition where the auditory impact preceded the visual impact by editing the start of the video to be 500ms before the impact occurred. In the third row, the waveform is the same one as TCSC to generate that TISC condition. The fourth row, the impact was extracted from another event, pulling out a tissue, and edited to occur with the same 500ms offset to generate the TISI condition. The enlargement of some of the movie frames is for illustrative purposes only.



Experimental Conditions. Each of the 36 events was presented in six conditions, two unimodal, auditory only (A_{only}) and visual only (V_{only}), and four multimodal conditions (TCSC, TISC, TCSI, TISI) that fully crossed temporal congruency and semantic congruency between the auditory and visual modalities.

In each movie, the event depicted consisted of 2-4 discrete impacts, allowing the movies to be edited so that the temporal and semantic relationships could be disentangled from one another. For the temporally synchronous conditions, the visual impacts occurred at the same time as the auditory impacts, but the semantics could be manipulated by either playing the original sound with the movie or by extracting an impact from the auditory track of another event and editing it to occur at the same time as the visual impact. For the temporally incongruent conditions, the movies were edited so that the impacts were temporally offset by 500 milliseconds. Half of the movies were edited so that the auditory impact preceded the visual impact by 500 milliseconds and half where visual impact preceded the auditory impact by 500 milliseconds. Again, the auditory impact was either taken from the same movie (semantically congruent) or extracted from another event and edited to occur with a 500 millisecond offset (semantically incongruent). An example of a temporally asynchronous trial where the auditory impact precedes the visual impact is shown in Figure 7.

Consequently, there were four multimodal conditions. The TCSC movies showed audio and video that were temporally congruent and from the same event. The TISC movies showed audio and video that were temporally incongruent, with a phase offset of 500 milliseconds, but both the audio and video came from the same event (e.g., shaking a maraca). The TCSI movies showed the video impacts from one event (e.g., shaking a

maraca) in temporal congruency with the sound from a different event (e.g., knocking over dominos). The TISI movies showed audio and video that were temporally incongruent, phase offset of 500 milliseconds, and the audio impact was from one event (e.g., shaking a maraca) and video came from a different event (e.g., knocking over dominos).

Experimental Task. Prior research has suggested that making an explicit congruency judgment can effect the obtained pattern of neural responses (van Atteveldt et al., 2007), so here we had subjects perform a target detection task that was independent of the congruency manipulations.

A trial began with a 200 millisecond red fixation cross to cue the start of the 2000 millisecond event (auditory only, visual only, or multimodal) followed by an 800 millisecond green fixation cross that indicated that the subject should respond whether they detected a target during the movie. The target was either visual or auditory to ensure that the participant was attending to both modalities, and the target could occur at anytime during the movie. The visual target was a purple asterisk that always appeared near the action of the movie (to minimize eye movements) and the auditory target was a middle C tone (400 Hz) that was heard for 500 milliseconds. Subjects made a yes/no response for every trial, using their left or right index finger. Response assignment was counterbalanced between subjects. As expected, accuracy was equally high for all five conditions across both modalities (97% +/- .02%).

Experimental Procedure. Before the scan, each subject practiced the three tasks using a set of movies that were not seen during the actual experiment. Subjects then studied a list of the 18 unique physical event types to appear in the experiment, performed a written recall task, and then read aloud the name of any event type not recalled. This procedure ensured that all events were familiar and identifiable to all subjects during the scanning session.

During the scanning session, high resolution anatomical MP-RAGE images were collected first, followed by four functional EPI scans using a fast event-related design. Each functional scan lasted 5 minutes and 18 seconds and consisted of 104 trials: 60 experimental trials and 44 baseline fixation trials. Within the 60 experimental trials, 54 were no target trials (9 each of the 6 experimental conditions: A, V, TCSC, TISC, TCSI, TISI) and 9 were target trials (randomly selected from the six conditions with at least one from each). The trial order and number of baseline trials for the functional scans were optimized using the genetic algorithm (Wager and Nichols, 2003). After completing four scans of this task, the subjects completed additional scans that are not discussed in this chapter.

fMRI Details. MRI data were acquired using a 3T Siemens TIM Trio scanner (Erlangen, Germany). Visual stimuli were presented through a rear projection system with a mirror mounted to the head coil, and audio stimuli presented through an Avotec SS-3100 Silent Scan audio system (Stuart, FL). Participants made behavioral responses with their left and right index fingers using the two outside buttons on a 4 button response pad (Mag Design & Engineering, Sunnyvale, CA). The high resolution MPRAGE scan includes

160 interleaved slices with 1mm isotropic voxels in a 256x256 matrix with a TR=1900msec, TE=2.98msec, and flip angle=9degrees, and the T1-weighted echoplanar images (EPI) covered the whole brain with 48 interleaved slices, 3mm isotropic voxels, TR=3000msec, TE=28msec, and flip angle=90degrees. The EPI scans for the TargetDetection task consisted of 104 volumes. Prior to analysis in AFNI (Cox, 1996), the functional data were pre-processed in AFNI to correct for slice timing differences and 3-D head movement, smoothed with a FWHM kernel of 1 voxel (3mm), and normalized to allow the beta coefficients to reflect percent signal change by dividing each time series by its mean.

fMRI Data Analysis

Prior to warping the data to the standard atlas (Talairach and Tournoux, 1988), the data for each individual was fit with a general linear model (GLM) to estimate the activity at each voxel using a regressor for each of the six experimental conditions (two unimodal and four multimodal conditions) as well as regressors of no interest to account for third-order polynomial trends. A gamma hemodynamic response function was assumed. The output of each individual participant's GLM was then warped to the standard atlas.

For the whole brain analysis, the group data was analyzed with a two-way within-subject (repeated-measures) ANOVA that included the time and semantics as fixed factors and subjects as a random factor. Activation maps were generated to show regions with a main effect time, a main effect semantics, or an interaction of two factors. All statistical maps were thresholded at $p < 0.05$ with a requirement that all regions of

interest (ROIs) have at least 24 voxels (connected by 3mm, i.e., one side of a voxel) to ensure an alpha level of 0.05 (AFNI's AlphaSim program written by B. Douglas Ward, run with 5000 Monte Carlo simulations).

The whole brain analysis was complemented with an unbiased region of interest analysis in order to subject the data to a more stringent statistical analysis. As depicted in Figure 8, a map was first generated of all task active voxels using a conjunctive “or” so that a voxel had to be active in a least one of the four conditions to be included (the orange voxels in Figure 8). Next, the max value for each of the task active voxels was used to identify a voxel with the peak activity that was closest to the regions of interest in the whole-brain analysis (the peak voxel closest to the L IFG region is shown in Figure 8). This peak voxel was then used as the centroid for a 9mm sphere-ROI to approximate the regions identified in the whole-brain analysis, and only the task active voxels within this sphere were used in the subsequent 3-way ANOVA of sphere-ROI, time, and semantics. This analysis was conducted on six regions from the frontal cortex and six regions from the posterior cortex. In both groupings, the 3-way interaction was significant, so 2-way ANOVAs of time and semantics were conducted within each sphere-ROI to investigate the nature of the 3-way interactions. In short, these 2-way analyses revealed the response profile of each brain region. To correct for multiple comparisons for the 2way ANOVAs computed in the 12 sphere-ROIs (denoted as m), a false discovery rate was computed using an alpha (α) level of .05 in the equation: $\alpha(m+1)/2m$, giving an FDR, corrected for 12 tests, of $q=.0271$ (Benjamini and Yekutieli, 2001).

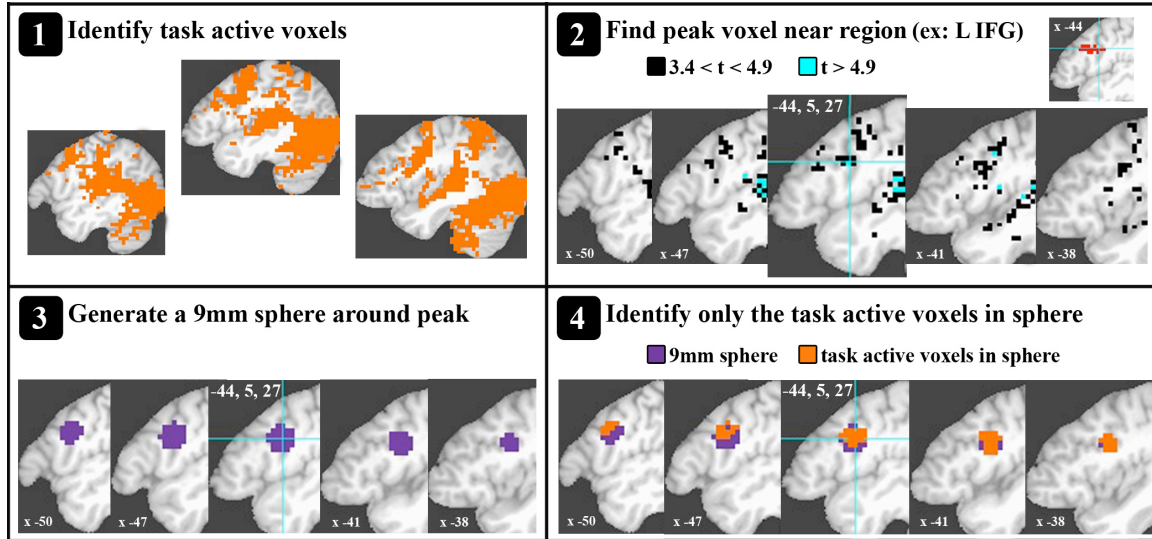


Figure 8 *Region-of-Interest Analysis Method*

This figure illustrates the method used to identify the 9mm spheres in the region-of-interest analysis, using the L IFG region as an example. In the step 1, voxels that were active in any of the four conditions are shown in orange ($p < 0.05$, minimum cluster of 24 voxels). In the step 2, the task voxels are thresholded ($t > 3.4$, $p < 0.001$) and shown in black with the peak voxels ($t > 4.9$) shown in teal blue. The peak voxel closest to the L IFG is identified at $-44, 5, 27$. In step 3, a sphere with a 9mm radius was drawn around the peak coordinate and is shown in purple. In step 4, the task active voxels in the sphere are identified, colored orange, and only these voxels are used in the sphere-ROI by time by semantics ANOVA.

3.3 Results

A whole brain analysis revealed regions throughout the brain that were differentially sensitive to one of the two integration cues, semantic congruency or temporal synchrony, as well as regions that were sensitive to an interaction between the two integration cues. This analysis was then supplemented with a more stringent region-of-interest (ROI) analysis in order to explore how the four multimodal conditions modulated the activity in the regions identified by the whole brain analysis.

Whole-Brain Analysis: Main Effect of Semantics

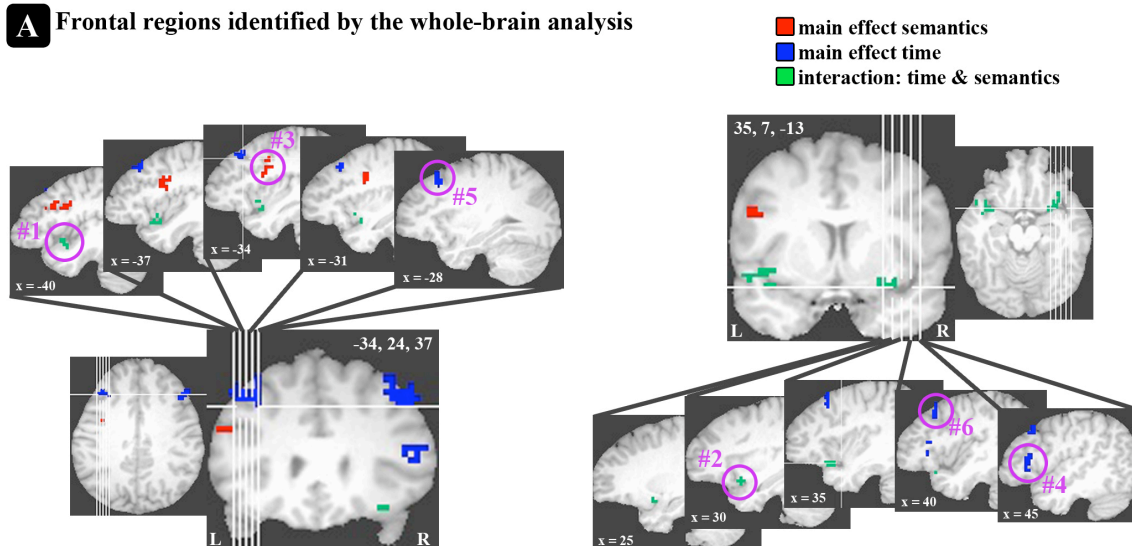
To identify regions with a main effect of semantics, the two multimodal conditions with congruent semantics between the auditory and visual modalities (TCSC and TISC) were contrasted with the two conditions with incongruent semantics (TCSI and TISI). Two regions were identified: one in left inferior frontal gyrus (pictured in red, Figure 9A) and one in right post-central gyrus (pictured in red, Figure 10A). A 2-way ANOVA within these regions confirmed that there was no interaction between time and semantics ($p > .05$). More details about these regions can be found in Table 2.

Figure 9 *Region-of-Interest Analysis: Frontal Cortex*

(A) The six frontal regions identified in the whole brain analysis ($p < 0.05$, minimum cluster of 24 voxels) are pictured on the brain from one subject in the study with the coordinates of the crosshairs listed on the coronal image. Regions identified with a main effect of semantics are shown in red, main effect of time in blue, and an interaction between semantics and time are shown in green. The regions are numbered with a circle around the region in one of the sagittal slices: #1 L inferior frontal gyrus (IFG)/anterior temporal region; #2 R inferior frontal gyrus (IFG)/anterior temporal region; #3 L IFG; #4 R IFG; #5 L middle frontal gyrus (MFG); #6 R MFG.

(B) For six regions identified by number in A, the results are shown for the time by semantics ANOVA in the sphere-ROI analysis, with significant effects shown in red for semantics, blue for time, and green for an interaction between semantics and time. Results that approach significance (less than $p < 0.05$ but do not survive the FDR correction) are listed in yellow, with non-significant results shown in black. Below the ANOVA results, a plot of the percent signal change is shown for each level of the two factors (i.e., response magnitudes for the four multimodal conditions: TCSC, TCSI, TISC, TISI).

