The Role of Synesthetic Correspondence in Intersensory Binding: Investigating an Unrecognized Confound in Multimodal Perception Research

A Dissertation
Presented to
The Academic Faculty

by

Julia DeBlasio Olsheski

In Partial Fulfillment
of the Requirements for the
Doctor of Philosophy Degree in the
School of Psychology

Georgia Institute of Technology
December 2013

The Role of Synesthetic Correspondence in Intersensory Binding:
Investigating an Unrecognized Confound in Multimodal Perception Research

Approved by:

Dr. Bruce N. Walker, Advisor School of Psychology *Georgia Institute of Technology*	Dr. Gregory Corso School of Psychology *Georgia Institute of Technology*
Dr. Paul Verhaeghen School of Psychology *Georgia Institute of Technology*	Dr. Paul Corballis School of Psychology *The University of Auckland*
Dr. Dr. Richard Catrambone School of Psychology *Georgia Institute of Technology*	

Date Approved: July 8, 2013

ACKNOWLEDGEMENTS

I wish to thank my advisor, Bruce Walker, for his continued support, my Sonification lab-mates for their commiseration, and my committee for their guidance throughout this process. I am indebted to my undergraduate assistants, Hannah Fletcher and Angela Chiang, who helped me to collect countless hours worth of data. Gratitude and love to my husband, Carson Olsheski, for sticking with me through several ‘expected graduation’ dates that came and went; to my mother, Noreen DeBlasio, for cheering me on through daily phone calls; and to my son, Cedric, for making life precious again as we experience it through his eyes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TOJ	Temporal Order Judgment
SC	Superior Colliculus
ERP	Event-Related Potential
SPL	Sound Pressure Level
SOA	Stimulus Onset Asynchrony
ACS	Attentional Control Setting
fMRI	Functional Magnetic Resonance Imaging
PET	Positron Emission Tomography
PSS	Point of Subjective Simultaneity
JND	Just Noticeable Difference
RT	Reaction Time
LED	Light Emitting Diode
SDT	Signal Detection Theory
ITL	Intersensory Temporal Locking
AV	Audiovisual
ANSI	American National Standards Institute
QUEST	Quick Estimate by Sequential Testing
PST	Psychophysics Toolbox
MANOVA	Multivariate Analysis of Variance
SNARC	Spatial Numerical Associations of Response Codes
MNL	Mental Number Line
ASF	Auditory Spatial Facilitation

EXECUTIVE SUMMARY

Multimodal integration research did not become a prominent theme in the study of perception until quite recently. Over the last couple decades, the once-popular modular approach has rapidly been replaced with an understanding that information arising from multiple modalities must be integrated in order to form a cohesive representation of external events. The *unity assumption* asserts that this integration is driven by amodal commonalities between signals. Although there are other binding agents, commonality in time and space are the two that are most emphasized in the literature. From our very early interactions within the environment, we learn that external signals arising from a common time and location typically belong to a unitary object or event. Co-occurrences generally bring about redundant information. Our cognitive system facilitates pooling of neural activation arising from information redundancy. This multimodal neural pooling can lead to behavioral enhancement known as intersensory facilitation.

Synesthesia is an abnormal interrelation of the senses where stimulation of one sensory modality also triggers sensations in another sensory modality. The research-based evidence supports that synesthesia is truly a perceptual phenomenon, despite a history of being called into question. Further, there is evidence that all individuals experience a milder form of cross-sensory correspondences, sometimes referred to as *weak synesthesia*. This more common form manifests in language, perceptual similarity judgments, and perceptual interactions. These manifestations of weak synethesia are evidenced by measureable changes in performance. Though the correspondences of weak synesthesia are more abstract than those experienced by synesthetes, weak synesthetic links posses a quantifiable degree of *dimensional overlap*.

Attention is closely linked with perception and there are many important distinctions to be made within attention: automatic versus controlled processing, exogenous versus endogenous

control, and covert versus overt shifts. Interactions between the modes of attention must be understood to effectively design empirical tests of attention. Attention may be focused on spatial location, objects within external space, or information arising from one modality versus others. Generally it is supported that focusing attention on a particular location, object, or modality speeds up processing for that corresponding location, object, or modality (i.e., “prior entry”).

Studies of human performance have clearly demonstrated that the presence of information arising from one modality can affect processing of information arising from another modality. However, the influence between two modalities is not always symmetrical. While many asymmetries in multimodal links have been identified, their cause has not been firmly established. One major clue comes from intersensory bias, where one modality can dominate another by a comparatively heavier weighting during integration. Sometimes this dominance is complete, but more often, the bimodal percept is distinct from either input in isolation. In audiovisual interactions, vision tends to dominate spatial tasks whereas audition tends to dominate temporal tasks. However, one must be careful not to overemphasize a dichotomous classification of tasks as either spatial or temporal, since many tasks are more appropriately categorized as spatiotemporal. A continuum model is presented to account for a variety of findings that do not fit current theoretical predictions.

The influence of amodal commonalities other than space and time are also considered. Synesthetic correspondence is offered as a formerly overlooked amodal binding property. It is suggested that the degree of dimensional overlap determines, at least in part, the strength of binding between multimodal features of a given percept. The program of research presented here was designed to evaluate this suggestion as well as to examine the spatiotemporal continuum

model and to demonstrate that failure to consider synesthetic correspondence in the design of perception research has led to gaps in current perceptual theory.

The current program of research tests the following main hypotheses: 1) Synesthetic correspondence is an amodal property that serves to bind intersensory signals and manipulating this correspondence between pairs of audiovisual signals will affect performance on a temporal order judgment (TOJ) task; 2) Manipulating emphasis during a TOJ task from spatial to temporal aspects will strengthen the influence of task-irrelevant auditory signals; 3) The degree of dimensional overlap between audiovisual pairs will moderate the effect of synesthetic correspondence on the TOJ task; and 4) There are gaps in current perceptual theory due to the fact that synesthetic correspondence is a potential confound that has not been sufficiently considered in the design of perception research. The results support these main hypotheses. Finally, potential applications for the findings presented here are discussed.

1. INTRODUCTION

In his influential text, *The Principles of Psychology,* James referred to interrelations of the senses, or what he termed *synesthetic phenomena*, as “strange idiosyncrasies” (1983/1890, p.29). Similarly, Helmholtz taught that sensory data of the different senses are *totally incompatible* with each other. Despite an abundance of empirical evidence to the contrary, the influence of these seminal ideas on the approach to perception research has persisted. Indications that cases of multimodal integration are not merely perceptual flukes have been available for nearly a century. Schiller (1935) presented one of the first compilations citing numerous empirical works that demonstrated interrelations of the senses. Thus, even the early years of this field produced evidence that each modality does not function in isolation. Still, the importance of multimodal integration remains underplayed. As a result, perception has traditionally been oversimplified by a ‘modular approach’, treating each modality as though it were operating independently (e.g. Driver & Spence, 2004; Shams, Kamitani, & Shimojo, 2004). Only recently has the understanding emerged that multimodal interactions are the rule rather than the exception (Shimojo & Shams, 2001).

Transduction of environmental energy into meaningful code is essential for survival. Few of the defining characteristics of what can be considered living (e.g., metabolism, reproduction, motility) could be accomplished without a means of gaining information concerning external events. The perceptual system is capable of strategically exploiting the torrent of available input. This is no small feat considering both the ratio of irrelevant to critical information in the environment and the noise inherent in the sensory system. Consider the concept of sensor fusion. In order for a single sensor to avoid misses, sensitivity can be increased. However, increasing sensitivity is problematic because it also increases the rate of false alarms. The logical paradox between increasing sensitivity while minimizing false alarms can be solved by coupling different types of sensors (i.e., those that are tuned to different forms of environmental energy) in a ‘sensor fusion’ device (Stein, Stanford, Wallace, Vaughan, & Jiang, 2004). Activation criteria can be specified for each sensor, in addition to a requirement for temporal synchrony of activation across sensors.

Integration of multiple sensory signals was used by our single-celled progenitors and was likely among the earliest forms of sensory processing. Complex multicellular organisms evolved sensory receptors that could segregate information derived by type of receptor, thus increasing the ability to detect (sensation) and interpret (perception) environmental changes. Environmental pressures helped select the number and type of sensory receptors, and this became a driving force in speciation (Stein, et al., 2004). Sensory receptor diversification allows monitoring of a variety of environmental cues in parallel and also allows for the substitution of one system for another (e.g., using touch and sound to navigate in darkness). Parallel sensory processing can decrease detection thresholds and increase accurate identification (i.e., *intersensory facilitation)*. This is especially so during *valid co-occurrences,* or cases where different sensory modalities bring redundant information about the same external event (Bertelson & de Gelder, 2004). Organisms with the ability to assess multiple sensory modalities can utilize both segregated and pooled information by way of neural integration (Boenke, Deliano, & Ohl, 2009). Over time, certain brain areas of multisensory organisms became specialized with the ability to process unimodal information and other parts with multimodal information (Stein, et al., 2004).

1.1 A Brief History of Multimodal Research

Though multimodal perception did not receive the attention of researchers it now boasts until quite recently, the phenomenon was noted in early psychological studies. For instance, Brewster (1839) found that prisms causing visual misperception of inverted objects led to corresponding tactile misperception. Muller (1838) noted that the existence of the *ventriloquist effect* implies integration of auditory and visual inputs, thus limiting the degree of their respective autonomy (as cited in Bertelson & de Gelder, 2004). The first systematic multimodal research movement did not occur until the late 1950s. At this time, active movement research received the most attention. This line of research led to the discovery that adaptation occurred when a viewer observed his or her own hand in active movement through prisms, but adaptation did not occur when a viewer's hand was moved by experimenter or else remained immobile (e.g., Held, 1965). Recalibration was also observed when participants were presented with simultaneous noise bursts and prisimatically displaced light flashes (e.g., Canon, 1970). Wallach first developed a general view of perceptual adaptation as based on 'informational discrepancy' in the late 1960s (Wallach, Bacon, & Schulman, 1978).

Audiovisual spatial interaction has been studied in more detail than any other case of bimodal interaction because both sensory modalities are exteroceptive and allow for a degree of control over stimulus parameters. (*Note that this degree of empirical control is in contrast to the earlier active movement investigations as these involved proprioception.)* A well-studied and popular example of audiovisual interaction arising from informational discrepancy is the *Ventriloquist Effect* (Howard & Templeton, 1966). This illusion is experienced when a ventriloquist "throws" his or her voice such that the sound seems to be originating from the ventriloquist's dummy (or when sounds seem to be coming from the on-screen event when viewing a movie at the theater). The ventriloquist effect demonstrates a form of audiovisual conflict. That is, the auditory modality sends signals regarding the location of sound origination while the visual modality presents a conflicting (and in this case invalid) location of sound origination. This phenomenon is a case of perceptual bias where the spatial separation of discordant auditory and visual signals is underestimated or ignored up to a separation of 15-20° (Colin, Radeau, Soquet, Dachy, & Deltenre, 2002).

Meanwhile, another line of research was taking place that involved *on-line effects*. On-line effects are immediate responses to multimodal stimulation. One such effect is termed '*spatial fusion'* and is the impression that two discordant inputs belong to the same event (i.e., an invalid co-occurrence). More generally, *fusion* results in a congruous, unitary perception formed from incongruent inputs (Bertelson & Aschersleben, 2003). *Binding* is a broader term that includes illusory fusion as well as veridical multimodal integration. The resulting fused percept is typically distinct from either of the incongruent inputs (Bertelson & de Gelder, 2004). Another distinct form of on-line effect is termed '*immediate crossmodal bias'*. This is observed when participants are asked to make judgments involving stimuli occurring in one modality while ignoring distractors in a separate modality. For example, a selective localization task might require participants to make a forced left/right choice regarding the location of a visual target while ignoring the location of auditory distractors. Such immediate bias measurements have become the standard paradigm of the current field (Bertelson & de Gelder, 2004).

The earlier works considered thus far have been concerned with the perception of spatial attributes. However, a more recent line of investigation pertinent in a historical survey of multimodal perception research involves event classification (e.g., speech recognition). This is where the study of phenomena such as the *McGurk Effect* fit in (McGurk & MacDonald, 1976). The McGurk effect is an example of the class of fusion percepts in which the resulting precept is distinct from either input. In this case, the visual signal is the spoken phoneme 'ga' and conflicts with the auditory signal of the pronounced phoneme 'ba' resulting in the distinct percept 'da'.

Examples of environmental fusions that are outside of the linguistic realm have also been discovered. For instance, during musical note perception (Saldaña & Rosenblum, 1993) and the *parchment skin illusion* (Jousmäki & Hari, 1998). The first is an extension of the McGurk effect using audiovisual stimuli from a plucked versus a bowed cello string (Saldaña & Rosenblum, 1993). Observer perceptions are effected not only by the auditory stimulus, but also by the visual stimulus presented, as evidenced by reports during audiovisual cello signal conflict. The *parchment skin illusion* is a audiotactile fusion where presenting auditory feedback that accentuates the high frequency sounds of an observer rubbing his or her hands together near a microphone increases the observer's tactile perception of palm dryness (Jousmäki & Hari, 1998). Fusion studies have more recently been extended into the influence on emotion. Evidence has been put forth to show that the combination of tone of voice and facial expression can effect perception of emotion (de Gelder & Vroomen, 2000).

2. MULTIMODAL INTEGRATION

Much of what is understood about perception has been derived from unimodal studies, as these have traditionally been the focus within perception research. Though the body of unimodal studies have contributed greatly to our current understanding of perception, its overemphasis has left critical gaps in perceptual theory (Driver & Spence, 1998). Daily interactions within the environment are almost never truly unimodal. Rather, the experience is of a clamorous world where events engender stimulation in multiple modalities simultaneously. As a bus approaches an intersection, an observer may see the color, shape, and size of the bus, hear sound of the engine, detect vibration of the sidewalk, feel the heat coming off the bus, smell the exhaust fumes, and perceive the passage of time as the bus arrives. Even the highly controlled example of locating a light in a sensory deprivation chamber would require integration between visual, vestibular, and kinesthetic information (i.e., oculomotor feedback regarding the position of the eyes within their orbit as well as muscle feedback regarding the position of the trunk in relation to the head). Although both processing and external signals are noisy and can sometimes lead to perceptual errors, more often intermodal redundancy enhances performance. This is due to the pooling of corresponding neural activation (Bertelson & de Gelder, 2004). Though the sensory channels are differentiated, the experience is not of a fragmented collection of sensations, but rather an integrated perception (Soto-Faraco, Spence, & Kingstone, 2004). Sensory information from multiple modalities must be integrated online and continuously despite vast differences in initial cognitive encoding (Driver & Spence, 1998). Empirical and theoretical development in multimodal integration is crucial to achieving a more complete understanding of perception.

2.1.1 Assumption of Unity Theory

The *assumption of unity* holds that the greater the number of amodal properties shared by environmental signals, the more likely it is that they have arisen from a common external event (Vroomen & Keetels, 2006; Welch & Warren, 1980). Amodal properties are those that can be determined without reference to any one specific sensory modality (Brown, 2005). The most important of these amodal properties are commonality in time and space (Vroomen & Keetels, 2006). Other, sometimes debated, amodal properties include number, semantic content, temporal patterning, intensity, shape, and orientation to name a few (Brown, 2005; Colin, et al., 2002; Vatakis & Spence, 2006). Following this assumption, multisensory integration should be reduced or absent when signals are separated too far in time or space. Indeed, our sensory experience of the world is that sights, sounds, smells, and other feedback originating from a single event generally occur around the same time and in a common location. Empirical evidence also supports the assumption of unity theory. Stimulation that occurs within a specified temporal and spatial window is generally bound into a single perceptual event (Koppen & Spence, 2007). The unity assumption is posited as the mechanism by which both illusory and veridical binding occur.