A Frontal regions identified by the whole-brain analysis



B Neural responses in 9mm spheres that approximate 6 regions shown in A

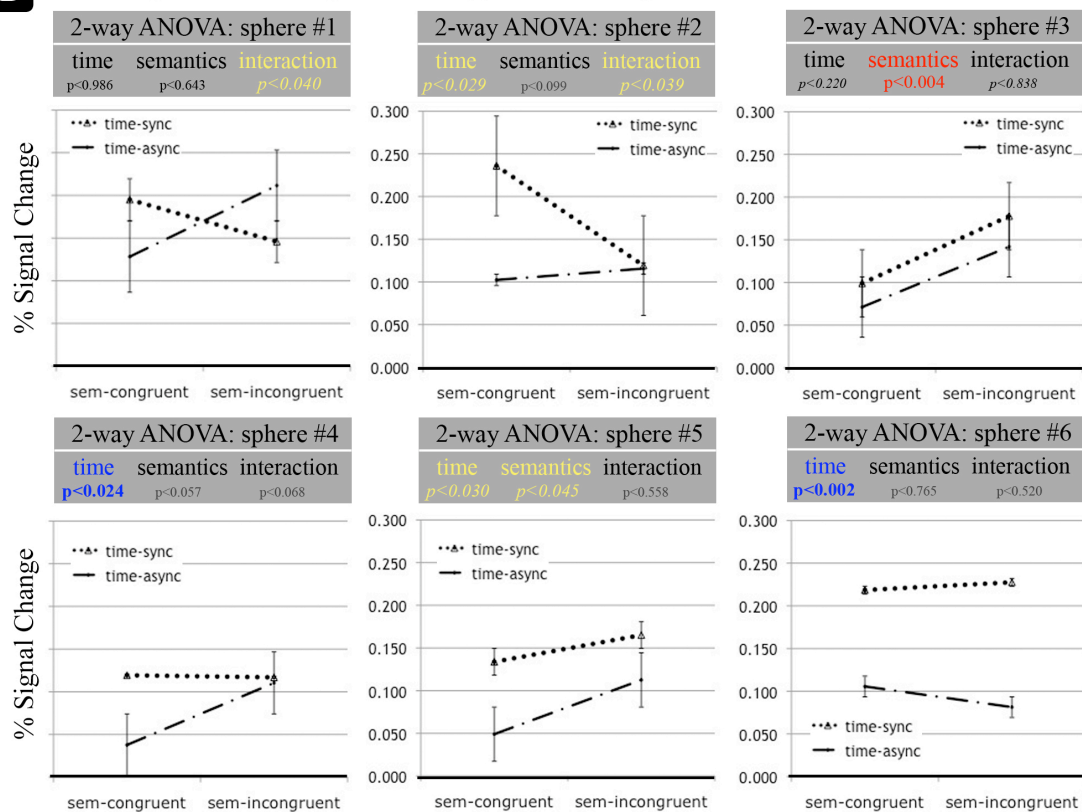
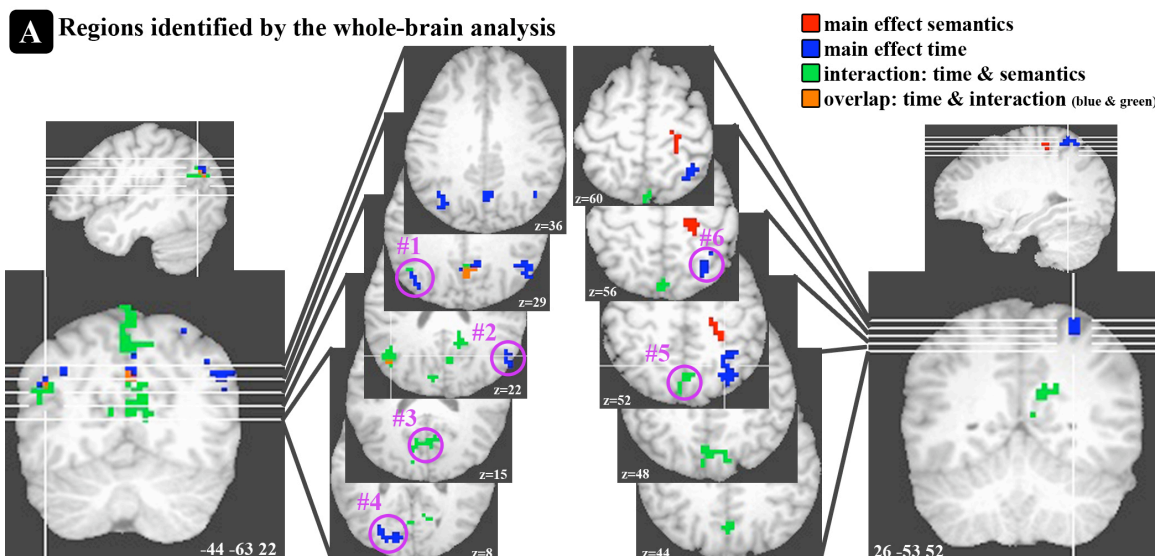


Figure 10 *Region-of-Interest Analysis: Posterior Cortex*

(A) Seven of the posterior regions identified in the whole brain analysis ($p < 0.05$, minimum cluster of 24 voxels) are pictured on the brain from one subject in the study with the coordinates of the crosshairs listed on the coronal image. Regions identified with a main effect of semantics are shown in red, main effect of time in blue, and an interaction between semantics and time are shown in green. The regions are numbered with a circle around the region in one of the sagittal slices: #1 L posterior middle temporal gyrus (MTG); #2 R MTG; #3 medial posterior cingulate; #4 L middle occipital; #5 medial precuneus; #6 R precuneus. The R post-central gyrus region (shown in red) is not labeled with a number because the sphere-ROI analysis could not be run in this region due to insufficient numbers of task active voxels in this region (see Results).

(B) For six regions identified by number in A, the results are shown for the time by semantics ANOVA in the sphere-ROI analysis, with significant effects shown in red for semantics, blue for time, and green for an interaction between semantics and time. Results that approach significance (less than $p < 0.05$ but do not survive the FDR correction) are listed in yellow, with non-significant results shown in black. Below the ANOVA results, a plot of the percent signal change is shown for each level of the two factors (i.e., response magnitudes for the four multimodal conditions: TCSC, TCSI, TISC, TISI).



B Neural responses in 9mm spheres that approximate 6 regions plotted above

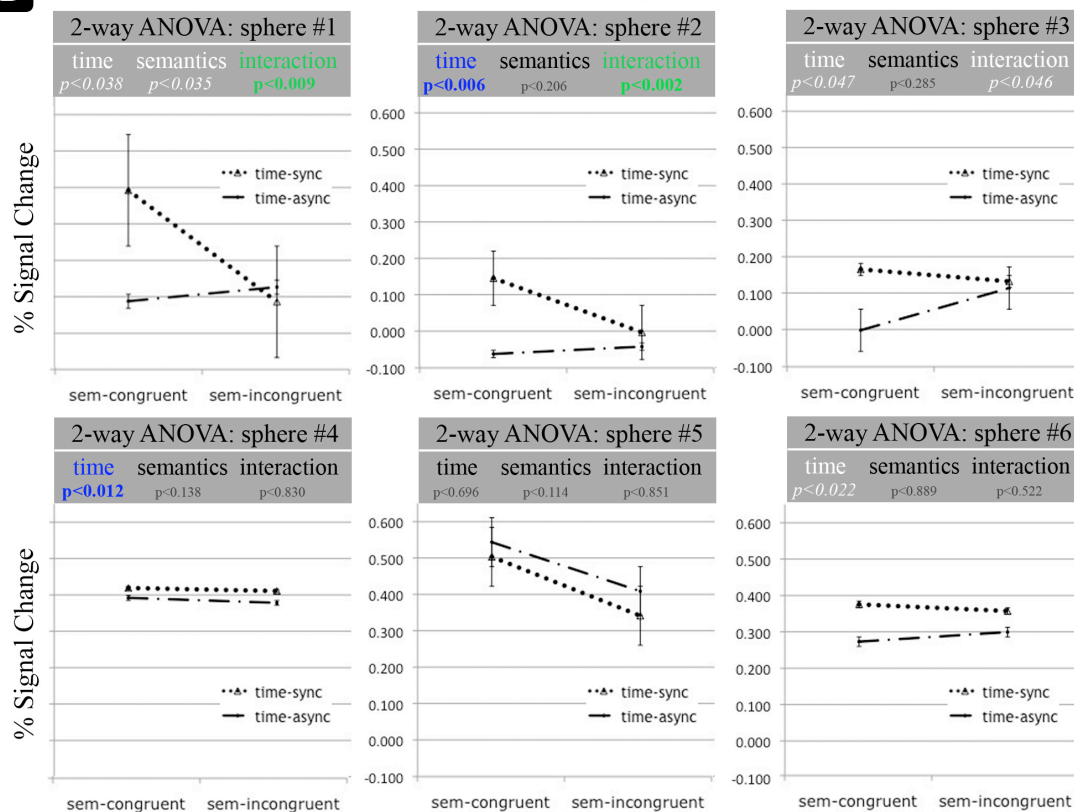


Table 2 Whole-brain Analysis: Anatomical Location, Size, and Peak Coordinates

Row#	Anatomical Location	Hemi.	Pictured	Overlap	# of Voxels	x	y	z
Main Effect: Semantics								
1	Inferior Frontal	L	Figure 9		69	39.4	-6	28.3
2	Post-central gyrus	R	Figure 10		31	-19.8	27.4	56.6
Main Effect: Time								
3	Inferior Frontal	R	Figure 9		31	-45.4	-26	10
4	Middle Frontal	L	Figure 9		41	31	-26.7	47.4
5	Middle Frontal	R			32	-32.2	-51.5	0.6
6	Middle Frontal	R	Figure 9		47	-43.2	-21.9	44.9
7	Medial Frontal	L/R			223	-3.4	-30.2	48.4
8	Middle Temporal	R	Figure 10		46	-49.3	63.4	28.6
9	Middle Temporal	L	Figure 10	row 24	32	38.4	66.9	32.8
10	Posterior Cingulate	L/R		row 25	30	-0.5	52.6	13.5
11	Precuneus	L/R	Figure 10		44	-1.3	61.6	33.4
12	Parietal	R	Figure 10		38	-29.4	52.6	55.6
13	Middle Occipital	L	Figure 10		24	20.3	80.9	8.2
14	Inferior Occipital	L			64	23.6	89.8	-10.4
15	Lingual	R			39	-0.9	83.9	-2.1
16	Lingual	L			35	7.9	76.5	-17.3
17	Caudate	R		row 31	100	-9	-0.9	13
18	Caudate	L			25	12.1	7.2	18.8
19	Cerebellum	L			101	34.1	72.3	-28.4
Interaction: Time & Semantics								
20	Inferior Frontal	L	Figure 9		33	39.1	-7.1	-9.5
21	Inferior Frontal	R	Figure 9		28	-30.3	-12.8	-12.6
22	Medial Frontal	L/R			76	-0.7	-52.9	6.3
23	Superior Frontal	L			28	16.7	-57.7	24.1
24	Middle Temporal	L	Figure 10	row 9	24	45	62.9	24.5
25	Posterior Cingulate	L/R	Figure 10	row 10	140	0.4	65.7	19.3
26	Posterior Cingulate				34	1.1	50.5	25.1
27	Precuneus				39	-2.5	89.6	21.8
28	Medial Parietal	L/R	Figure 10		59	1.3	64.7	56.6
29	Cingulate				47	0.7	30.9	39.2
30	Caudate	L			43	6.2	-3.1	0.1
31	Caudate	R		row 17	82	-2.1	-28.7	0.7

a. All regions were identified with a $p < .05$, minimum cluster size of 24 voxels, alpha .05

b. Abbreviations: Hemi. = Hemisphere; # of Voxels = Number of 3mm voxels in region

c. Overlap column lists the row number of the region that overlaps

Whole-Brain Analysis: Main Effect of Time

To identify regions with a main effect of time, the two multimodal conditions with temporal synchrony between auditory and visual modalities (TCSC and TCSI) were contrasted with conditions that were temporally asynchronous (TISC and TISI). A total of 17 regions were identified and details about each can be found in Table 2. These regions were distributed throughout the brain, including several in the frontal cortex, bilateral regions in posterior temporal cortex (pictured in Figure 10A), medial regions in posterior cingulate, precuneus, parietal cortex, and several regions in inferior lingual and occipital gyri. Bilateral regions were found in the caudate as well as a region in left cerebellum. Interestingly, the right inferior frontal region is almost bilateral with the left hemisphere region identified in the main effect of semantics contrast, and just superior to these inferior frontal regions, bilateral regions were also identified in the middle frontal gyrus (Figure 9A).

Whole-Brain Analysis: Interaction between Time & Semantics

Significant interactions between semantics and time were also identified in several regions throughout the brain, and details about these 12 regions are listed in Table 2. Several regions were found in frontal cortex, including bilateral regions in the inferior frontal cortex that were inferior to the bilateral regions identified in the semantics (left hemisphere) and time (right hemisphere) contrasts. Three of the regions also partially overlapped with regions identified in the main effect of time contrast: a region in the left posterior temporal cortex, one in the posterior cingulate cortex, and another in the right

caudate. Additional regions were also found in the parietal cortex and the cingulate cortex (superior & anterior to the posterior cingulate).

Region-of-Interest (ROI) Analysis

To supplement the whole-brain analysis, an ROI analysis was performed to subject the regions identified in the main effect and interaction activation maps to a more stringent statistical analysis (see Methods & Figure 8 for details). One set of six ROIs were selected in the frontal cortex and another set of six in the posterior cortex. These ROIs were chosen based on their similarity with the regions identified by the overlap analysis in Experiment 1 that included bilateral areas in middle frontal gyrus, a medial region in the posterior cingulate, and a medial region in the parietal cortex. Since these ROIs were identified based on the peak of the task active voxels, there is no bias for either of our two factors, time and semantics, and consequently, these ROIs allow us to independently test three things: (1) investigate whether the hemispheric preference for semantics in the left hemisphere and time in right hemisphere is reliable, (2) validate the main effects & interactions found with the whole-brain analysis, and (3) plot the response magnitudes for our four multimodal conditions and explore the underlying processing and computational constraints of each region.

Region-of-Interest (ROI) Analysis: Frontal ROIs

In the whole-brain analysis, bilateral regions in the inferior frontal gyrus (IFG) were found that were very similar to the bilateral frontal regions identified with an overlap of regions in the AV contrasts in Experiment 1, where the left showed a

preference for semantics and right showed a preference for time. Similarly, here in Experiment 2, the left IFG region showed a main effect of semantics and the right IFG showed a main effect of time. In addition, we found additional bilateral regions in the frontal cortex: a set in the middle frontal gyrus (MFG) that was superior to the bilateral IFG regions and a set that was inferior, located more medially in the IFG and spreading in to the anterior superior temporal gyrus. For the ROI analysis, the 9mm spheres were centered as proximal to these six regions as possible.

The six frontal ROIs were analyzed with a 3-way ANOVA of time, semantics, and ROI. The results showed a main effect of time ($F(1,13) = 9.575, p < 0.0085$), an interaction between time and ROI ($F(5,65)=4.006, p<0.0031$), an interaction between semantics and ROI ($F(5,65)=0.0013, p<0.0013$), and a 3-way interaction ($F(5,65)=2.788, p<0.0242$). None of the other effects were significant: main effect of semantics $F(1,13)=1.23, p>0.05$; main effect of ROI $F(5,65)=0.85, p>0.05$; interaction of time and semantics $F(1,13)=2.92, p>0.05$.

Since the 3-way ANOVA identifies differential effects across the six ROIs, a 2-way ANOVA of time and semantics was performed to further explore the nature of the 3-way interaction in each of the six ROIs. The results of the 2-way ANOVAs in each of the ROIs is shown in Figure 9B, and the results are colored to reflect the significance: blue if the main effect of time is significant, red if main effect of semantics is significant, green if interaction is significant, and yellow if the result is less than $p < 0.05$ but fails to survive the FDR correction for multiple comparisons among the 12 ROIs (6 frontal and 6 posterior). Figure 9B also contains a graph for each ROI that plots the percent signal

change for each of the four conditions to explore the processing constraints of the region for a multimodal stimulus.

The two most inferior bilateral frontal regions, that extend in to the anterior temporal lobe, both approach significance for a 2-way interaction of time and semantics (L ROI [Figure 9B #1]: $F(1,13)=5.17$, $p<0.040$ and R ROI [Figure 9B #2]: $F(1,13)=5.26$, $p<0.039$), matching the whole-brain analysis that identified these regions in the interaction contrast. In addition, the region of interest analysis reveals that a main effect of time also approaches significance in the right ROI [Figure 9B #2]: $F(1,13)=6.04$, $p<0.029$. The plots of the response magnitudes for the four multimodal conditions in these bilateral ROIs show a similar response profile, with the strongest activity when both integration cues are either congruent or both cues are incongruent. These ROIs are less active when the cues are in conflict with one another, e.g., temporal *congruence* in conflict with semantic *incongruence*. Interestingly, in the right ROI, the response is dominated by a stronger response for the temporally synchronous and semantically congruent condition, that is, the condition that shows intact movies of real-world events.

Superior to these ROIs, the bilateral regions in the IFG both confirmed their main effects from the whole-brain analysis in the 2-way ANOVA: the L ROI [Figure 9B #3] revealed a main effect of semantics ($F(1,13)=12.09$, $p<0.004$) and the R ROI [Figure 9B #4] revealed a main effect of time ($F(1,13)=6.58$, $p<0.024$). As expected from prior research (Belardinelli et al., 2004; Laurienti et al., 2003; Hein et al., 2007), the semantically incongruent conditions were more strongly activated than the congruent conditions in the L ROI. The pattern of activation in the R ROI showed equivalent

activation among 3 of the 4 conditions, except for less activation for the temporally incongruent but semantically congruent condition.

Moving dorsally again, the final two ROIs are located bilaterally in the middle frontal gyrus, with the L ROI [Figure 9B #5] approaching significance for a main effect of time (matching the whole-brain analysis, $F(1,13)=5.96$, $p<0.030$) as well as approaching significance for a main effect of semantics ($F(1,13)=4.95$, $p<0.045$). The R ROI [Figure 9B #6] matched the main effect of time result from the whole-brain analysis ($F(1,13)=14.87$, $p<0.002$). Both regions show greater activity for the temporally synchronous conditions, with the left region also showing more activation for the semantically incongruent conditions than the semantically congruent conditions (the same expected pattern for the semantic relationships as found in the L IFG region located just inferior to this region).

Region-of-Interest (ROI) Analysis: Posterior ROIs

The whole brain analysis also identified similar posterior regions to those found in the overlap analysis for Experiment 1, namely the posterior cingulate and the parietal activity focused in the medial and right precuneus. Consequently, for the posterior analysis, we included the posterior cingulate, a middle occipital region just adjacent to the posterior cingulate, two of the three parietal regions, and the bilateral posterior temporal regions. The temporal regions were included based on the central role of the posterior temporal region in prior studies, as discussed in Chapter 2 with the functionally-defined integration regions, so it seemed relevant to investigate their role in this study. Unfortunately, we were unable to include the region identified with a main effect of

semantics in the whole-brain analysis because this region was not near task active voxels, so the 9mm sphere centered near this whole-brain region only had 1 voxel when the sphere was constrained to only the task active voxels (Fig 8, step 4). The middle occipital region was thus included based on its proximity to these regions in order for the posterior analysis to have the same number of regions as the frontal analysis. As before, the 9mm spheres for the sphere-ROI analysis were centered as proximal to these six regions as possible.

The six frontal ROIs were analyzed with a 3-way ANOVA of time, semantics, and ROI. The results showed a main effect of ROI ($F(5,65)=8.547$, $p<0$), an interaction between time and semantics ($F(1,13)=8.56$, $p<0.012$), an interaction between semantics and ROI ($F(5,65)=2.75$, $p<0.026$), and a 3-way interaction ($F(5,65)=2.45$, $p<0.043$). None of the other effects were significant: main effect of time $F(1,13)=3.62$, $p>0.05$; main effect of semantics $F(1,13)=3.45$, $p>0.05$; and an interaction between time and ROI $F(5,65)=1.33$, $p>0.05$.

Since the 3-way ANOVA identifies differential effects across the six ROIs, a 2-way ANOVA of time and semantics was performed to further explore the nature of the 3-way interaction in each of the six ROIs. Similar to the ROI analysis of the frontal regions, the results of the 2-way ANOVAs in each of the ROIs is shown in Figure 10B, and the results are colored to reflect the significance: blue if the main effect of time is significant, red if main effect of semantics is significant, green if interaction is significant, and yellow if the result is less than $p<0.05$ but fails to survive the FDR correction for multiple comparisons among the 12 ROIs (6 frontal and 6 posterior). Figure 10B also contains a

graph for each ROI that plots the percent signal change for each of the four conditions to explore the processing constraints of the region for a multimodal stimulus.

The bilateral posterior temporal regions both show an interaction (L ROI [Figure 10B #1]: $F(1,13)=9.46$, $p<0.009$ and R ROI [Figure 10B #2]: $F(1,13)=14.61$, $p<0.002$). Both of the main effects for time and semantics approach significance in the L ROI (time $F(1,13)=5.33$, $p<0.038$; semantics $F(1,13)=5.55$, $p<0.035$), while only the main effect of time is significant in the R ROI ($F(1,13)=10.88$, $p<0.006$). The response profiles for the two regions are almost identical, showing the highest activation for the real-world combination of temporally synchronous and semantically congruent AV stimuli. The different result in the left region for a main effect of semantics appears to be driven by a more significant trend for a stronger decrease for the semantically congruent but temporally asynchronous stimuli.

The posterior cingulate ROI [Figure 10B #3] approaches significance for both a main effect of time ($F(1,13)=4.83$, $p<0.047$) and an interaction between time and semantics ($F(1,13)=4.88$, $p<0.046$), which matches the whole-brain analysis that found a posterior cingulate region showing an interaction that was mostly inferior but with a small overlap with the posterior cingulate region identified for a main effect of time. This slightly different localization of the whole-brain regions may contribute to the weaker effect in the sphere-ROI analysis due to partial overlap with these two slightly different regions in the whole-brain analysis. The response profile in this region looks very similar to the posterior temporal regions, with the strongest activity for the real-world temporally and semantically congruent condition.

Adjacent to the posterior cingulate ROI, the middle occipital cortex ROI [Figure 10B #4] shows a main effect of time ($F(1,13)=8.51$, $p<0.012$), just like the whole-brain analysis, and the response profile shows greater activity for the temporally synchronous events.

The two final regions included in the posterior ROI analysis were located in the parietal cortex. The medial precuneus ROI [Figure 10B #5] was the only ROI to reveal *no significant* effects in the 2-way ANOVA ($p>0.05$). Curiously, this ROI was the most similar to location of the overlap region identified in Experiment 1, so the absence of an effect in Experiment 2 is somewhat surprising and may be connected to the limited amount of overlap between the whole-brain analysis region and the task active voxels in the sphere-ROI in this analysis. The right-lateralized precuneus ROI [Figure 10B #6] matched the whole-brain analysis, showing a main effect of time ($F(1,13)=6.77$, $p<0.022$), driven by greater activation for the temporally synchronous conditions.

3.4 Discussion

This study replicates the major findings of Experiment 1 that time and semantics recruit separable networks of brain regions for processing each type of cue integration, including a preference in the left hemisphere for semantics and a preference in the right hemisphere for time. Furthermore, these results survive a more stringent statistical region-of-interest analysis that identifies a significant 3-way interaction between ROIs, temporal synchrony, and semantic congruency among six bilateral ROIs in the frontal cortex as well as six ROIs in posterior cortex. The response profiles of these regions help

to reveal the underlying functional computations performed by different brain regions to support integration of real-world, dynamic events. While the four most superior of the frontal regions show main effects for the two integration cues, the two inferior frontal regions that extend in to the anterior temporal lobe, as well as the bilateral posterior temporal regions and the posterior cingulate, reveal interactions among the cues, with the real-world TCSC condition typically showing the strongest activity. The parietal activity was less robust, and thus, the role of these regions requires additional research to explore their functional significance.

This discussion focuses on the similarity between the network of regions identified by Experiment 1 and the regions identified here. We also include a comparison of the activation response profiles for four multimodal conditions to further speculate about the underlying computations associated with different brain areas, as well as potential hemispheric preferences for different integration cues.

Convergence between Experiments 1 & 2

Despite the experimental design differences, Experiments 1 and 2 exhibit strong convergence for the network of brain regions involved in processing temporal and semantic relationships between auditory and visual modalities as well as the underlying neural activation response profiles in these areas.

Frontal Regions

Four of the six frontal regions used in the region-of-interest analysis overlap with regions identified by the AV contrasts in Experiment 1, while the remaining two frontal

regions are new to Experiment 2.

The right inferior frontal region that extends in to the anterior temporal region and approaches significance for a main effect of time and an interaction (Figure 9B #2) is similar to the right anterior temporal region identified in the Temporal Processing Contrast in Experiment 1 (Figure 4, R aMTG), both showing greater response for temporally and semantically congruent events compared to temporally synchronous but semantically incongruent events. The region in the left hemisphere is new in Experiment 2 and may reflect more power to detect semantic congruence with the addition of the TISC condition that reveals the interaction of time and semantics in these inferior frontal regions, especially given the expected role of the anterior temporal lobe in semantic processing (Pobric et al., 2007).

Similarly, the two most superior bilateral regions in the middle frontal gyrus only match in one hemisphere between Experiments 1 and 2. Here, it is the left hemisphere region (Figure 9B #5) that matches a region in the Congruency Processing contrast (Figure 6, L MFG), showing the same response profile with greater activity for TISI than TCSC, while the bilateral region in the right hemisphere is new in Experiment 2. The new region shows a main effect of time, so it may reflect the refinement of the temporal incongruence in this second study to rely on a phase offset between the impacts in each modality rather than a mixture of temporal incongruence parameters used in the first study (see Introduction for more details).

Finally, both of the bilateral inferior frontal regions located in the middle of these other four frontal regions match the regions identified in the Overlap analysis in Experiment 1 (Figure 3B), and again, the left hemisphere shows a preference for

semantics and the right reveals a preference for time. Again, the addition of the TISC condition in Experiment 2 may have enhanced the localization of the semantic processing, as the L IFG region in Experiment 2 is larger than Experiment 1 and extends more anteriorly; however, in both experiments, the response profile shows more activity for semantically incongruent conditions than semantically congruent ones. This left hemisphere preference for semantics converges with decades of research on language processing in the brain (Josse and Tzourio-Mazoyer, 2004) as well as prior multimodal research on events (Lewis et al., 2005; Doehrmann et al., 2008). The higher response for semantically incongruent compared to congruent also converges with prior multimodal research (Hein et al., 2007; Taylor et al., 2006; van Atteveldt et al., 2007).

The right hemisphere preference for time has also been suggested in prior research on the temporal onset synchrony of simple stimuli, such as flashes and beeps (Bushara et al., 2001; Calvert et al., 2001). The response profile does differ between Experiments 1 and 2 for the regions sensitive to the manipulation of temporal synchrony. For the regions identified in Experiment 1, the temporally asynchronous stimuli elicited a stronger response than the synchronous stimuli, while Experiment 2 reveals the opposite with synchronous greater than asynchronous. This difference is true for all regions sensitive to time, not just in the frontal areas, but in the temporal areas as well (with the exception of the anterior temporal region in Experiment 1 which may be more semantically-driven as discussed above). As discussed in the Introduction to this chapter, the nature of the temporal incongruence did differ between the two experiments, so it is possible that the phase offset between the auditory and visual modalities used in Experiment 2 as compared to the mixture of frequency and phase difference in

Experiment 1 drives this difference between the studies. However, future research is needed to better understand this difference in the response profile.

In short, the six frontal regions identified in Experiment 2 converge with both the anatomical location and response properties of the frontal regions in Experiment 1. The two frontal regions novel in Experiment 2 may result from the inclusion of the TISC condition, which boosted the sensitivity to semantic congruence, while the differences in the response profile for the regions sensitive to time may reflect the refinement of the temporal asynchrony here in Experiment 2 to be based on a fixed phase offset of 500 milliseconds for all events.

Posterior Regions

Similar to the group of frontal regions, four of the six posterior regions used in the region-of-interest analysis overlap with regions identified by the AV contrasts in Experiment 1, while the remaining two posterior regions are new to Experiment 2.

The two regions in posterior temporal cortex are similar to regions identified in Experiment 1. The left region in Experiment 2 (Figure 10B #1) overlaps the temporal region identified in the Overall Processing contrast in Experiment 1 (Figure 6 L MTG), and the two regions show the same response profile with TCSC showing greater activation than TISI though the absolute values differ. The right region in Experiment 2 is actually a bit superior and posterior to the right region identified in Experiment 1 in the Temporal Processing contrast (Figure 4 R pMTG), although both reveal a sensitivity to time. Again, the different localization of the regions may reflect the mixed types of temporal incongruence in Experiment 1 compared to a fixed phase offset used in the

temporally asynchronous condition in Experiment 2.

The posterior cingulate region in Experiment 2 matches the location of the overlap region in Experiment 1. This region showed a mixture of responses in Experiment 1 which may be reflected in the interaction shown in Experiment 2. In addition, the Experiment 1 overlap suggested a preference for time in the superior half of the identified region, and again, this matches Experiment 2 where the whole brain analysis found a main effect of time to be slightly superior to the region identified in the interaction activation map. The region-of-interest analysis map further confirms the tendency for both an interaction as well as a main effect of time in the posterior cingulate. Both studies find more activity for TCSC than the incongruent conditions.

The middle occipital region in Experiment 2 is not seen in Experiment 1; however, its response profile indicates a main effect of time, and thus, it may reflect the detection of the phase offset between the timing of the impacts in the auditory and visual modalities. Although they were not explored in the region-of-interest analysis, Experiment 2 also identified additional regions in the occipital cortex (listed in Table 2) that showed a main effect of time. Likewise, Experiment 1 found regions throughout the occipital cortex that showed stronger response to the TISC condition compared to either the TCSC or the TISI conditions. Once again, most of these regions were localized to different subregions of the occipital cortex, except for one region in the right Lingual Gyrus (Table 2 row 15 & Figure 4 R Ling G.). This suggests that different types of temporal incongruence may be detected in different sub regions of cortex, but to our knowledge, only two fMRI studies have looked at temporal relationships between modalities (Bushara et al., 2001; Calvert et al., 2001), focusing on onset synchrony

between arbitrary visual objects and auditory sounds which lack the richness of the dynamic temporal structure in the events used here. Recent behavioral research has highlighted that detection of a temporal structure between modalities does not depend on perfect onset synchrony for each of the flash/beep pairs (Denison, Driver, & Ruff, 2009). Consequently, future research is needed to explore the neural substrates supporting different types of temporal incongruence between modalities (onset synchrony, temporal structures with different tempos, etc).