2.1.2 Maturation of Multimodal Integration

Sensory experience has a profound role on the maturation of intersensory processing. Two conflicting theories regarding the development of multimodal integration are prominent. The first asserts that the senses start out independent from one another and then intersensory associations develop with experience during the first months and years of life (King, 2004). The second argues that the senses are unified at birth and gradually become differentiated from each other over the course of development (Bower, 1971). The empirical evidence regarding infant perception of the relation between modalities provides some support for both ideas. Some intersensory perceptual abilities are present at birth while others emerge gradually. Newborn infants associate crossmodal sensory information on the basis of temporal relations while spatial relations come to play a more important function later in life (King, 2004). Auditory functioning is more fully developed at birth than is vision. By an age of six months, infant audition reaches nearly adult functional levels (DeCasper & Spence, 1986). Taking the comparative developmental maturity and the superior temporal resolution of audition into account, it logically follows that perhaps newborns rely more on temporal relations than on spatial ones because their auditory systems are more reliable than their immature visual systems. That is, the heavier reliance on temporal relations through the first several months of life may be due to greater variance inherently produced by the less developed visual system.

2.1.3 Behavioral Consequences of Multimodal Integration

Enhanced performance in the presence of bimodal versus unimodal stimuli is widely reported. Generally it is thought that the redundancy of multimodal stimuli, or the so-called *target redundancy effect,* leads to performance enhancement. Facilitatory performance effects include faster reaction time, increased accuracy, and decreased detection thresholds (Teder-Salejarvi, Di Russo, McDonald, & Hillyard, 2005; Wallace, 2004). *Intersensory facilitation of reaction time* is a term specifically given to the behavioral result of more rapid RTs to multisensory compared to unisensory events. This specific term was first used decades before the more recent surge of interest in multimodal research (Bernstein & Edelstein, 1971). Two competing explicative models have been offered for intersensory facilitation of reaction time. The *race model* posits that faster bimodal RTs arise because the more quickly detected of the two cues triggers the overall response. The *coactivation model* suggests that parallel processing of multiple sensory channels interact at the sensory processing, response selection, and/or motor execution level and that these interactions facilitate performance (Giard & Peronnet, 1999).

2.1.4 Neural Basis of Multimodal Integration

Intersensory facilitation has more recently been investigated at the neural level. The superior colliculus (SC), a midbrain structure most noted for its role in initiating and controlling gaze shifts, is a polysensory area containing a variety of afferent inputs from areas traditionally associated with early unimodal processing. These afferents converge onto single neurons, rendering them multisensory (Stein, et al., 2004). Superior colliculus multisensory neurons are topographically ordered such that the resulting receptive fields overlap in sensory space. Single-cell recordings of multisensory SC neurons have revealed that integration is most often achieved through summation of the individual inputs. The resulting neural activity is sometimes superadditive (i.e., greater than the sum of unimodal inputs in isolation) and rarely subadditive (Stein, et al., 2004). Note, however, that even subadditive multimodal integration will yield greater neural activity than either modality in isolation. For example, a nonzero result of the bimodal waveform minus the summation of unimodal waveforms is often used to define the processing stage at which multimodal integration has occurred in event-related potential (ERP) studies (Gondan, Vorberg, & Greenlee, 2007).

The degree of multimodal neural enhancement can be described by the following rules: 1) the spatial rule; 2) the temporal rule; and 3) the rule of *inverse effectiveness* (Smits, 2009). The spatial rule outlines that multimodal neural response increases with degree of spatial overlap because stimuli fall within increasingly aligned receptive field excitatory regions. Conversely, response depression can result from spatially disparate presentations where one stimulus lands within the inhibitory region of the other modality’s receptive field. The temporal rule outlines that multimodal neural response will increase with perceived synchrony of the multimodal inputs due to maximal overlap in respective periods of peak activity. Finally, the rule of *inverse effectiveness* states that enhancement is greatest when multimodal stimuli are presented at or near their respective response thresholds (Meredith & Stein, 1983). Taken together, the behavioral and neurological evidence fit well with the assumption of unity's emphasis on spatial and temporal commonality across multimodal paired inputs.

2.2 Synesthesia

Synesthesia, meaning "to perceive together", is an atypical form of interrelation of the senses. Specifically it is a condition in which an otherwise normal person experiences sensations in one modality when a second modality is stimulated (Martino & Marks, 2001; Ramachandran & Hubbard, 2001b). This definition describes what is sometimes termed *crossmodal synesthesia*, but there are forms of *intermodal synesthesia* as well (Grossenbacher & Lovelace, 2001). In either case, the *inducer* is the stimulation that triggers the synesthetic perception, called the *concurrent*. The specific concurrent induced is consistent across multiple presentations of a given inducer, but the inducer-concurrent pair varies from one synesthete to the next. These perceptions are not experienced by the majority of people under comparable conditions (Grossenbacher & Lovelace, 2001). Female synesthetes are more common than males with a ratio of 6:1 (Baron-Cohen, Burt, Smith-Laittan, Harrison, & Bolton, 1996). Prevalence estimates vary greatly from 1 in 20 to 1 in 20,000 individuals, partially due to varying methodologies and definitional criteria (Ramachandran & Hubbard, 2001b).

There exist many varieties of synesthesia. These are classified by the category of triggering inducers and the respective category of induced concurrents. For example, sound-color synesthesia (i.e., seeing hues when hearing certain auditory attributes) has been documented as far back as Greek antiquity when philosophers sought to quantify the ‘color of music’ (van Campen, 1997). Currently, grapheme-color synesthesia (i.e., experiencing hues upon viewing certain printed symbols) is widely accepted as the most commonly occurring form among synesthesia researchers (Ramachandran & Hubbard, 2001b). Historically though, the scientific community as a whole has largely discounted the phenomenon altogether. The more common dismissals include that the ‘so-called’ synesthetes are crazy, looking for attention, having hallucinations elicited by prior drug use, merely remembering childhood associations (e.g., a childhood book with the number 3 printed in blue), or simply being metaphorical (e.g., sharp cheese or bitter cold). While it is true that pharmacological synesthesia can be experienced while under the influence of certain drugs (e.g., LSD or mescaline), this form does not explain developmental synesthesia or acquired synesthesia. Developmental synesthesia appears to emerge in early childhood, run in families, and have an X-linked dominant mode of transmission (Grossenbacher & Lovelace, 2001). Acquired synesthesia is more rare and is the result of brain injury (Grossenbacher & Lovelace, 2001). The lingering doubts regarding the validity of synesthesia as a truly perceptional phenomenon are surprising given the abundance of compelling empirical evidence.

2.2.1 Evidence that Synesthesia is a Truly Perceptual Phenomenon

Inducer-concurrent pairings are stable within an individual synesthete over long periods of time. In one study, a group of nine speech-color synesthetes were compared to an age-, gender-, and IQ-matched group of nine controls. All participants were asked to describe the color sensation experienced on hearing each of a list of 130 items. The control participants were informed they would be retested one **week** later. The experimental group members (i.e., synesthetic participants) were not told they would be retested and were brought back one **year** later. The results showed that the synesthetes were far more accurate (92.3%) one year later than the control group participants were (37.6%) one week later (Baron-Cohen, Harrison, Goldstein, & Wyke, 1993). Though these results can be used as evidence against several of the common rebuttals, they cannot refute the possibility that the associations are based on memory.

Perhaps the most compelling evidence for the validity of synesthesia as a perceptual phenomenon, rather than one based on memories, comes from studies of perceptual grouping and the pop-out effect. Indeed, these two tasks are often used to diagnostically test whether a phenomenon is truly perceptual. Ramachandran and Hubbard (2001a) found that when presenting participants with a black and white array of ‘5’s with mirror image ‘2’s embedded and arranged so as to form a shape, the synesthetic participants were much more accurate than control participants (81.25% versus 59.4%) in detecting the embedded shapes (see Figure 1). Though the control participants did perform better than chance, the synesthetes reported much lower task difficulty due to pop-out arising from the induced colors associated with the ‘5’s and the ‘2’s respectively (Ramachandran & Hubbard, 2001a).