Finally, the regions in parietal cortex provide a much less clear description of the role this brain area may play in multimodal integration. The medial precuneus region identified in Experiment 2 (Figure 10B #5) partially matches the overlap region of Experiment 1 found in medial precuneus (Figure 3B), but the response of the Experiment 2 region did not survive the more stringent region-of-interest analysis, although the whole-brain analysis found an interaction between the two integration cues that is similar to the response profile found for the medial precuneus in Experiment 1. In addition, a right precuneus region was also identified in Experiment 2 that showed a main effect of time, but no equivalent region was found in Experiment 1. The region identified by the main effect of semantics activation map was just anterior to this right precuneus region in the post-central gyrus, but it was not included in the region-of-interest analysis because the method used to identify the statistically-unbiased ROIs depended on task active voxels which did not exist in this brain region. In short, the processing role of the regions in the parietal cortex were hard to characterize due to a lack of a robust activation response. A review by Doehrmann & Naumer (2008) suggests that the parietal cortex may supported selective and/or divided attention to each modality during a multimodal

stimulus, so perhaps the response is weaker here because no explicit attention was required for the experimental task other than to monitor for a visual or auditory target during each trial. Thus, the role of the parietal regions in the processing of time and semantic relationship between modalities is open for future studies to investigate.

In summary, the six posterior regions found in Experiment 2 display some convergence with Experiment 1. The posterior temporal and posterior cingulate regions are quite similar in location and response properties, while the variation in the location of the regions in the occipital cortex vary and may reflect particular types of temporal incongruence. The role of the parietal cortex is unresolved.

3.5 Conclusion

This study employed a 2x2 factorial design to investigate the interplay of temporal and semantic relationships between auditory and visual modalities during the integration of real-world environmental events. Overall, the present results converge with those of Experiment 1, revealing separable networks of regions sensitive to the two types of integration cues. Interestingly, there is an overall pattern in which we see a strong preference in the right hemisphere for temporal synchrony (in parietal, temporal, and frontal areas) and an equally strong preference in the left hemisphere for either semantics alone or the interaction of semantics with time. Thus, the use of better controlled movies of environmental events further reveals the interplay of two common cues to integration. As such, Experiment 2 advances our understanding of the neural underpinnings of the process by which modality-specific information is bound into coherent, multimodal event percepts.

4. Chapter 4:

Investigating the Timecourse of Multimodal Integration

4.1 Introduction

Experiments 1 and 2, both using fMRI, identified networks of brain regions that are separably recruited by time and semantics – two of the primary cues to integrating multimodal information. However, in that fMRI relies on the slow hemodynamic lag of the BOLD response, these two experiments do not address the processing timescale for each of these cues. As discussed in Chapter 2, the common temporal structure between the auditory and visual modalities is carried in the low-level information in perceptual signals, while the shared semantic associations across multimodal information are likely to depend on high-level experiential knowledge. Consequently, we posit that temporal synchrony plays an influential role in the early processing of a multimodal signal, while semantic congruence has a later influence on the integrative process. This hypothesis predicts that the TCSC and TCSI conditions in Experiment 2 prompt identical processing based on the temporal synchrony between modalities until the timepoint where high-level knowledge is retrieved and applied. However, once the semantic incongruence is identified, processing between the two conditions should diverge. Conversely, the TCSC and the TISC conditions are predicted to diverge early in processing based on the absence of the temporal synchrony. In that the modality-specific perceptual signals will continue to lack temporal synchrony (a characteristic of a percept with a common origin), it is unclear, once the signals have diverged, whether the modalities are bound at a later

timepoint when semantic congruence between them is identified. The goal of Experiment 3 is to test these predictions by employing EEG, a method that provides more fine-grained measurements of neural activity (sampled every 4 milliseconds versus the 3 seconds it takes to acquire a full brain volume in the two fMRI experiments). At the same time, as in the two fMRI experiments, the focus for Experiment 3 is on the interplay between semantic and temporal cues to integration during the processing of environmental events.

One of the first uses of EEG to explore the timecourse of multimodal processing relied on a metric used in single-cell recordings in the cat superior colliculus (Meredith and Stein, 1983): Is the response to the multimodal stimulus different than what is expected from the linear sum of the unimodal responses? Giard & Peronnet (1999) employed this metric in a study of multimodal object recognition and provided the first evidence of multimodal interactions using event-related potentials (ERPs). The auditory-visual interactions were measured with a difference wave between the ERP to the multimodal stimulus (an associated circle and tone) and the ERP to the sum of the auditory-only and visual-only ERPs. The study found early AV interactions over posterior cortex (40-90ms post-stimulus onset), over temporo-central cortex (~95ms post-stimulus onset), and over frontal cortex (140-165ms post stimulus onset). Interestingly, the short latency of the posterior interaction, thought to reflect primary visual areas, and temporo-central interactions, thought to reflect primary auditory cortex, suggested that these AV effects occurred at the same time as the earliest responses for a unimodal stimulus; that is, by the time the first evoked potential occurred for a modality-specific stimulus, the response already reflected a multimodal interaction. This important finding

fueled subsequent research aimed at uncovering the underlying mechanism for these early multimodal interactions.

Most subsequent multimodal studies have used AV speech stimuli, exploiting the precise relationship between the articulations of the mouth and the acoustic signal. Behaviorally, research has shown that audio-visual speech improves speech understanding (for a review, see Campbell, 2008); consequently most EEG studies have focused on identifying the timecourse of this multimodal performance enhancement (Bernstein et al., 2008; Besle et al., 2004; Klucharev et al., 2003; Lebib et al., 2003; Lebib et al., 2004; Musacchia et al., 2006; Saint-Amour et al., 2007; van Wassenhove et al., 2005). These studies have found consistent effects on the early auditory evoked potentials (P50, N1, P2) for the congruent AV trials (Besle 2004; Klucharev 2003; Lebib 2003; Lebib 2004; VanWassenhove 2005). Typical speculation has been that this modulation reflects the predictive relationship between the modalities, where the movement of the lips begins before the onset of the acoustic information allowing the visual to predict the auditory (although the visual signal was manipulated to not predict auditory in Besle et al, 2004; yet they still found an effect on N1/P2).

Stekelenburg and colleagues (2007) investigated whether this effect was specific to AV speech or if the early auditory effects would occur with environmental sounds where the visual predicted the timing of acoustic onset. Their study used only two events, clapping hands or tapping a pen on a glass. Like AV speech where lip movements precede the auditory signal, these events involved visual movement (hands moving towards one another or a pen moving towards a glass) that preceded the sound of the impact between the two hands or the two objects. The reduced amplitude for the auditory

N1 and P2 was independent of stimulus category, suggesting that a general predictive temporal relationship drives AV interactions across domains. Consequently, Stekelenburg ran two additional experiments to confirm this interpretation. In both experiments they manipulated the semantic congruence between the auditory and visual modalities as well as generating movies of events that lacked a temporally predictive relationship. First, the edited movies preserved the temporal congruency between the modalities, but eliminated the semantic congruency by having the sound of a hand clap occur when the movie showed the pen tapping the glass – equivalent to our TCSI condition. Early auditory effects were unaffected by the semantic incongruence, suggesting that temporal synchrony may underlie the auditory evoked potentials modulations. Second, they used movies of two new events where the visual stimulus did not predict the auditory, tearing paper or sawing wood. That is, the sound of the paper tearing was temporally coincident with the visual tearing motion of the hand. The effects on the auditory evoked potentials were eliminated. Thus, this study suggests that early multimodal interactions depend on the predictive temporal relationship between the modalities.

From these studies, it appears the AV interaction effects are specific to the early auditory evoked potentials; however, the experimental procedure may have contributed to the auditory specificity. These studies all displayed a static visual scene for several frames before the visual movement began. Thus, the onset of the visual stimulus, and the associated visual evoked potentials, occurred prior to the onset of visual movement in the video. The onset of the auditory stimulus, on the other hand, was temporally coincident with the onset of change in the acoustic information. As such, the experimental design of

these studies was optimized to investigate the effect of AV stimuli on the auditory evoked potentials.

A small number of studies have also investigated the role of multimodal effects in object recognition. In these studies, the onset of both auditory and visual modalities was temporally coincident, which is perhaps why they find AV effects in the early *visual* (not auditory) evoked potentials. Two studies paired animal pictures with animal vocalizations, either semantically congruent or incongruent, finding that a congruent pair reduced the visual N1 (Molholm et al., 2004; Yuval-Greenberg and Deouell, 2007). Paradigms that have used deforming circles combined with particular auditory tones to form multimodal objects have also found an AV effect on the visual N1 (Fort et al., 2002; Giard and Peronnet, 1999). In fact, the Giard & Peronnet (1999) study found that AV effects can occur in either the early visual evoked responses or in the auditory depending on the modality each participant relied on to complete the object identification task. Thus, the nature of early AV interactions may depend on both the experimental design as well as task strategy.

The interaction of semantics and integrative mechanisms is not limited to the early evoked potentials. The N400 marker of semantic congruence, originally identified for incongruent words in sentences (Kutas and Hillyard, 1980), is also found in priming paradigms with environmental events. Cummings and colleagues (2006) compared a condition where a picture was either semantically congruent or incongruent to a subsequent environmental sound or spoken word, finding that the N400 occurred a little bit earlier for the sounds than the events. Org and colleagues (2008) switched the order, first showing a word that was either semantically congruent or incongruent followed by

an environmental sound. The N400 was greater for the incongruent trials and could be enhanced when participants made an explicit judgment about congruency. The only study to look at the N400 for movies of everyday events showed silent films of actors performing every day tasks (ironing pants, cutting bread). Sitnikova and colleagues (2008) edited movies so that the action that began the movie was concluded with either a congruent or incongruent ending: the actor begins setting out a cutting board and a loaf of bread to cut, and the movie either ends with him cutting the bread (congruent) or ironing the bread (incongruent). The silent incongruent action elicited an increased amplitude for the N400 and also a shorter latency (just like Cummings et al, 2006). Together, these studies indicate that the N400 reflects semantic congruence of picture-sound pairs as well as semantic congruence of visual-only movies.

Using the same event stimuli as Experiment 2, Experiment 3 builds on this evidence for AV interactions by exploring the interplay of temporal and semantic congruence during the processing of multimodal movies of environmental events. The onset of auditory and visual information occurs simultaneously to avoid any bias for effects on the auditory evoked potentials, ensuring that all of the ERPs of interest based on the prior research (P50, N1, and P2 for auditory, P1 and N1 for visual, and N400 for semantics) can be explored for temporal and semantic influences. In addition, a spatio-temporal analysis is performed to directly investigate the timecourse of processing for each cue, examining whether temporal synchrony, that is, low-level information, has an earlier influence on processing than semantic congruence, that is, high-level knowledge.

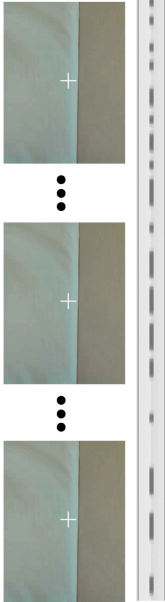
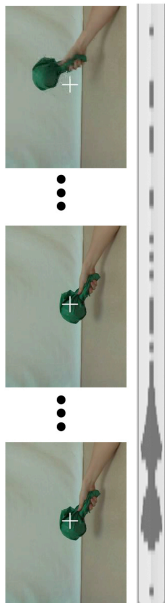
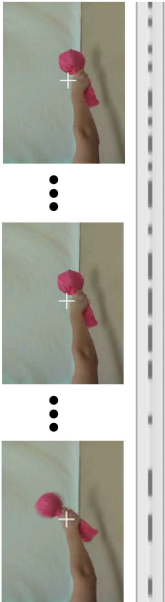
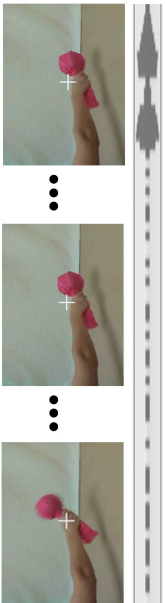
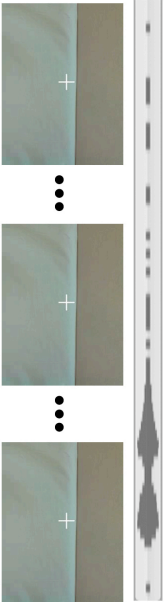
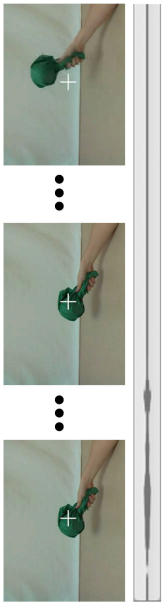
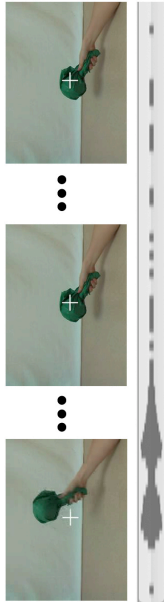
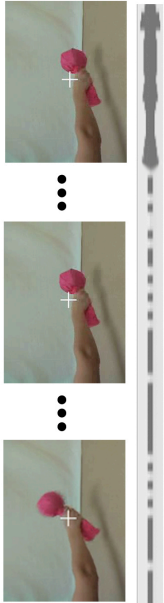
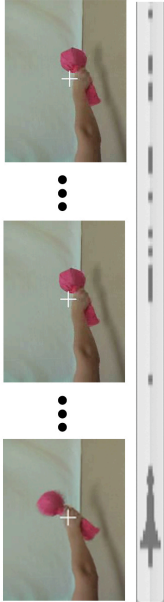
4.2 Methods & Materials

Participants. Thirty two right-handed, University of Colorado (Boulder, CO) students (14 female/18 male; mean age: 20; range: 18-26) participated in the study, but the dataset from one subject (female) was incomplete due to eye discomfort that terminated the session. Subjects received course credit or were paid for their participation, and all provided written, informed consent consistent with and approved by University of Colorado IRB. All had normal or corrected-normal vision and hearing.

Stimuli. The same thirty-six, two second AV movies (18 unique real-world, physical events with 2-4 discrete impacts) from Experiment 2 were again used for this study. The only changes were to minimize eye movements by rescaling the size of the movies to 320x240 and adding a central, white fixation cross. Examples can be seen in Figure 11.

Figure 11 *Nine Conditions and Target Examples*

Examples of the nine conditions used in Experiment 3 are illustrated with the shaking a maraca event. In the control condition, an empty visual scene was shown with ambient background noise. In the V_{only} condition, the event movie was shown with the ambient background noise. In the A_{only} condition, the empty visual scene was shown with the auditory track of the event. In the TCSC condition, the original movie and its associated waveform were shown. In the TCSI condition, the event movie was shown with a waveform edited with the impact from another event to be temporally congruent. In the TISC conditions, the impacts used in the auditory waveform were from the same event, but in the “av” condition, the beginning of the movie was shifted so that the auditory impact occurred before the visual impact. Conversely, in the “va” condition, the original movie was shown by the waveform was delayed by 500ms so that the visual impact occurred before the auditory impact. For the TISI conditions, the same process was used as the TISC conditions, except that the auditory impacts were extracted from other events and edited to maintain the same temporal offset as the TISC conditions. Finally, the right, bottom row shows an example of the purple visual target as well as the waveform for the 500ms auditory target.

Control		TISCav	
Vonly		TISCva	
Aonly		TISlav	
TCSC		TISiva	
TCSI		Targets	

Experimental Conditions. Based on the more fine-grained temporal resolution of EEG, sampling every 4 milliseconds compared to 3 seconds in the fMRI study, three modifications were made to the six conditions of Experiment 2 (A_{only} , V_{only} , TCSC, TISC, TCSI, & TISI): (1) all conditions now included a visual and auditory stimulus (2) a new control condition was added; and (3) the ordering the unimodal information in the temporally asynchronous stimuli was explicitly coded for each trial in order to constrain what subset of trials were used in the analyses.

The slight modification in condition parameters so that each trial contained an auditory and visual component ensured that the early evoked potentials for both modalities occurred on every trial, thereby equating the low-level stimulus responses on each trial. This allows us to explore differential effects that are related to the dynamic event action in the movies without pervasive signal differences due to the absence of these early component processes in one condition but not the other. As noted in the Introduction, this intentional design difference from most of the research on AV speech maximizes our ability to explore early multimodal interactions on both auditory and visual responses, rather than focusing on auditory only. A depiction of each condition is shown in Figure 11. The modifications from Experiment 2 occurred for the unimodal trials where an image of an empty scene is used on the A_{only} trials and the ambient background noise is heard during V_{only} trials.

The other condition change was the addition of a control trial that presented the static visual scene with the ambient background noise. These trials eliminate a possible confound for multimodal interactions due to a slow wave evoked potential that precedes

every trial (Teder-Sälejärvi et al., 2002); however, these trials will not be used in any of the analyses discussed in this writeup.

Finally, although the same physical stimuli were used, the two temporally asynchronous conditions, TISC and TISI, were explicitly coded for the temporal ordering of the unimodal information. In other words, TISCav/TISlav refers to the movies where the auditory impact is heard at 67 milliseconds post-stimulus onset and the visual impact is not seen until 567 milliseconds post-stimulus onset (a phase offset of 500 milliseconds), while TISCva/TISiva indicates that the visual impacts occurs at 67 milliseconds post-stimulus onset and auditory impact at 567 milliseconds post-stimulus onset. This recoding of the temporally incongruent conditions was essential when comparing waveforms to equate the low-level information that was available as the event action in the movie unfolded. Inadvertently, the two orderings of the unimodal information (av and va) modulated the time at which subjects knew it was a multimodal trial. As shown in Figure 11, the V_{only} and two temporally incongruent conditions with the visual-auditory ordering, TISCva & TISiva, were indistinguishable for the first 500ms of the trial due to the addition of the ambient background noise played with the V_{only} movie. Conversely, the A_{only} and two temporally incongruent conditions with the auditory-visual ordering, TISCav & TISlav, were differentiated in the first frame since the multimodal trial had a video with an actress and object while the A_{only} trial showed an empty scene. Due to this difference, the analysis for the temporally incongruent trials is restricted to the auditory-visual ordering, TISCav.

Each subject saw all of the 36 events in each condition 4 times, completing a total of 1024 trials (15% of which were control trials). The target trials were infrequent (4% of

trials), but they were equally likely to occur in all conditions, with half of the targets occurring aurally and half visually (see Figure 11 for examples). Target trials were not analyzed, providing a total of 980 trials per subject before artifact rejection.

Experimental Task. As in Experiment 2 and several prior studies (Senkowski et al., 2007; Stekelenburg and Vroomen, 2007; Teder-Sälejärvi et al., 2002), the subjects performed a target detection task that was independent of the congruency manipulations. Once again, the target was either visual or auditory to ensure that the participant was attending to both modalities, and examples of the targets are shown in Figure 11. Unlike the fMRI study, subjects did not respond on every trial in order to eliminate decision- and response-related processing and to maximize our power to explore the timecourse of multimodal processing; instead, subjects responded as soon as they detected a target, and these target trials were excluded from the analysis. Subjects performed very well, with an overall accuracy of $99\% \pm .003\%$.

Experimental Procedure. Before the recording session, each subject performed a behavioral task with the experimental stimuli, ensuring that all events were equally familiar and identifiable during the recording session.

Scalp EEG voltage data were collected with a 128-channel HydroCel Geodesic Sensor Net connected to an AC-couple 128-channel, high-input impedance amplifier (200 M Ω , Net Amps, Electrical Geodesics Inc., Eugene, OR). Amplified analog voltages (.1-100 Hz bandpass) were digitized at 250 Hz. Individual sensors were adjusted until impedances were less than 50 k Ω . The EEG was digitally low-pass filtered at 40 Hz.

Trials were excluded for four reasons: (1) if the trial contained a target; (2) if the subject made an incorrect response (identifying a non-target stimulus as a target stimulus); (3) if the trial contained an eye movement (electro-oculogram channel differences greater than $70 \mu\text{V}$); and (4) if more than 20% of the channels were bad (average amplitude over $100 \mu\text{V}$ or voltage fluctuations greater than $50 \mu\text{V}$ between adjacent samples). Individual bad channels were replaced on a trial-by-trial basis with a spherical spline algorithm.

Individual Recorded voltages were initially referenced to a vertex channel (Cz), but an average-reference transformation was used to minimize the effects of reference site activity and accurately estimate the scalp topography of the measured electrical fields.

The average reference was corrected for the polar average-reference effect (Junghöfer et al., 1999).

Event-related potentials (ERP) were time locked to one of two timepoints: the onset of the stimulus or the onset of the second impact. The onset of the stimulus occurred 67ms before the peak of the auditory impact sound or its equivalent frame in the visual stimulus. Thus, a time lock to the stimulus onset allows exploration of the early sensory evoked potentials as well as the first impact in the event. Note that this time is equivalent for all movies. A time lock to the peak of the auditory impact sound for the second impact was used in the N400 analysis to explore amplitude changes as more temporal and semantic congruence information accrued as the action in the events unfolded. Note that the time of the second impact is variable between events to ensure that all events didn't share a predictable rhythm. The average timepoint for the second impact across all 36 event movies was 1164ms (+/- 231ms).

After artifact rejection, the mean number of trials per condition per subject was 110/136 $\pm$ 5. ERPs were baseline-corrected to the 200ms prior to stimulus onset.

Data Analysis. The first analysis focused on the early sensory evoked potentials identified by prior multimodal research (see Introduction): the visual P1 and N1 as well as the auditory P50, N1, and P2. To identify the time window and location for each component, the location of the peak amplitude was identified, a cluster drawn around the peak electrode, the latency for the component across all subjects was computed, and the mean latency across subjects $\pm$ two standard deviations was used for the time window for the component. The components were scored in the follow time windows: visual P1 = 77-156ms, visual N1 = 135-182ms, auditory P50 = 46-94ms, auditory N1 = 117-183ms, and the auditory P2 = 194-251ms. These clusters and time windows were used to test for sensory effects of temporal and semantic incongruence by comparing TCSC, TCSI, and TISCav.

The second analysis explored the N400 component, identified as a marker of semantic congruence originally in language but extended to incongruent pictures and events. The N400 was scored from 300-500ms based on prior multimodal research (Cummings et al., 2008). Only the TCSC and TCSI conditions were compared in this analysis because the temporally offset visual impact in the TISCav does not occur until 567ms (after the N400 time window) and thus no semantic incongruence has occurred yet in the temporally incongruent condition.

The final analysis explored the time course of processing for each integration cue. To investigate temporal processing, point-by-point two-tailed *t*-tests on the difference

wave between the TCSC and TISCav was computed for each electrode from stimulus onset through 1500ms. The difference was considered significant when 12 or more consecutive time points (48 ms) were significantly different from zero ($p < .05$) (Guthrie and Buchwald, 1991). This analysis detected the time windows and scalp distribution for effects of temporal synchrony. The same process was used to test semantic processing using the difference wave between TCSC and TCSI.

4.3 Results

The data analysis involved three major stages. First, we investigated the early visual and early auditory evoked potentials (auditory P50, N1, and P2 & visual P1 and N1) for temporal and semantic effects using TCSC, TCSI, and TISCav. Second, semantic effects were examined on the N400. Third, we conducted a spatio-temporal analysis to examine the overall timecourse of processing for both temporal synchrony and semantic congruence.

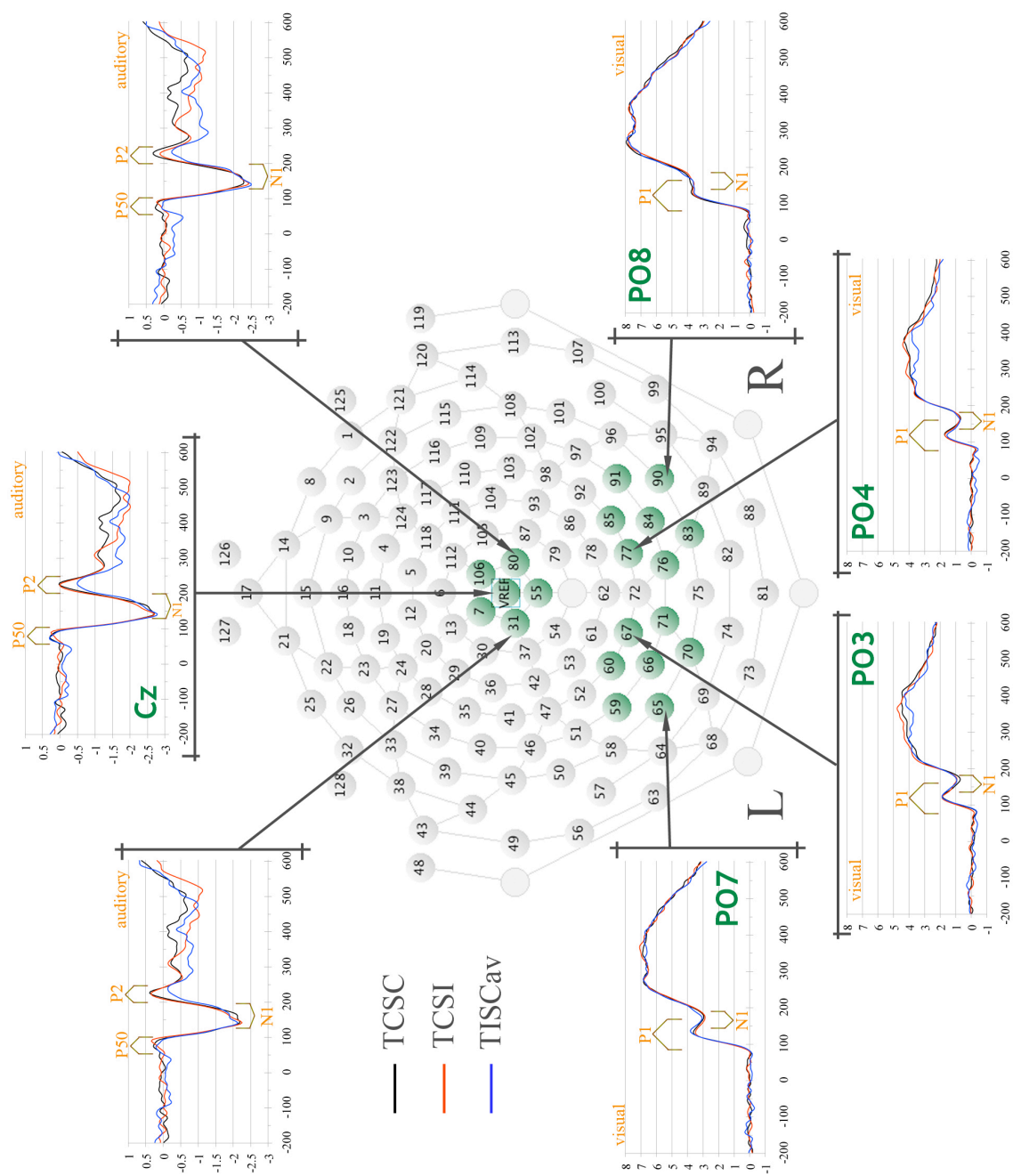
Temporal & Semantic Effects on early sensory ERPs

This analysis explores whether any differences can be seen on any of the early sensory evoked potentials when a multimodal stimulus occurs but either semantics (TCSI) or time (TISCav) is incongruent between the auditory and visual modalities. To explore effects on the two visual evoked potentials, P1 and N1, a condition (TCSC, TCSI, TISCav) by hemisphere (Left & Right) ANOVA was conducted in the two posterior clusters (see Methods & Figure 12). No significant effects were found for the P1

($p > 0.05$, biggest $F(2,60) = 1.37$) or the N1 ($p > 0.05$, biggest $F(2,60) = 2.09$). The auditory evoked potentials were also investigated in one cluster centered around Cz with a one-way ANOVA of condition (TCSC, TCSI, TISCav). Again, no significant effects were found on the P50 ($p > 0.05$, $F(2,60) = 0.66$), the N1 ($p > 0.05$, $F(2,60) = 0.72$), or the P2 ($p > 0.05$, $F(2,60) = 0.12$). A representative plot of the ERPs for each condition in each cluster is shown in Figure 12.

Figure 12 *Early Sensory Evoked Potentials*

The evoked related potentials (ERPs) timelocked to the onset of the event movie are shown for seven electrodes, with the auditory evoked potentials depicted at the top and visual evoked potentials at the bottom. In each trace, the black line is the TCSC condition, the red line is the TCSI condition (semantics), and the blue line is the TISCav condition (time). Electrodes 31, VREF (equivalent to the 10-10 electrode Cz), and 80 are shown from the auditory cluster (centered around VREF where the peak was found for the early auditory components) with orange markers showing the time windows for three auditory components, P50, N1, P2 (see Methods). Electrode 65 (equiv. PO7) and electrode 67 (equiv. PO3) are shown from the left hemisphere visual cluster (centered around electrode 66 where the peak was found for the early visual components) with orange markers showing the time window identified for the two visual components, P1 and N1 (see Methods). Similarly, bilateral electrodes 77 (equiv. PO4) and electrode 90 (equiv. PO8) are shown from the right hemisphere visual cluster with the same orange markings. The montage of the 128 electrode net identifies the other electrodes used in each cluster (auditory, left visual, right visual) by coloring the electrodes green.



Semantic Effects on the N400

This analysis investigates whether semantic incongruence between auditory and visual movies of environmental events elicits the N400 component, a marker of semantic congruence. Prior research has shown that environmental sounds elicit an N400 that is more anterior over frontal regions than the centrally-located N400 originally identified for words (Orgs et al., 2006; Sitnikova et al., 2008; Van Petten and Rheinfelder, 1995; Van Petten and Luka, 2006). There is also evidence that the amplitude of the N400 during sentence processing is modulated as evidence about semantic constraints accrues for subsequent words (Van Petten and Luka, 2006). Consequently, based on these prior findings, this analysis explored three clusters of electrodes (Fz, Cz, and Pz) to investigate the anterior-posterior localization, and it also investigated the effect of impact number to examine if the N400 amplitude for the second impact would be modulated as the amount of semantic congruence evidence accrued during the movie. Finally, unlike the analysis on the early sensory components, this analysis only included the TCSC and the TCSI conditions because the offset of the semantically incongruent visual stimulus in the TISCav condition is not seen until 567ms, a timepoint outside of the 300-500ms window identified for the N400.