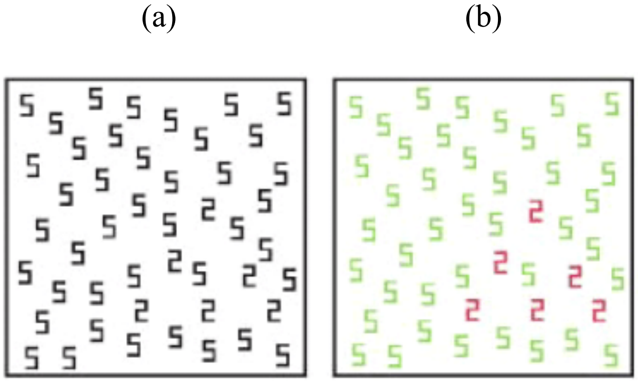
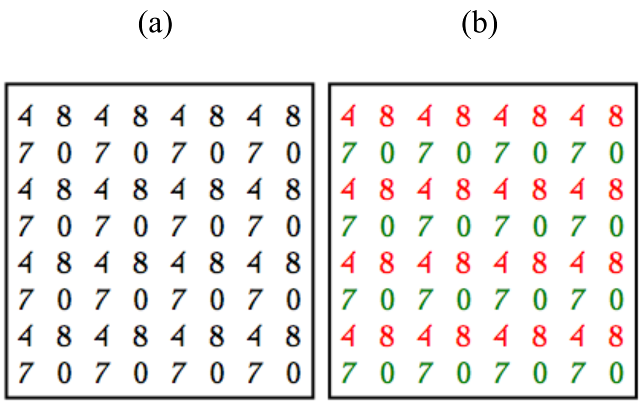


Figure 1. Example display used in Ramachandran and Hubbard's synesthetic pop-out test. Panel (a) depicts the black and white array as presented to participants. Panel (b) is an illustrative example of what synesthete participants report to perceive.

Ramachandran and Hubbard (2001a) also presented synesthetic findings involving perceptual grouping principles. When presented with an array of numbers and given a forced-choice between reporting horizontal rows or vertical columns, the probability that control participants will choose one orientation over the other is at chance. However, participants can be biased to report one orientation over the other on the basis of shape similarity (e.g., 4 and 7 or 0 and 8 are more readily grouped than 4 and 0). Matrices were carefully constructed such that shape-based grouping would lead to one orientation while (induced) color-based grouping would lead to the alternative orientation (see Figure 2). As hypothesized, synesthete participants grouped based on induced colors approximately 90% of the time while control participants grouped based on shape approximately 60% of the time.



4	8	4	8	4	8	4	8
7	0	7	0	7	0	7	0
4	8	4	8	4	8	4	8
7	0	7	0	7	0	7	0
4	8	4	8	4	8	4	8
7	0	7	0	7	0	7	0
4	8	4	8	4	8	4	8
7	0	7	0	7	0	7	0

4	8	4	8	4	8	4	8
7	0	7	0	7	0	7	0
4	8	4	8	4	8	4	8
7	0	7	0	7	0	7	0
4	8	4	8	4	8	4	8
7	0	7	0	7	0	7	0
4	8	4	8	4	8	4	8
7	0	7	0	7	0	7	0

Figure 2. Illustration of Ramachandran and Hubbard's synesthetic perceptual grouping test. In this example, (panel a) non-synesthetes would tend to report that the numbers can be perceptually grouped in columns based on shape, (panel b) whereas a synesthete with the specific number-color mappings depicted would tend to report that the numbers can be perceptually grouped in rows based on induced color.

2.2.2 Proposed Neural Models of Synesthesia

Grossenbacher and Lovelace (2001) propose that a disinhibited feedback mechanism is the neurological underpinning of synesthesia. Signals from multiple sensory pathways converge on specific multimodal areas of the brain and are reciprocated to unimodal areas via feedback connections. These authors suggest that, although in most people the feedback connections are inhibited to avoid synesthetic perception, in synesthetes the feedback is disinhibited. This means that inducer sensory signals are sent along the unimodal pathway and feedback signals are propagated back along the concurrent pathway (see Figure 3).

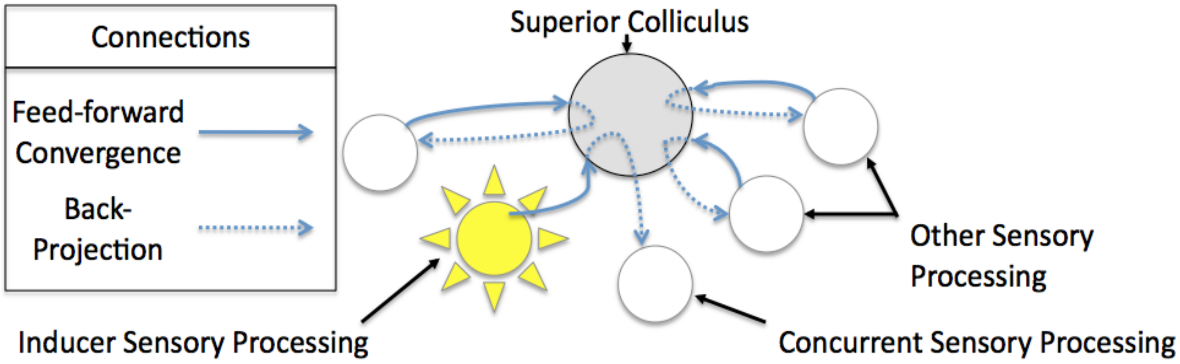


Figure 3. Illustration of Grossenbacher and Lovelace's Disinhibited Feedback Mechanism demonstrating how an event processed along the inducer sensory pathway is sent to the concurrent sensory processing area.

Ramachandran and Hubbard (2001b), on the other hand, propose that synesthesia may be the result of some form of cross-wiring. Specifically, these authors point to the substantial number of prenatal connections between different brain areas that exist in the immature brain. These connections are removed through a process called pruning prior to birth (Ramachandran & Hubbard, 2001b). They suggest that there may be a genetic mutation leading to a failure of pruning between synesthetically corresponding neural areas. Potentially, both theories may explain some part of synesthetic perception. Regardless, the abundance of evidence pointing to the legitimacy of this phenomenon should not be ignored by mainstream neuroscience because of the obvious implications for understanding normal sensory functioning (Ramachandran & Hubbard, 2001b). This is especially so in regards to the study of normal multimodal integration.

2.3 Normal Crossmodal Correspondences: “Weak Synesthesia”

Martino and Marks (2001) used the terms *weak synesthesia* and *strong synesthesia* to distinguish a form of intersensory correspondence that is much more common than that experienced by synesthetes. Whereas only a synesthete demonstrates the strong form as experienced through perception of the concurrent sensation, weak synesthesia is experienced by most normal observers. Weak synesthesia is expressed through language (i.e., synesthetic, or crossmodal metaphor), perceptual similarity (i.e., congruence relations), and perceptual interactions during information processing (i.e., Garner interference). Whereas the inducer-concurrent relationship delimitative of strong synesthesia is literal, unidirectional, and absolute, the analogous crossmodal correspondences of weak synesthesia are metaphorical, bidirectional, and contextual (Martino & Marks, 2001).

2.3.1 Language

One expression of weak synesthesia is evidenced through language. Synesthetic metaphors are common expressions in which words appropriate for perception in one modality are used to describe an experience in another modality (Marks, 1982). Examples in language such as ‘sharp cheese’, ‘bitter cold’, or ‘loud shirt’ do not come across as odd even though, for instance, cheese is in fact soft (tactile) while we speak of it as being sharp (gustatory). Such examples are prevalent in everyday language because there are underlying mechanisms that drive these synesthetic connections in a naturalistic way (Ramachandran & Hubbard, 2001b).