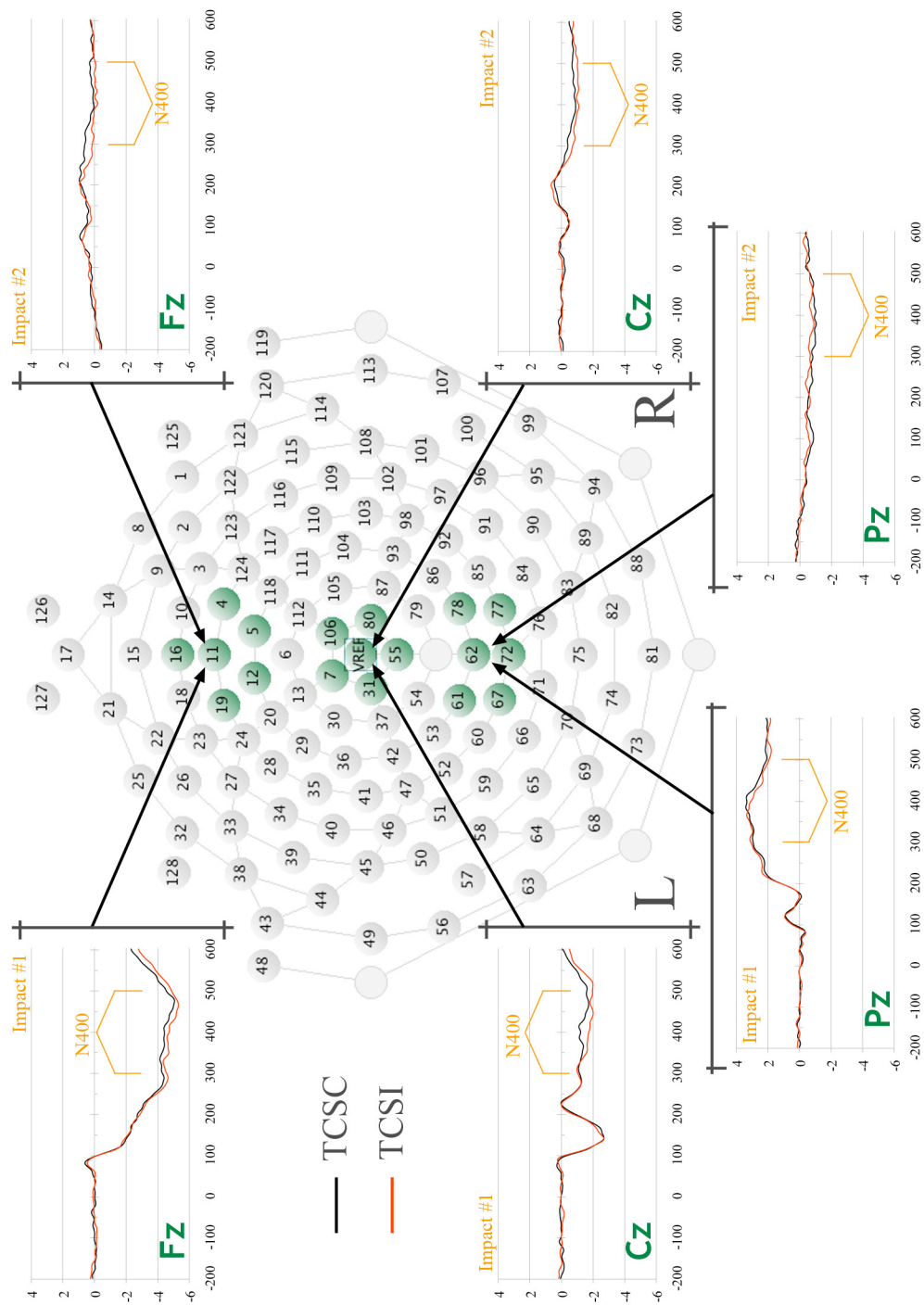
A 3-way ANOVA of condition (TCSC, TCSI), cluster (Fz, Cz, Pz), and impact number (first, second) showed a main effect of condition ($F(1,30)=4.39$, $p<0.045$), a main effect of cluster ($F(2, 60)=74.05$, $p<10^{-16}$), and an interaction between cluster and impact number ($F(2,60)=132.73$, $p<10^{-16}$). None of the other effects were significant (smallest p : $F(2,60)=0.73$, $p<0.48$ for the 3-way interaction).

To explore these effects further, subsequent 2-way ANOVAs of condition by impact number were conducted separately in each of the three clusters. In the most anterior cluster, Fz, a main effect of condition was found ($F(1,30)=5.23$, $p<0.029$), but no other effects were significant (impact number: $F(1,30)=0.94$, $p>0.05$) nor was the interaction: $F(1,30)=0.139$, $p>0.05$). In the central cluster, Cz, the main effect of impact number was significant ($F(1,30)=148.69$, $p<3.73^{-13}$), most likely driven by the early auditory potentials that are localized on Cz at the start of the movie and almost coincident with the first impact. The main effect of condition in the Cz cluster approached significance ($F(1,30)=3.84$, $p<0.059$), while the interaction did not ($F(1,30)=1.29$, $p>0.05$). In the most posterior cluster, Pz, only the main effect of impact number was significant ($F(1,30)=84.55$, $p<3.11^{-10}$; condition: $F(1,30)=0.026$, $p>0.05$ and interaction: $F(1,30)=0.19$, $p<0.66$). Again, the strong impact effect is likely driven by the early visual potentials localized just inferior and the early auditory potentials localized just anterior.

The ERPs for each condition at each location are shown in Figure 13.

Figure 13 *The N400 response for Impacts 1 & 2*

The evoked related potentials (ERPs) are shown for two analyses: on the left, the graphs show the ERP timelocked to the onset of the event movie (coincident with the first impact in all events), while the graphs on the right show the ERP timelocked to the onset of the second impact (variable time from stimulus onset across the events). Responses amplitudes were tested in three clusters, shown by the green electrodes in the montage, and the ERPs are shown for one representative electrode from each cluster: electrode 11 (equivalent to the 10-10 electrode Fz), electrode VREF (equiv. Cz), and electrode 62 (equiv. Pz). Orange markers indicate the time window used for the N400 component.



Spatio-temporal Analysis

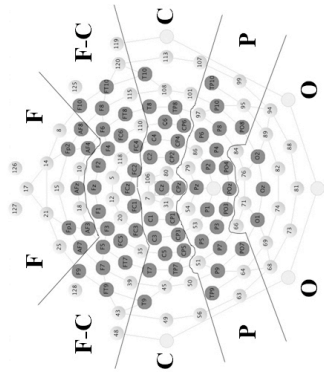
This final analysis emphasizes the overall timecourse and spatial localization for temporal and semantic effects during multimodal processing. Point-wise t-tests were computed at every electrode, comparing the amplitude of two of the multimodal conditions. The comparison between the TCSC and the TCSI condition provides a spatio-temporal map of semantic congruency processing, while the TCSC and TISCav provides a map of temporal synchrony processing.

As shown in Figure 14, the earliest effect of semantic congruence occurs over central scalp regions between 400 and 550ms. The effects then shift anteriorly, revealing semantic effects over frontal regions between 800-900 ms as well as between 1100-1400 ms.

Also in Figure 14, the earliest effect of temporal synchrony is seen in central as well as posterior electrodes between 300-400 ms, occurring just before the centralized semantic congruency effects. Effects of temporal synchrony are also seen between 1000-1150ms, approximately the average time point for the second impact in the events (see Methods), suggesting that this activity may reflect the real-time processing of the event dynamics. In both instances, the temporal synchrony effects immediately precede semantic congruency effects. This earlier effect for time may reflect the fact that temporal synchrony information is carried in the perceptual signal (a low-level integration cue), while semantic information is dependent on accessing higher-level knowledge.

Figure 14 *Spatio-temporal Analysis of Semantic and Temporal Effects*

The montage at the top left shows what electrodes are included in the topographic groupings (F, F-C, C, P, O) labeled on the y axis of the spatio-temporal plot (here, dark gray indicates the classic 10-10 layout). Within each topographic grouping, the electrodes are ordered from left through center to right. The spatio-temporal plot shows point-by-point two-tailed t-tests for difference waves that were greater than zero for at least 12 consecutive time points ($p < 0.05$). As illustrated above the plot, the blue lines indicate a significant difference between the TCSC and TISCav conditions (i.e., effects of temporal synchrony), while the red lines indicate a significant difference between the TCSC and TCSI conditions (i.e., effects of semantic congruence). Finally, topo maps of the difference wave are shown for several time points, as indicated by the gray line extending from the timeline to each column of maps. The topo maps for the time difference wave are in the top row (on top of the blue line), while the topo maps for the semantics difference wave are in the bottom row (on top of the red line). In the bottom left of the figure, the legend for the topo maps is shown.



10-10 in dark gray, overlaid on 128-channel HydroCel GSN; all electrodes used in analysis

Spatio-Temporal Analysis

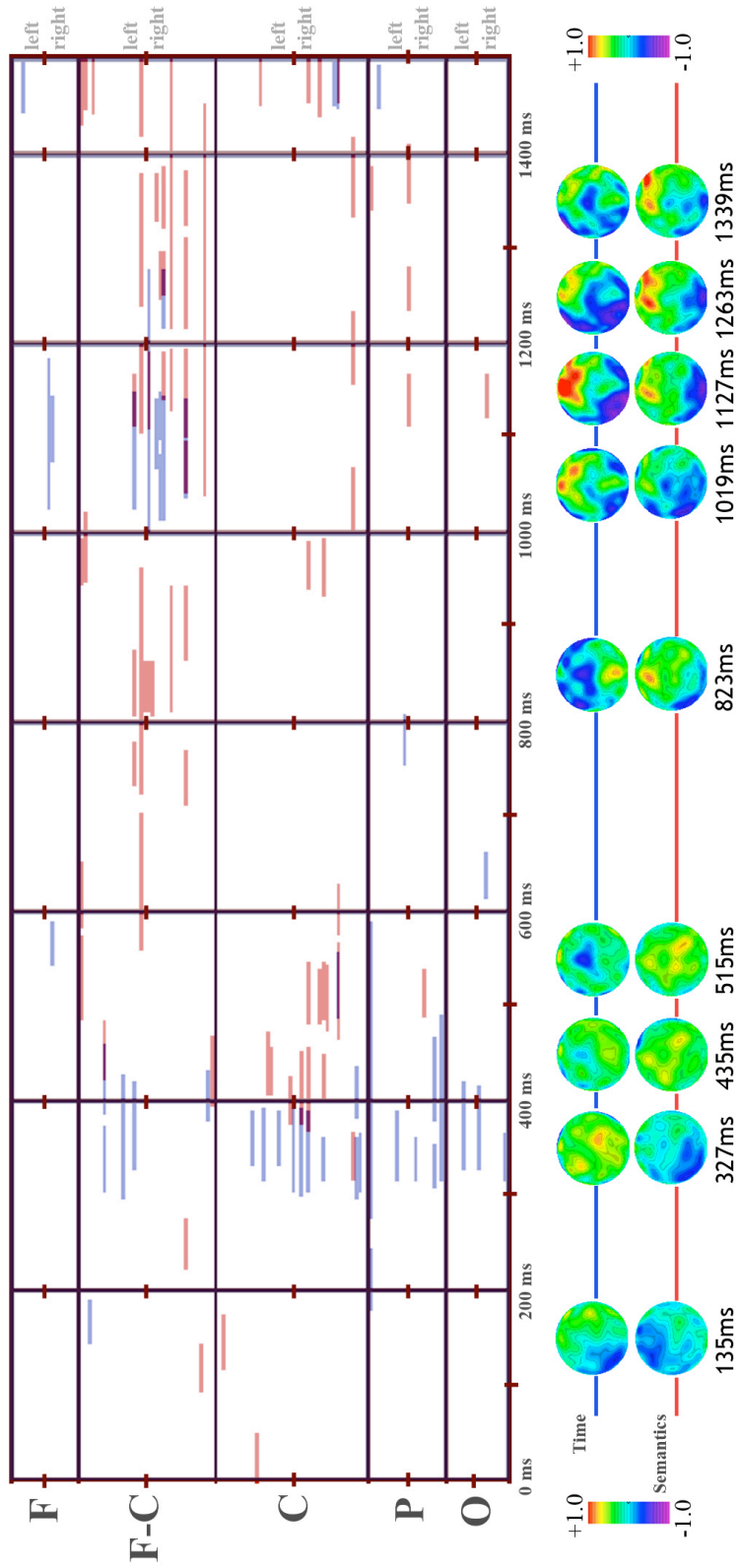
12 (or more) consecutive timepoints, $p < 0.5$

Time Difference Wave

Time	Semantics	vs.	Time	Semantics
✓	✓		✗	✓

Semantics Difference Wave

Time	Semantics	vs.	Time	Semantics
✓	✓		✓	✗



4.4 Discussion

This study investigated the timecourse and spatial localization of semantic and temporal congruency effects during the processing of dynamic movies of environmental events. The three analyses explored effects of the two integration cues on the early sensory evoked potentials, on the modulations of the N400 component known to reflect semantic constraints, and on the spatio-temporal properties of multimodal processing. The earliest temporal synchrony effects were seen between 300-400ms with a posterior distribution over the scalp. The earliest effects of semantic congruency immediately followed between 400-550ms, within the timeframe of the N400, in more fronto-central regions. Both cues showed effects in frontal regions around the average time point of the second impact, with temporal effects once again preceding the semantic effects. In short, the results indicate different temporal latencies for the processing of two cues to integration. In particular, low-level temporal information influences multimodal processing before the higher-level semantic information.

For the remainder of this discussion, we focus on the results from each of the three analyses, speculating about what the presence or absence of effects for the two integration cues may indicate about that underlying processing mechanisms for multimodal integration.

Absence of Effects on Early Sensory ERPs

Most of the prior literature has explored the multimodal interactions found on the early evoked potentials by comparing the simultaneously presented auditory-visual stimuli with the summed response of the auditory and visual stimuli presented alone, i.e., AV versus A+V (Fort et al., 2002; Giard and Peronnet, 1999; Molholm et al., 2002; van Wassenhove et al., 2007). As discussed in the Introduction, multimodal interactions have been identified most frequently on the early auditory components (P50, N1, P2), though modifications of the commonly used experimental design reveal effects on the early visual components (P1, N1) as well.

However, the AV to A+V comparison used in earlier studies highlights the advantage of redundant multimodal information, while here, we focused on the differential effects of semantic and temporal congruency on multimodal processing. Consequently, it is not expected that effects would be seen on the early components for all three conditions (TCSC, TCSI, and TISCav), in that the two temporally congruent conditions contain redundant multimodal information. The only study, to the best of our knowledge, that is comparable to our design investigated a TCSI condition with similar movies of impact events, finding the same multimodal interactions for the TCSI condition as was found for the TCSC condition (as compared with unimodal stimuli) (Stekelenburg and Vroomen, 2007). This finding suggests that the semantic relationship between modalities is *not* able to influence the early sensory evoked potentials. Our results support this finding as well as introduce a novel experimental condition that is temporally asynchronous.

In this context, the fact that we did not identify a temporal synchrony effect on early evoked potentials was a bit surprising. A typical explanation about the underlying mechanism of early multimodal interactions argues that they depend on the temporal synchrony (and spatial correspondence) between modalities, an explanation inspired by the seminal work in the superior colliculus of cats (Meredith and Stein, 1983). This explanation predicts the TISCav condition would not show multimodal interactions on the early sensory component because temporal synchrony is not present in the multimodal stimuli. Therefore, if the response to TCSC and TCSI shows the multimodal enhancement due to temporal and spatial synchrony, a comparison among all three conditions (TCSC, TCSI, and TISCav) would be expected to show a difference since TISCav lacks the multimodal enhancement. However, our study finds no difference among these three multimodal conditions. One explanation for the absence of an effect could be that the underlying assumption about the mechanism is incorrect; that is, the early sensory effects do not reflect temporal and spatial congruence between auditory and visual modalities, but instead, the early effect reflects the complexity of the stimulus. That is, two visual stimuli (intramodal stimuli) would have the same modulation on the early sensory potentials as a visual-auditory stimulus.

This idea is drawn from the fMRI study by Hocking & Price (2008) that argues that the posterior superior temporal sulcus region thought to be an important gateway for multimodal integration responds just as well to two stimuli in the same modality as it does for stimuli in two different modalities. An equivalent explanation for the early ERP interactions has not been explicitly tested. The only preliminary evidence that the same early effect would not occur with intramodal stimuli is the third experiment in the study

by Stekelenburg and colleagues (2007). This experiment showed that multimodal interactions were not found when the visual information did not predict the timing of the auditory impact. Thus, two modalities were present yet multimodal interactions were not seen. However, as also noted in the Introduction, this experimental design presents a static visual stimulus for several hundred milliseconds before visual movement begins and an auditory stimulus is heard. Thus, it is not clear that these results would hold when both modalities share the same onset, and the sensory processing is initiated for both modalities simultaneously.

Independent of experimental design, another possibility exists that may explain why we found no difference among the three multimodal conditions on early evoked potentials. The integration process must account for the inherent temporal delay between auditory and visual information due to different sensory processing times from ear/eye to the cortical areas. Therefore, the full duration of this delay time window has to pass before the system can detect a temporal asynchrony. Behavioral research has shown that subjects are unable to detect onset asynchronies of up to 80ms when the auditory stimulus precedes the visual information (Bushara et al., 2001; Lewkowicz, 1999). Thus, this inherent tolerance of temporal asynchrony may necessitate more processing time than has occurred by the early evoked potentials, suggesting that effects of temporal synchrony would not be seen until after this inherent temporal delay has passed (a duration of time that is subject for future investigation). We will return to this idea in Chapter 5; however, admittedly, null effects are hard to interpret.

Semantic Effect on the N400

Although the N400 was originally identified for incongruent words in sentences (Kutas and Hillyard, 1980), it has been extended to picture-word priming paradigms (Cummings et al., 2006; Orgs et al., 2008) as well as silent movies that ended with an incongruent final scene (Sitnikova et al., 2008). We found that this marker also reflects semantic incongruence during the dynamic unfolding of event actions. In line with this prior research, the N400 was localized in the anterior Fz cluster for the events, rather than the more centralized localization of the N400 with words (Orgs et al., 2006; Sitnikova et al., 2008; Van Petten and Rheinfelder, 1995; Van Petten and Luka, 2006). Several of these authors speculate that this more anterior shift in the scalp reflects slightly different neural generators for processing the semantics of event actions compared to processing the semantics of language comprehension, but the poor spatial resolution of EEG makes this interpretation hard to examine directly in any of these studies.

This analysis also explored whether the amplitude of the N400 would increase by the time of the second impact as the semantically incongruent information continues to accrue during the movie. Research on sentence comprehension has found that the amplitude of the N400 can be modulated by the semantic context (Van Petten and Luka, 2006), but no difference in N400 amplitude was revealed between the first and second impact. However, it is not clear that two impacts would be sufficient to result in N400 modulation. Subjects hear many words during sentence comprehension that gradually build up the semantic context, whereas we are testing here after the second impact, not after many impacts, so there may be insufficient evidence to develop a rich semantic context. This is certainly an interesting question for future research.

Timecourse of Multimodal Processing

The spatio-temporal analysis provides an overall picture of when and where the effects of temporal and semantic congruency occur. The earliest temporal effects are distributed over posterior cortex before moving to more anterior locations by the second impact, while semantic effects have a fronto-central scalp distribution. Interestingly, when the temporal effects and semantic effects are localized in similar electrode clusters, the effects of temporal asynchrony *always* directly precede the effects of semantic congruency. These findings are in accordance with the predicted timecourse for a low level cue, such as temporal synchrony, that is present in the perceptual signal, and, thus, can have an earlier influence on multimodal processing. The waveforms for the TCSC and the TCSI conditions are initially indistinguishable perhaps due to the temporal synchrony between auditory and visual modalities, a relationship that suggests a common origin for the modality-specific information. However, as multimodal processing continues, higher-level semantic information about the stimulus can begin to influence processing, perhaps overriding the temporally synchronous information. Therefore the system will no longer bind the modality-specific information together. Thus, the semantic congruence can influence and refine integrative processes that may be incorrectly initiated due to the early temporal synchrony across modalities.

4.5 Conclusion

The goal of Experiment 3 was to explore the processing timecourse of temporal and semantic congruency for dynamic, multimodal events. More specifically, we were interested in whether a low-level integration cue, temporal synchrony, would have an earlier influence on multimodal processing than a higher-level cue, semantics, that relies on experiential knowledge. Confirming this conjecture, Experiment 3 revealed that temporal effects do in fact precede semantic effects. This indicates that the two cues have different temporal latencies in addition to recruiting separable networks of brain regions identified in the two fMRI experiments. More generally, Experiment 3 reveals that multimodal processing is an ongoing process for several hundred milliseconds following stimulus onset. During this time window, different integration cues influence the processing on different timescales, revealing that multimodal integration is a dynamic process, influenced by both low-level and higher-level information, perhaps in interactive stages.

5. Chapter 5. Discussion

We experience most events in the world in multiple modalities. Prior research on multimodal integration, both behavioral (De Gelder and Bertelson, 2003) and neural (Driver and Noesselt, 2008), identified the critical role of spatial, temporal, and semantic congruency between modalities as factors guiding the integration between modalities. Converging evidence from both animals and humans, as well as data from various methods, identified a consistent network of brain regions in posterior, temporal, and frontal cortices recruited during multimodal processing (Kayser and Logothetis, 2007).

The series of studies in this dissertation were primarily focused on three questions: (1) how do the integration cues interact with one another? (2) what functional roles do the regions recruited in the multimodal network serve in multimodal processing, and are the convergence zones the primary sites of neural integration? (3) what is the timecourse of the interplay between the integration cues during multimodal processing?

These questions guide this discussion, as we review the main results for each of the three experiments, exploring how these data inform our understanding about the neural integration of multimodal events.

5.1 Experiment 1

The first fMRI study verified that our novel set of environmental event stimuli recruited the common regions consistently identified in prior multimodal research, with the addition of a posterior cingulate region that may be specific to our particular stimuli (Lewis et al., 2004). Experiment 1 had two major findings: (1) regions identified by a

functional definition of integration (hypothesized to be convergence zones) were not sensitive to either integration cue; and (2) each of the integration cues recruited separable regions, with a preference for semantics in the left frontal hemisphere and a preference for time in the right frontal hemisphere.

The first result addresses the question of the underlying neural architecture discussed in chapter 1, suggesting that convergence zones identified by a functional contrast may not reveal regions responsible for integrative processes. Functional contrasts have served as a fundamental principle of multimodal integration research for decades. The race model confirms multimodal interactions in behavioral studies (Miller, 1982), while the principles of super- and sub-additivity in neurophysiological research confirm multimodal interactions in single cell research (Stein and Meredith, 1993). The initial fMRI studies were heavily inspired by the neurophysiological research (Calvert, 2001), leading to a debate about the most appropriate functional definition for identifying integration regions from the BOLD signal (Laurienti et al., 2005). Subsequently, the alternatives were methodically compared (Beauchamp, 2005), identifying the functional contrast, $0 < A < AV > V > 0$, as the most appropriate (which we used in Experiment 1). The pervasive assumption was that a greater response for multimodal stimuli indicated a primary role in multimodal integration, as reflected in Account C of the proposed neural architectures (Figure 1) that predict convergence zones as feedback regions to primary sensory regions. However, our study is the first to empirically investigate how manipulating the congruency of the integration cues modulates the neural responses in functionally-defined integration regions (convergence zones). If these functionally-defined regions are responsible for integration between modalities, the response of the

region should differentiate among congruent temporal and semantic relations (as indicators of a coherent, multimodal percept) and incongruent temporal and semantic relations (as indicators that modality-specific information should not be integrated). However, no differentiation for congruency, for either integration cue, was seen in any of the four functionally-defined integration regions. This result converges with a study by Hocking and Price (2008) that showed that the neural response to two stimuli in the same modality was equal to the response to two stimuli in different modalities. Consequently, there may be relevant convergence zones, as proposed in Account C (Figure 1), but the prevalent view that functionally-defined regions in fMRI successfully identifies these convergence zones seems to be inaccurate. An alternative approach is needed to investigate the underlying neural architecture.

The second result reveals that the two integration cues do not interact with one another at all stages of processing since separable regions are identified for each cue. Furthermore, the overlap analyses suggested functional roles for the bilateral frontal activation. The right hemisphere may be specialized for processing the temporal relations between auditory and visual stimuli, while the left hemisphere may be specialized for semantic relationships between the modalities. Similarly in the posterior cortex, the overlap analyses suggested particular subregions of the posterior cingulate and the parietal cortex may also be functionally specialized for the processing of either time or semantics.

Experiment 1 suggests that explicitly manipulating integration cues may provide better localization of regions engaged in multimodal processing rather than relying on functional contrasts. Although Experiment 1 found that convergence zones, i.e.,

functionally defined integration regions, were not sensitive to integration cues, this does not rule out Account C (Figure 1) for the underlying neural architecture; however, these results do suggest that a method to identify convergence zones is currently unknown for fMRI studies. In addition, the results do not really fit Account A or B for the neural architecture either, as those proposed models are predominantly focused on the early sensory areas, and primary auditory and visual cortices were not identified. This could reflect the large-scale resolution of fMRI, or it could suggest that the underlying architecture is more widespread throughout cortex.

5.2 Experiment 2

The factorial design of the second fMRI study enabled a more thorough investigation of each region's functional role, including the identification of regions revealing interactions between the two cues. Importantly, the results of the second study converged with the major findings of the first study: each cue recruited separable brain regions, and the preference for time in the right hemisphere and semantics in the left hemisphere extended to posterior regions in addition to replicating Experiment 1's results for this hemispheric preference in the frontal regions. Thus, these hemispheric preferences for particular integration cues may be an organizing principle for the multimodal network, leading to a provocative conjecture about the underlying neural architecture. Currently, the three accounts proposed in Figure 1 all focus on the individual sensory modalities, and the accounts vary on how cross-talk occurs between the primary sensory cortices. In other words, these architectures propose that the multimodal network describes how the neural circuitry specialized for each sensory

modality interacts with one another. Perhaps the underlying circuitry is not for modalities, but it is for integration cues instead. That is, we replace the specialized substrates for modalities with specialized substrates for integration cues, such as spatial correspondence, temporal synchrony, and semantic congruence. These specialized substrates may be identified as those regions showing a main effect of time, a main effect of semantics, etc. This alternative account would replace the “audiovisual” and “audiotactile” regions in Account B with “temporal-semantic”, “temporal-spatial”, “spatial-semantic” regions. In short, the underlying neural circuitry for the multimodal network is specialized for processing shared relationships across modalities, not individual sensory modalities, which is reminiscent of “amodal properties” discussed in the Introduction. Returning to the Experiment 2 data, we discuss its possible relation to research on amodal stimulus properties.

One of the most striking results in Experiment 2 (Table 2) was the stark contrast in the number of regions showing a main effect of semantics (2) with the number of regions showing a main effect of time (17). The participants performed the same target detection judgment for all conditions, and accuracy was in the high nineties, so there is no evidence that this reflects difficulty with the experimental task for one integration cue but not the other. One possible explanation relates to the distinction of time as a low-level cue and semantics as a higher-level cue. The temporal structure of the event is carried in the perceptual signal itself, and thus, the temporal pattern can be identified independent of a particular modality, an amodal stimulus property. Adults reliably use temporal relationships in multimodal behavioral tasks (Welch, 1999), and, more interestingly, infants show sensitivity to the temporal relation in behavioral tasks as well (Spelke,

1976). Several researchers have argued that amodal properties, like time, are the developmental basis for how infants learn what modalities should be bound in to coherent percepts (Bahrick and Lickliter, 2002; Lewkowicz, 1999; Spelke, 1976). If temporal relations play a fundamental role in the neural integration of multimodal events, it may not be surprising that time permeates most of the regions in the multimodal network reflecting its strong influence on the process. Semantics, on the other hand, is a higher-level cue that may also be a concomitant part of the process of integration, but its effects may be confined to later stages of processing and, thus, it wouldn't have the same effect throughout the entire multimodal network. This hypothesis could be tested by looking at the role of spatial congruence as an integration cue, as spatial location is also carried in the perceptual signal and has been identified as an amodal stimulus property (although infant work suggests sensitivity to spatial relations occurs later than sensitivity to temporal structure (Lewkowicz, 1999)).

The regions sensitive to time in Experiment 2 also differed from the regions identified in Experiment 1. As discussed in Chapter 3, these differences likely reflect that the temporal incongruence in Experiment 2 was a 500 millisecond phase offset between the occurrence of the impact in each modality, while the temporal incongruence in Experiment 1 depended on a mixture of phase offsets, frequency differences, etc. However, this explanation assumes that different types of temporal incongruence would recruit distinct brain regions. One of the most intriguing pieces of evidence for this differentiation again comes from the developmental literature, where infants show different sensitivities to different types of temporal relations during the course of development (Lewkowicz, 1999). Initially, they can only detect onset synchrony between

two modalities. Then, they detect onset synchrony as well as the duration of the stimulus, followed by sensitivity to the temporal rate, and finally, they can detect rhythm congruence between modalities (Lewkowicz, 1999). In short, it is possible that these two fMRI studies may have only explored a fraction of the functional specialization for temporal relations within the multimodal network.

There is an intriguing synergy between our results showing separable networks for each cue and the developmental literature showing differential sensitivity to amodal properties during development. It is possible that the trajectory of development is linked to the specialization of the neural circuitry for each integration cue. Certainly, it is worth exploring whether the different types of temporal relations (synchrony, duration, rate, rhythm) recruit unique neural regions or shared resources. The data is still very sparse, and the amount of speculative content is high, but the possibility exists that the underlying neural architecture is focused on integration cues (or amodal properties) and not individual sensory modalities.

5.3 Experiment 3

As a complement to the fMRI studies, the EEG study focused on the timecourse of processing for each of the two integration cues, explicitly testing whether temporal synchrony influenced multimodal processing earlier than semantic congruence. The results showed the first effects of temporal synchrony occurred at 300ms post-stimulus onset, compared to 400ms for semantic congruence. In addition, the second time window that revealed an effect of temporal synchrony, 1000ms, also preceded the window of time identified for an effect of semantic congruence, 1100ms. These earlier temporal effects

could reflect the ability of low-level cues carried in the signal to exert earlier effects on multimodal processing.