Further evidence of weak synesthetic links comes from the study of the evolution of language. Köhler first developed a set of abstract stimuli and asked participants to assign a given name to each of them (as cited in Werner & Wapner, 1952). These figures were further developed and explored by Werner and Wapner (see Figure 4). When given the pair of abstract shapes illustrated in Figure 4 and asked to assign one shape the name ‘kiki’ and the other ‘bouba’, participants will consistently assign ‘kiki’ to the abstract figure with sharp corners and ‘bouba’ to the rounded one. These results are consistent across 95% of observers though they have no prior experience with either the visual stimuli or the invented names. This is because the sharp visual changes of one shape mimic the abrupt tongue/palate and phonemic inflection of the sound “kiki”. This study provided the first strong evidence of naturalistic constraints on sound-object mapping. These and similar findings have implications on the development of proto-languages.

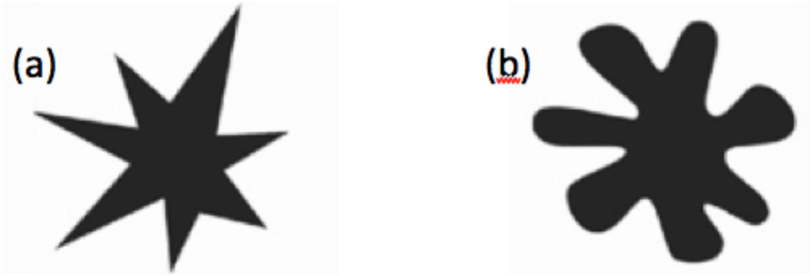


Figure 4. The name Kiki is most often assigned to a trapezoid with more jagged lines and acute points (panel a), while the name Bouba is most often assigned to the more rounded shape with smooth curves (panel b).

2.3.2 Perceptual Similarity

Weak synesthesia is also evidenced through observer judgments of perceptual similarity, specifically during crossmodal magnitude matching. *Magnitude estimation* is a psychophysical technique where observers are first asked to assign a value to a stimulus at a moderate reference intensity and then to continue assigning values as intensity is varied randomly from the reference (Schiffman, 2001b). For example, a 1000 Hz sound stimulus may be presented at 40 dB while an observer is asked to assign a value between 10 and 100 to that stimulus. The sound is then presented over many trials at 1000 Hz alternating between the reference and randomly varying sound pressure levels (SPL). The observer’s magnitude estimations are then correlated with actual intensity to gauge the relationship between perception and sensation (Stevens, 1936). While Fechner first discovered that there is a calculable relationship between stimulus attributes and their perceptual consequences, Stevens later found that though most of these are linear (e.g., weight and heaviness), some are logarithmic. For example, brightness of a light increases by the intensity raised to one-third power ($\mathbf{I}^{1/3}$) whereas pain from electric shock increases by a power of three ($\mathbf{I}^3$) (Schiffman, 2001b).

Similar to magnitude estimation, *magnitude matching* is a psychophysical technique where observers are asked to adjust the level of one stimulus attribute to match varying levels of another presented attribute (Marks, 2004). For example, participants might be asked to match the intensity of a hue to the size of an object. *Crossmodal matching* is similar to magnitude matching; only now, the stimulus attributes being compared are from two separate modalities. For example, an observer may be asked to adjust the frequency of a sound stimulus to match the hue of a visual stimulus (Marks, 2004). *Congruence relations*, the sharing of relative values of perceptual attributes from different stimuli, generally demonstrate a polar relationship. These relations have been found between many perceptual pairings such as pitch and color, loudness and brightness, or vertical position and pitch.

Congruence relations may be intramodal or crossmodal. Intramodal congruence relations are seen between features within a single modality that share a common dimension. For instance, vertical spatial position and direction indicated by a visual arrow are consistently related such that high is matched with up-pointing, and low is matched with down-pointing, arrowheads. However, the congruence relations between crossmodal pairings (and even certain intramodal pairings) are abstract. An example of an abstract intramodal pairing is that of auditory pitch and loudness. Albeit the pair shares a common modality, they have no common referent so the relation is abstract. Crossmodal congruence relations, however, remain the most abstract, figurative, or even metaphorical. Evidence suggests that crossmodal congruence relations drive similarity perception at the holistic, rather than the attribute, level. This would imply that similarity perception occurs at a relatively early perceptual level (Melara, 1989). Some authors remain dubious of this holistic view because it would imply that information remains undifferentiated with regard to modality during early perceptual processing (e.g., Marks, 2004). Regardless, there is agreement that there must be some form of early processing, as indicated by results of perceptual interaction studies.

2.3.3 Perceptual Interactions

Finally, weak synesthesia is demonstrated via perceptual interactions. Crossmodal perceptual interaction is evidenced when the interpretation of data in one sensory modality is influenced by information available in another sensory modality (Bertelson & de Gelder, 2004). Research has focused on two related, but distinct types of crossmodal perceptual interaction: *Garner interference* and *congruence effects.* One or both types of perceptual interaction emerge during *speeded classification tasks*. Speeded classification tasks involve selective attention where observers must identify specific attributes of a target stimulus as quickly as possible while ignoring other, irrelevant attributes (Ben-Artzi & Marks, 1995).

Garner interference. Garner's early studies utilizing the speeded classification paradigm were focused on intramodal attributes. For example, participants might be asked to selectively attend to size while ignoring color of a visual stimulus (Garner, 1977). In many cases, the orthogonal variation of an irrelevant attribute yields slower reaction time and decreased accuracy in the speeded classification of the relevant attribute when compared to baseline. This trend is termed 'Garner interference' (Melera & Marks, 1990). Garner interference traditionally has been used to distinguish separable versus integral dimensions. Stimulus attributes are said to be separable if the orthogonal variation of one does not affect classification of the other. In other words, if observers are able to selectively attend to one attribute while successfully filtering out the irrelevant attribute then the attributes are separable (Marks, 2004). Conversely, when the orthogonal variation of an irrelevant attribute does effect speeded classification of the target attribute, the two are defined as integral.

Garner (1977) found that color and size are separable dimensions, as demonstrated by nearly identical performance on baseline tasks (irrelevant dimensions constant) versus filtering tasks (irrelevant dimensions orthogonally varied). This indicates that orthogonal variation of the irrelevant attribute did not affect classification of the target attribute. Auditory pitch and loudness, however, are integral variables and although they share no common referent, the variation of one drastically interferes with classification (speeded or otherwise) of the other. In fact, perceived auditory intensity, or loudness, is highly sensitive to frequency variation (see Figure 5).

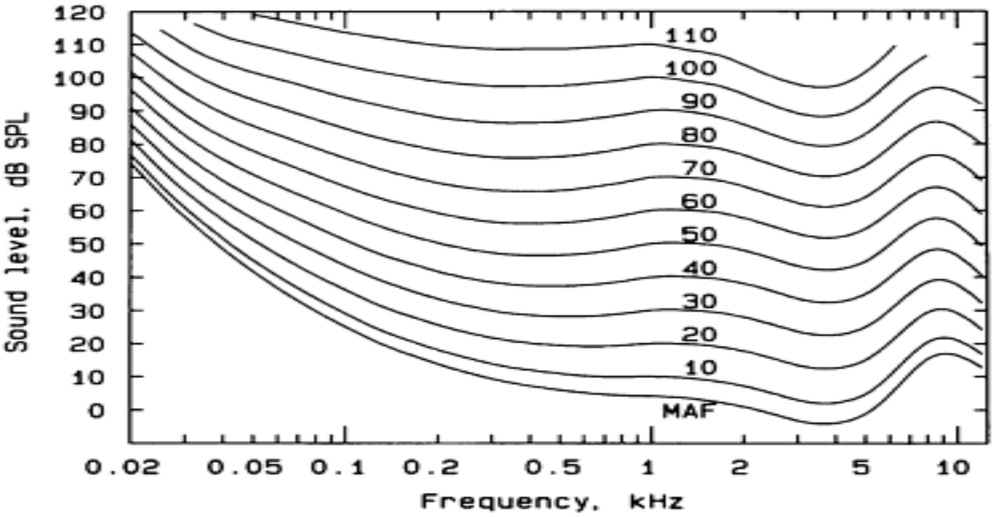


Figure 5. An equal-loudness contour graph depicts the perceived loudness of frequencies at varying SPLs. Each contour on the graph is measured in phons where a phon is defined as the number of dB SPL of a standard tone (1000-Hz) that would be perceived as equally loud as a test tone (Schiffman, 2001).