Although the temporal effects do occur earlier than semantic effects, it was surprising that they did not influence the multimodal processing until 300ms post-stimulus onset. Prior EEG research has identified multimodal interaction effects on the early sensory potentials, thought to be the first detectable response to a unimodal stimulus. Thus, if temporal relations are fundamental to how modalities are integrated, a lack of temporal synchrony could be expected to affect the earliest evoked potentials; however, the early potentials were explicitly tested, and no difference was found between the temporally asynchronous condition and the temporally synchronous one. This was particularly surprising since the study by Steklenberg and colleagues (2007) suggested that the multimodal interactions on the auditory evoked potentials reflected when the visual modality predicts the timing of an action in the auditory modality. Their two events were very similar to ours, but the experiments differ based on the experimental design. In their study, a static visual image of the scene was presented for several hundred milliseconds before the onset of the impact sound, while in ours, the onset of the video and audio were simultaneous. Thus, it is possible that earlier effects of temporal synchrony are overshadowed by concurrent processing related to scene analysis, movement processing, etc. All of these concurrent processes may also be overshadowing any effect of temporal synchrony on the evoked potentials in our study, while these additional processes are eliminated in the design used by Steklenberg and colleagues.

The onset of temporal effects at 300ms may also reflect the duration of the acceptable time window for onset asynchronies between modalities that are still

perceived as temporal synchronous, a time period referred to as the intermodal temporal contiguity window, ITCW (Lewkowicz, 1999). Research with circles arbitrarily paired with tones has shown that participants are unable to detect asynchronies less than 80ms when an auditory stimulus precedes the visual stimulus, the ordering used in our TISCav condition (Bushara et al., 2001). Thus, perhaps the system would assume that the modalities are temporally synchronous (given the onset of the ambient background noise with the onset of the first visual frame of the movie), and no temporal synchrony effect would be expected until after the ITCW had passed. Even given the unknown delay between the neural generator and when the component would reach the scalp, a temporal synchrony effect at 300ms still seems a bit delayed for an ITCW of 80ms. However, it is a strong assumption to presume that the ITCW for simple stimuli is the same as the ITCW for semantically-rich, real-world events. Although not discussed in the analyses for this dissertation, we also collected behavioral data of explicit timing judgments about whether the stimuli were temporally synchronous or not, and despite the 500 millisecond offset, participants were not at ceiling. In fact, some participants were at chance levels for the synchrony judgment. Of course, decisional processes may be independent of the regions sensitive to the low-level temporal information in the perceptual signal; however, this does suggest that semantically-rich stimuli may be able to modulate the length of the ITCW. If so, the expected timeframe for the effects of temporal synchrony on multimodal processing may be closer to 300ms post-stimulus onset. This may reflect a strong interaction between temporal and semantic processing.

The absence of effects on the early sensory potentials also has intriguing implications for the provocative proposal that it is integration cues, and not sensory

modalities, that underlies the neural architecture of the multimodal network. Effects on the early sensory modalities are expected based on the current accounts that multimodal interactions rely on primary sensory areas; however, if the multimodal circuitry is specialized for integration cues, not modalities, the absence of effects on sensory potentials is less surprising. It is certainly the perceptual signal that carries the low-level information, and the multimodal circuitry is likely to be intertwined with the primary sensory cortices. However, multimodal interactions may occur in the circuitry specialized for the integration cues, not in the primary sensory areas, following a timecourse of its own.

5.4 Conclusion

This series of three studies establish a framework for investigating the functional role of regions consistently identified in the multimodal network. The generation of a novel set of environmental events enabled two integration cues to be manipulated independently of one another, providing an avenue to explore how the integration cues interact with one another. Here, we found that two integration cues, time and semantics, recruited separable networks of neural regions, identifying a preference in the right hemisphere for processing temporal synchrony and a preference in the left hemisphere for processing semantic congruency. In addition, effects of temporal synchrony preceded effects of semantic congruency, suggesting that low-level information carried in the signal may have earlier effects on multimodal processing.

These data suggest an alternative account for the underlying neural architecture; instead of relying on the interactions of the primary sensory areas (as suggested in Figure 1), the multimodal network may involve the interactions among neural circuitry specialized for particular integration cues. This proposal has intriguing synergy with behavioral research on amodal properties in both infants and adults, including a developmental trajectory of intermodal perception in infants that may reflect the gradual specialization of the multimodal network. Investigating this alternative account requires a focus on the integration cues, not functional contrasts, to identify the components of the multimodal network. By using integration cues, we can continue to explore the functional role of the regions consistently identified in multimodal research. In short, these experiments offer a framework better suited to revealing how integration cues interact with one another to form our integrated experience of events.

References

- Allman BL, Bittencourt-Navarrete RE, Keniston LP, Medina AE, Wang MY, Meredith MA (2008) Do cross-modal projections always result in multisensory integration? *Cereb Cortex* 18:2066-2076.
- Alpert GF, Hein G, Tsai N, Naumer MJ, Knight RT (2008) Temporal characteristics of audiovisual information processing. *J Neurosci* 28:5344-5349.
- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559-571.
- Bahrick LE, Lickliter R (2002) Intersensory redundancy guides early perceptual and cognitive development. *Adv Child Dev Behav* 30:153-187.
- Barracough NE, Xiao DK, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cognitive Neurosci* 17:377-391.
- Beauchamp MS (2005) Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3:93-113.
- Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809-823.
- Belardinelli M, Sestieri C, Di Matteo R, Delogu F, Del Gratta C, Ferretti A, Caulo M, Tartaro A, Romani GL (2004) Audio-visual crossmodal interactions in environmental perception: an fMRI investigation. *Cognitive Processing* 5:167-174.
- Benjamini YB, Yekutieli DY (2001) The Control Of The False Discovery Rate In Multiple Testing Under Dependency. *The Annals of Statistics* 39:1165-1188.
- Bernstein LE, Auer ET, Wagner M, Ponton CW (2008) Spatiotemporal dynamics of audiovisual speech processing. *Neuroimage* 39:423-435.
- Besle J, Fort A, Delpuech C, Giard MH (2004) Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci* 20:2225-2234.
- Bischoff M, Walter B, Blecker CR, Morgen K, Vaitl D, Sammer G (2007) Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia* 45:578-586.

- Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cereb Cortex* 17:2172-2189.
- Brosch M, Selezneva E, Scheich H (2005) Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *J Neurosci* 25:6797-6806.
- Buckner RL, Koutstaal W, Schacter DL, Rosen BR (2000) Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain* 123 Pt 3:620-640.
- Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* 21:300-304.
- Bushara KO, Hanakawa T, Immisch I, Toma K, Kansaku K, Hallett M (2003) Neural correlates of cross-modal binding. *Nat Neurosci* 6:190-195.
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110-1123.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649-657.
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14:427-438.
- Campbell R (2008) The processing of audio-visual speech: empirical and neural bases. *PHILOS T R SOC B* 363:1001-1010.
- Cummings A, Ceponiene R, Dick F, Saygin AP, Townsend J (2008) A developmental ERP study of verbal and non-verbal semantic processing. *Brain Res* 1208:137-149.
- Cummings A, Ceponiene R, Koyama A, Saygin AP, Townsend J, Dick F (2006) Auditory semantic networks for words and natural sounds. *Brain Res* 1115:92-107.
- Dhamala M, Assisi CG, Jirsa VK, Steinberg FL, Kelso JA (2007) Multisensory integration for timing engages different brain networks. *Neuroimage* 34:764-773.
- Doehrmann O, Naumer MJ (2008) Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Res* 1242:136-150.

- Doehrmann O, Naumer MJ, Volz S, Kaiser J, Altmann CF (2008) Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia* 46:2776-2786.
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron* 57:11-23.
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429-433.
- Fort A, Delpuech C, Pernier J, Giard MH (2002) Early auditory-visual interactions in human cortex during nonredundant target identification. *Brain Res Cogn Brain Res* 14:20-30.
- Fuster JM, Bodner M, Kroger JK (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405:347-351.
- De Gelder B, Bertelson P (2003) Multisensory integration, perception and ecological validity. *Trends Cogn Sci* 7:460-467.
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004-5012.
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278-285.
- Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473-490.
- Guthrie D, Buchwald JS (1991) Significance testing of difference potentials. *Psychophysiology* 28:240-244.
- Hein G, Doehrmann O, Müller NG, Kaiser J, Muckli L, Naumer MJ (2007) Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J Neurosci* 27:7881-7887.
- Hein G, Knight RT (2008) Superior Temporal Sulcus-It's My Area: Or Is It? *J Cogn Neurosci* 20:2125-2136.
- Hershenson M (1962) Reaction time as a measure of intersensory facilitation. *J Exp Psychol* 63:289-293.

- Hocking J, Price CJ (2008) The Role of the Posterior Superior Temporal Sulcus in Audiovisual Processing. *Cereb Cortex*.
- Howard, Templeton (1966) Human spatial orientation. London: Wiley.
- Jones JA, Callan DE (2003) Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14:1129-1133.
- Josse G, Tzourio-Mazoyer N (2004) Hemispheric specialization for language. *Brain Res Brain Res Rev* 44:1-12.
- Junghöfer M, Elbert T, Tucker DM, Braun C (1999) The polar average reference effect: a bias in estimating the head surface integral in EEG recording. *Clin Neurophysiol* 110:1149-1155.
- Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information? *Brain Struct Funct* 212:121-132.
- Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. *Neuron* 48:373-384.
- Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional imaging reveals visual modulation of specific fields in auditory cortex. *J Neurosci* 27:1824-1835.
- Klucharev V, Möttönen R, Sams M (2003) Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Res Cogn Brain Res* 18:65-75.
- Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207:203-205.
- Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279-292.
- Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE (2005) On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp Brain Res* 166:289-297.
- Laurienti PJ, Wallace MT, Maldjian JA, Susi CM, Stein BE, Burdette JH (2003) Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Hum Brain Mapp* 19:213-223.

- Laurienti PJ, Burdette JH, Wallace MT, Yen YF, Field AS, Stein BE (2002) Deactivation of sensory-specific cortex by cross-modal stimuli. *J Cogn Neurosci* 14:420-429.
- Lebib R, Papo D, de Bode S, Baudonnière PM (2003) Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neurosci Lett* 341:185-188.
- Lebib R, Papo D, Douiri A, de Bode S, Gillon Dowens M, Baudonnière PM (2004) Modulations of 'late' event-related brain potentials in humans by dynamic audiovisual speech stimuli. *Neurosci Lett* 372:74-79.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci* 25:5148-5158.
- Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA (2004) Human brain regions involved in recognizing environmental sounds. *Cereb Cortex* 14:1008-1021.
- Lewkowicz DJ (1999) The development of temporal and spatial intermodal perception. *Advances in Psychology* 129:395-420.
- Macaluso E, George N, Dolan R, Spence C, Driver J (2004) Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage* 21:725-732.
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746-748.
- Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221:389-391.
- Miller J (1982) Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol* 14:247-279.
- Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci* 25:5884-5893.
- Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb Cortex* 14:452-465.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14:115-128.

- Morgan ML, Deangelis GC, Angelaki DE (2008) Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* 59:662-673.
- Musacchia G, Sams M, Nicol T, Kraus N (2006) Seeing speech affects acoustic information processing in the human brainstem. *Exp Brain Res* 168:1-10.
- Naumer MJ, Doehrmann O, Müller NG, Muckli L, Kaiser J, Hein G (2009) Cortical Plasticity of Audio-Visual Object Representations. *Cereb Cortex* 19:1641-1653.
- Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27:11431-11441.
- Ojanen V, Möttönen R, Pekkola J, Jääskeläinen IP, Joensuu R, Autti T, Sams M (2005) Processing of audiovisual speech in Broca's area. *Neuroimage* 25:333-338.
- Olson IR, Gatenby JC, Gore JC (2002) A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Brain Res Cogn Brain Res* 14:129-138.
- Orgs G, Lange K, Dombrowski JH, Heil M (2006) Conceptual priming for environmental sounds and words: an ERP study. *Brain Cogn* 62:267-272.
- Orgs G, Lange K, Dombrowski JH, Heil M (2008) N400-effects to task-irrelevant environmental sounds: further evidence for obligatory conceptual processing. *Neurosci Lett* 436:133-137.
- Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, Sensolo S, Fazio F (2003) A functional-anatomical model for lipreading. *J Neurophysiol* 90:2005-2013.
- Pobric G, Jefferies E, Ralph MA (2007) Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proc Natl Acad Sci U S A* 104:20137-20141.
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001) Inaugural Article: A default mode of brain function. *Proceedings of the National Academy of Sciences* 98:676.
- Saito DN, Yoshimura K, Kochiyama T, Okada T, Honda M, Sadato N (2005) Cross-modal binding and activated attentional networks during audio-visual speech integration: a functional MRI study. *Cereb Cortex* 15:1750-1760.

- Saint-Amour D, De Sanctis P, Molholm S, Ritter W, Foxe JJ (2007) Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia* 45:587-597.
- Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory' processing. *Curr Opin Neurobiol* 15:454-458.
- Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res* 14:187-198.
- Senkowski D, Saint-Amour D, Kelly SP, Foxe JJ (2007) Multisensory processing of naturalistic objects in motion: a high-density electrical mapping and source estimation study. *Neuroimage* 36:877-888.
- Sestieri C, Di Matteo R, Ferretti A, Del Gratta C, Caulo M, Tartaro A, Olivetti Belardinelli M, Romani GL (2006) "What" versus "where" in the audiovisual domain: an fMRI study. *Neuroimage* 33:672-680.
- Sitnikova T, Holcomb PJ, Kiyonaga KA, Kuperberg GR (2008) Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *J Cogn Neurosci* 20:2037-2057.
- Skipper JI, van Wassenhove V, Nusbaum HC, Small SL (2007) Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex* 17:2387-2399.
- Spelke ES (1976) Infants' intermodal perception of events. *Cogn Psychol* 8:533-560.
- Stein BS, Meredith MM (1993) *The Merging of the Senses*. Boston, MA: MIT Press.
- Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci* 19:1964-1973.
- Sugihara T, Diltz MD, Averbeck BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neurosci* 26:11138-11147.
- Talairach J, Tournoux P (1988) *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers.
- Taylor KI, Moss HE, Stamatakis EA, Tyler LK (2006) Binding crossmodal object features in perirhinal cortex. *Proc Natl Acad Sci U S A* 103:8239-8244.

- Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Res Cogn Brain Res* 14:106-114.
- Ulrich R, Miller J, Schröter H (2007) Testing the race model inequality: an algorithm and computer programs. *Behav Res Methods* 39:291-302.
- van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17:962-974.
- van Atteveldt NM, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43:271-282.
- van Atteveldt NM, Formisano E, Goebel R, Blomert L (2007) Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36:1345-1360.
- Van Petten C, Luka BJ (2006) Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang* 97:279-293.
- Van Petten C, Rheinfelder H (1995) Conceptual relationships between spoken words and environmental sounds: event-related brain potential measures. *Neuropsychologia* 33:485-508.
- van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A* 102:1181-1186.
- van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45:598-607.
- Wager TD, Nichols TE (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *Neuroimage* 18:293-309.
- Welch RB (1999) Meaning, Attention, and the "Unity Assumption" in the Intersensory Bias of Spatial and Temporal Perceptions. *Advances in Psychology* 129:371-387.
- Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. *Psychol Bull* 88:638-667.
- Yuval-Greenberg S, Deouell LY (2007) What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *J Neurosci* 27:1090-1096.

- Zacks JM, Speer NK, Vettel JM, Jacoby LL (2006) Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychol Aging* 21:466-482.
- Zhou YD, Fuster JM (2000) Visuo-tactile cross-modal associations in cortical somatosensory cells. *Proc Natl Acad Sci U S A* 97:9777-9782.