Using a reference tone of 1000 Hz at 40 dB SPL, the average listener will perceive other frequencies to be of equal loudness at intensities varying from approximately 90 dB SPL at 20

Hz (the lower frequency threshold of hearing), down to approximately 33 dB SPL at 4000 Hz (the approximate peak of human auditory sensitivity), and gradually rising again near the upper frequency threshold of hearing to approximately 48 dB (Schiffman, 2001b). If plotting this example, all points falling on this contour would measure 40 phons. Note that it is no coincidence that the frequency range of greatest sensitivity overlaps with the frequency range of human speech and other sounds critical for human survival (i.e., 2-5 kHz). Finally, the speeded classification paradigm has been used more recently to discover crossmodal integral variables. For example, pitch and spatial position demonstrate Garner interference when orthogonal variation of one decreases performance on speeded classification of the other when compared to baseline performance (e.g., Melara & O'Brien, 1987).

Congruence effects. A second type of crossmodal perceptual interaction revealed by the speeded classification paradigm is evident by enhanced performance when stimuli have congruent aspects when compared to performance for stimuli with incongruent aspects. 'Congruent aspects' are stimulus dimensions that are matched in at least some loose, general way when compared to 'incongruent aspects' whose dimensions are mismatched (Marks, 2004). Using a previously described example, an up-pointing arrowhead would be considered congruent, or matching, with a high spatial position, and incongruent, or mismatching, with a low spatial position (see Figure 6). Some describe congruent aspects in more detail by referring to 'dimensional overlap' (Kornblum, Hasbroucq, & Osman, 1990). In the case that there is no dimensional overlap in a stimulus pair, then the related task is non-congruent. (This is associated, though not synonymous, with the idea of separable variables discussed previously). If there is dimensional overlap in a stimulus pair, then the related task is considered congruent when the mapping is such that the polarities are matched, and incongruent if the mapping is such that the polarities are mismatched. The degree of overlap is a more concrete way to discuss the variability of dimensional match, which can range from quite literal to much more abstract, metaphorical, or vague. Also, the concept of dimensional overlap delineates integral variables according to the degree to which they affect performance.

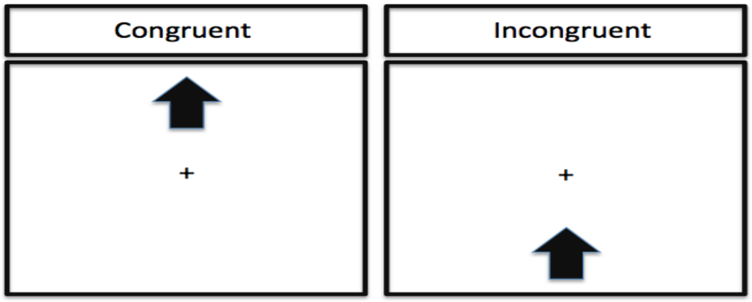


Figure 6. Example of congruent versus incongruent stimulus dimensions. The left panel illustrates congruent dimensions of the up-pointing and high spatial position of the arrow on the left. The right panel illustrates incongruent dimensions of up-pointing but low spatial position of the arrow on the right.

Though there are some researchers who suggest that congruence effects and Garner interference arise from a common mechanism (Ben-Artzi & Marks, 1995), the two are distinct perceptual interactions. Whereas congruence effects are observed as performance enhancement when comparing congruent to incongruent stimulus pairs, Garner interference is a performance decrease observed when comparing orthogonal variation of an irrelevant, but integral dimension compared to baseline. Often, but not always, the two forms co-occur. Garner interference is more likely to occur in isolation than are congruence effects (Marks, 2004). There are, however, exceptions; for example, when viewing an illustration of a hand, participants were better able to discriminate whether a near-threshold tactile stimulus was delivered to their index finger versus the pinky when the illustration included an arrow pointing to the congruent, rather than incongruent finger. Further, performance in both the congruent and incongruent conditions exceeded that during baseline, or the no-pointer condition (Pasalar, Ro, & Beauchamp, 2010).

This is an illustration of congruence effects with no accompanying Garner interference. Whereas Garner interference pertains to a global decrease in performance across levels of the variables, congruence effects represent within-stimulus-level performance variability.

3. DISTINCTIONS WITHIN ATTENTION

Despite James' assertion that "everyone knows what attention is" (1983/1890, p.403), the term 'attention' is used in varying psychological contexts and its meaning is not universally agreed upon (Pashler, 1998). There is agreement that attention is closely linked to perception. Titchner's law of prior entry states that attended stimuli are processed more rapidly than unattended stimuli (Titchner, 1908). Driver and Spence (2004) elaborate that 'selective attention' is a catchall term encompassing cognitive activities that allow deeper processing of certain information. Information that receives deeper processing generally has more behavioral relevance to our current goals, higher intrinsic salience, and/or greater biological significance. Additionally, attention is largely regarded as a finite, cognitive resource whose devotion to some task leaves less of that resource available for other tasks (Galotti, 2004).

3.1 Automatic versus Controlled Processing

Automatic processing is associated with easy or over-learned tasks and operates in parallel, indicating no strain on capacity limitations (Galotti, 2004). Automatic processing initiates a learned sequence that proceeds without observer control (Schneider & Shiffrin, 1977). The activation of this sequence is almost guaranteed given a particular input configuration, but only after an appreciable amount of training (Schneider & Chein, 2003). Controlled processing is required for more cognitively demanding tasks. Controlled processing usually operates serially, is capacity limited, and is under conscious control (Galotti, 2004). Unlike automatic processing, the sequence of elements are not retrieved from long term memory, but instead are temporarily activated by the observer in a manner that requires attention (Schneider & Shiffrin, 1977). The benefit of this type of capacity-costly processing is that it can be applied in novel situations as it does not require endless hours of training (Schneider & Chein, 2003).

3.2 Exogenous versus Endogenous Control of Attention

Voluntary attentional shifts are termed *endogenous* and are controlled by perceiver generated signals and top-down processes (Klein & Shore, 2000). Endogenous attention begins with central focus; it is characterized as voluntary or strategic, and allows for selection of information that is relevant to our current conscious goals (Driver & Spence, 2004; Goldsmith & Yeari, 2003; Klein & Shore, 2000). The effects of endogenous attention tend to be larger and longer-lasting (Driver & Spence, 1998). Conversely, *exogenous* attention begins with a widely-spread focus and is reflexive, automatic, and stimulus-driven (Goldsmith & Yeari, 2003). Attention may be drawn under exogenous control when spatially or temporally asymmetric stimulation occurs (Klein & Shore, 2000); especially when the stimulation is received simultaneously by more than one sensory channel (Spence, McDonald, & Driver, 2004). In his pioneering text, William James (1890) referred to these two distinct forms as active and passive attention. Qualitative differences between the two forms of attention have been enumerated and it is believed that each arises from a distinct neural substrate (Driver & Spence, 1997; Peelen, Heslenfeld, & Theeuwes, 2004). Therefore, it is important that studies of attention consider endogenous and exogenous forms separately.

Posner's (1980) cuing paradigm has been used extensively to distinguish exogenous and endogenous visual attention. Typically with this cuing paradigm, predictive central cues are used to prompt endogenous attention while non-predictive peripheral cues are used to attract exogenous attention. The general results indicate that performance is higher for targets appearing in cued rather than uncued locations for both endogenous and exogenous attention

(Peelen, et al., 2004). However, in the exogenous condition only, a pattern emerges called *inhibition of return*. Namely, when the cue precedes the target by more than about 250 ms, performance for targets in the correctly cued location is decreased (Posner & Cohen, 1984).

Receptor orienting has been shown to be more rapid under exogenous attentional control. Further, exogenous attention appears to be unaffected by memory load or predictive validity of signaling stimuli, unlike endogenous attention (Klein, 1994). Early reports suggested that both forms of attention utilized the same attentional resources, but attributed the behavioral differences to the 'transportation' mechanism. Transportation is fast and automatic for exogenously controlled attention while slow and voluntary for endogenously controlled attention. Klein (1994) suggests that it is not only the metaphorical vehicle that differs for the two forms of attention, but also the "passenger" (p. 169). Specifically, that the mechanisms, pathways, and processing characteristics are distinct between attention under the two forms of control. Some refer to attention as being endogenously pushed versus exogenously pulled.

3.3 Covert versus Overt Attention

Exogenous and endogenous attention are often mistakenly aggregated or else studied independently. However, normal function involves a complex interplay between the two and their control over both overt and covert orienting in a manner that may be cooperative or competitive (Klein, 1994; Klein & Shore, 2000; Peelen, et al., 2004). Overt attention involves physical orienting of sensory receptors to focus on an external event. Covert attention involves an attentional shift that occurs despite a lack of physical movement. For instance, covert visual attention is shifted without eye movement (Driver, 2001). There is agreement that within exogenous attention, covert and overt attention are strongly linked, and that this may be because both covert shifts and receptor orientation are cued by similar external events (Klein & Shore, 2000). However, whether or not endogenously controlled covert attention is simply a manifestation of unexecuted motor plans (i.e., programmed but uninitiated overt shifts) is a subject of much debate (Klein & Shore, 2000).

Variations of the basic Posner cuing paradigm have been used to isolate covert and overt visual attention through the use of eye-tracking devices (Klein, 1994; Spence & Driver, 1997). Alternatively, some researchers lacking access to eye monitoring equipment have designed studies utilizing stimulus onset asynchronies (SOAs) too short to allow for saccadic eye movement. These researchers defend this methodology by pointing out that the display would be blank by the time a potential overt eye-movement were completed (Klein, 1994).

3.4 Interactions Between the Modes of Attention

3.4.1 Overt

The interplay between exogenous, endogenous, overt, and covert attention is complex and only partly understood. There is evidence for cognitive networks of inhibitory gating mechanisms that can prevent overt orienting. When an external source of continuous stimulation is removed, it is hypothesized that the sudden offset can trigger an exogenous deactivation of the inhibitory gating, thus allowing for the extremely rapid endogenous overt orienting (e.g., express saccades) that has been observed under such circumstances (Klein & Shore, 2000; Trappenberg, Dorris, Munoz, & Klein, 2001). Further, it is suggested that endogenous disengagement from a fixation stimulus is a necessary preparatory step of overt orienting. It has been shown that when a constant unique target stimulus is presented with changing distractors, participants follow instructions and endogenously orient to the target singleton. However, when a new distractor is added at the time of target presentation, exogenous attention is pulled toward the onset of the new distractor (see Figure 7).

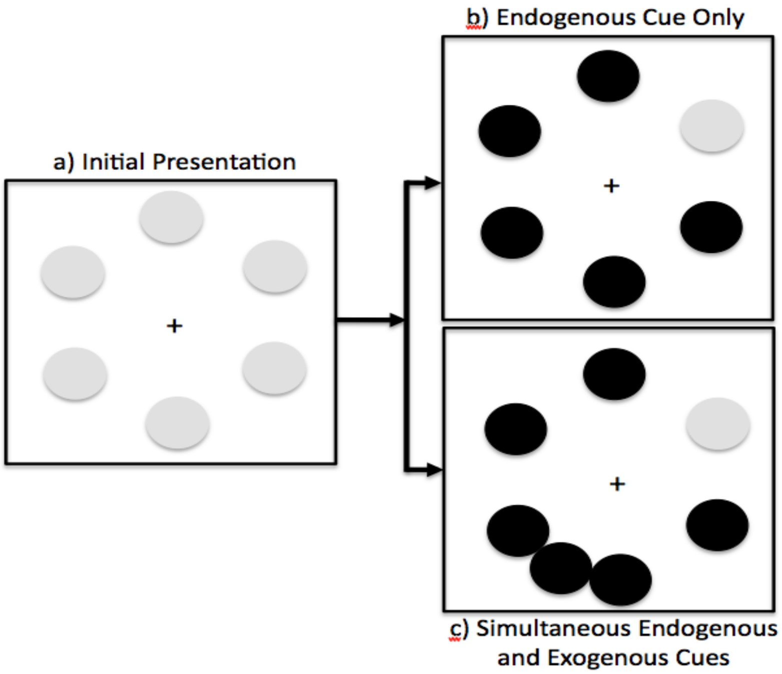


Figure 7. When presented with circles arranged symmetrically around a central fixation (panel a) that subsequently all but one change in color, participants endogenously orient to the unchanging singleton (panel b). However, when a new distractor is added at the same time as the distractor feature change, exogenous attention is simultaneously cued to its onset (panel c).

The results show that performance decreases for target detection, a large proportion of eye movements are made toward the irrelevant distractor despite intention to move toward the target, most of these errors are immediately redirected toward the target, and the observer lacks awareness of ever having oriented towards the distractor. Theeuwes and colleagues (1998) have suggested that endogenous and exogenous shifts are programmed in parallel. In this case, the slower acting endogenous shift inhibits and overwrites the quicker, but erroneous, exogenous shift. This paradigm is distinct from the similar *antisaccade task*, where participants are presented with an exogenous peripheral stimulus that endogenously instructs participants to orient to a target in a different location. The antisaccade peripheral distractor is task-relevant, unlike the task-irrelevant abrupt onset distractor demonstrated here.

3.4.2 Covert

As with endogenous-exogenous interactions in overt attention, there remains debate regarding the extent to which salient peripheral external events can capture attention or be ignored when they are task irrelevant (Klein & Shore, 2000). Seemingly contradictory results have supported that covert attention can be exogenously drawn by certain task-irrelevant, though salient, stimuli (e.g., stimulus onset/offset or presence of a singleton) despite successful endogenous filtering of other stimuli (Klein & Shore, 2000). The SOA for exogenous disruption of endogenous filtering is approximately 150-200 ms. When the distractor precedes the target by more than 200 ms, it does not appear to affect performance. Further, exogenous distractor interference on endogenous target performance may depend on whether the distractor and target share a feature property critical to task performance. This concept is encompassed in the *contingent involuntary orienting hypothesis* (Folk, Remington, & Johnston, 1992). Folk and colleagues (1992) suggest that the mechanism that guides endogenous attention is the instantiation of an appropriate *attentional control setting* (ACS), or a generated rule that determines which signals will result in orienting.

3.5 Space-Based Attention versus Object-Based Attention

Researchers have provided seemingly contradictory evidence as to whether attention is allocated to unparsed regions of space or to perceptual groupings (Driver, 2001; Goldsmith & Yeari, 2003). Despite early debate, there is now a general consensus that space, objects, or both can be selected as relevant sources of information (Egly, Driver, & Rafal, 1994). Specifically, selective attention is a system used to allow preferential access to deeper processing and cognitive algorithms both to particular regions of space and to the objects within that space (Klein & Shore, 2000).

Vision has the highest spatial resolution of all the senses. Spatial attention can be selectively focused on a region of space subtending an area as small as 1° of visual angle (Johnston & Dark, 1986). Attention can be directed to the parafovea and regions that are spatially disparate, but only at the cost of precision, accuracy, and depth of processing. Attention can also be focused on the basis of objectness (i.e., perceptual groupings) rather than spatial location per se (Driver, 2001; Johnston & Dark, 1986). For example, early evidence was presented in the previously mentioned Rock and Gutman (1981) study where participants performed poorly on a memory test of shapes from an unattended stream even though they were presented in spatially overlapping arrays with shapes from an attended visual stream. Similarly, when shown super-imposed video of a hand game and a ballgame and asked to respond to a particular event occurring in only one of the streams, participants were unable to report events in the unattended stream, even when the events were quite strange (Neisser & Becklen, 1975). Distractors that can be perceptually grouped with targets because, for example, they share uniform connectedness within an enclosed region, show greater inference effects than equally distanced distractors that are not perceptually grouped with those same targets (Driver, 2001).

There remains some question as to how the attentional system shifts between or incorporates information from both space-based and object-based attention. There is evidence to support the idea that exogenous cues consistently induce object-based attention while endogenous cues generally induce space-based attention (Egly, et al., 1994). Goldsmith and Yeari (2003) suggest that endogenous and exogenous cues are confounded by the initial focus of attention. These are centrally focused for endogenous versus diffusely spread for exogenous attention. They argue that it is this initial focus, rather than the mode of cuing that determines whether attention will be space-based or object-based. Goldsmith and Yeari posit that for both endogenous and exogenous cuing, circumstances that foster diffuse focus are more likely to engender object-based effects. Both space-based and object-based attention place an emphasis on stimuli that are spatially discrete, thus further consideration is warranted for stimuli that are temporally discrete. Temporally discrete stimuli are dynamic objects and these can be termed ‘events’ (Berger, Martelli, & Pelli, 2003). Event-based attention is more complex and has not received much empirical attention.

3.6 Modal Attention

Although the space-based versus object-based distinction requires further clarification, there remains evidence of yet another form of attention allocation. Selective modality-based attention is the ability to attend to specific sensory modalities. Early belief regarding modality and attention was that input from each sensory modality could be processed independently (Driver & Spence, 2004). Behavioral evidence from dual-tasks, speeded classification tasks, and cuing tasks as well as neurobiological evidence from ERP and fMRI research indicates that the hypothesis that each sensory modality acts as a separate attentional resource is incorrect. Instead, it is now understood that selective attention must be coordinated across the different senses (Driver & Spence, 2004). Behavioral findings demonstrate that there are crossmodal interactions in selective attention, especially concerning space-based attention. Neurobiological findings bolster the behavioral findings by suggesting that crossmodal links involving spatial synergies are present in selective attentional processing (Eimer, 2004). *Multimodal prior entry,* an extension of Titchner’s *law of prior entry* (see Section 3), holds that attending to a specific sensory modality speeds up processing of stimulation arising in that modality (Spence, Shore, & Klein, 2001).

3.6.1 Behavioral Evidence of Modal Attention

Dual-task performance is enhanced when the relevant stream consists of multimodal stimuli originating from the same spatial location. However, performance is degraded when multimodal stimuli originate from disparate spatial locations. Note that the former of these echoes *intersensory facilitation*. The latter, on the other hand, is distinct. When sensory information in one modality is spatially disparate from a second, performance is worse than baseline for either modality presented in isolation. This result, the spatial modulation of dual-task performance, has been demonstrated when multimodal inputs are task relevant (e.g., attending to audio and video of a person speaking) and when one of the sensory inputs is irrelevant (e.g., monitoring for target characters in a random list while shadowing a speech stream). As discussed previously within the context of weak synesthesia, crossmodal distractor interference has been demonstrated in speeded classification tasks. When making speeded judgments in response to stimulation in one modality and stimulation from a second modality is also presented, performance can be affected. Explicitly, performance decreases when the crossmodal distractor is incongruent and increases when congruent (Driver & Spence, 2004). In a variation of the cuing paradigm, it has been demonstrated that when spatial expectancy is cued for one modality, detection performance in a second modality occurring in the cued location is also enhanced, though to an attenuated extent (Driver & Spence, 2004). For example, when cued for a visual target, reaction time for an auditory stimulus appearing in the cued location is significantly decreased, but not to the same degree as reaction time for a visual stimulus. Similar trends have been demonstrated for other modality pairings.

Crossmodal links in exogenous spatial attention have been examined in yet another variant of the cuing paradigm. Driver and Spence (1998) presented a orthogonal-cuing method where peripheral cues exogenously direct attention laterally, followed by target presentation requiring a vertical participant response (e.g., left/right cues prompt up/down responses). The equiprobable lateral cue is nonpredictive and does not induce response bias on the orthogonal elevation judgment (e.g., a bias toward responding 'right' following an exogenous rightward cue). Similarly, this pattern has emerged in endogenous paradigms where central, nonpredictive cues in one modality can serve to push spatial attention in other modalities to a cued region even when such a shift has proven counterproductive (e.g., Ho & Spence, 2005).

3.6.2 Neurobiological Evidence of Modal Attention

Electrophysiological studies have corroborated behavioral findings that there exist crossmodal links activated when currently irrelevant modalities are presented at the location of expected stimulation in a relevant modality (Eimer, 2004). Surprisingly, such unexpected events affect processing in the early sensory-perceptual stages with elevated ERP components in the first 200 ms after stimulus onset. These early components reflect activation within modality-specific (i.e., unimodal) cortical regions thought previously to be unaffected by multisensory integration. ERP components beyond 200 ms diminish, reflecting that crossmodal links do not affect post-perceptual processes and thus cannot simply be the result of strategy.

One thought is that intersensory facilitation might be a reflection of an attentional gradient in hemispace arising from asymmetric hemispheric activation (Kinsbourne, 1993). This simplistic 'hemispheric-activation account' can be tested by manipulating posture such that hands are crossed while the participant makes responses. According to the hemispheric-activation account, in a neutral arm position, spatial congruency effects would arise from a shift of spatial attention toward the side of the body contralateral to cognitive activation (e.g., rightward attentional shift when left hemisphere is activated). This possibility has been tested and refuted by studies that ask participants to cross their hands over the body midline while keeping their gaze fixed straight ahead, such that there is a mixed mapping between the senses (see Figure 8). Such studies have found that in a hands-crossed posture, attentional shift remains toward the relevant external side even though the anatomical side is switched (e.g., rightward attentional shift when left hand crossed to the right of body midline even though left hand corresponds with right hemispheric activation). These findings indicate that spatial coordinates of crossmodal interactions are not fixed by hemispheric projections, but rather receptors from one modality (i.e., hands) can be remapped with receptors from another modality (i.e., eyes) to maintain consistent mapping of external space (Driver & Spence, 1998). Indeed, it is now believed that afferent signals arising from the separate modalities are represented in the superior colliculus by corresponding maps of sensory space that are also in register with motor representations (Giard & Peronnet, 1999).

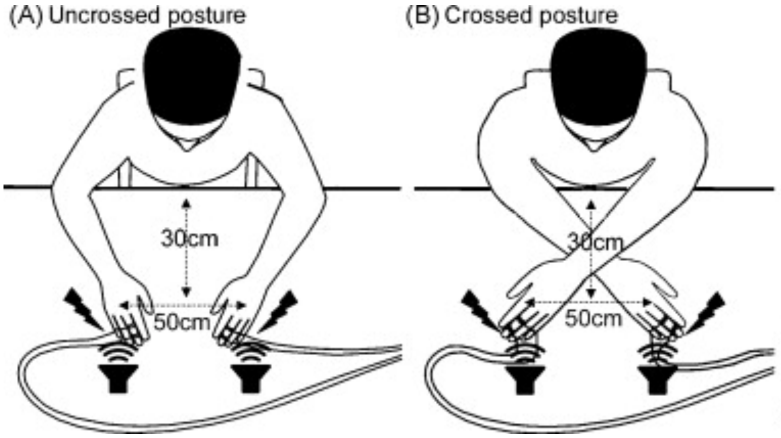


Figure 8. Illustration of a research method comparing neutral (panel A) and hands-crossed (panel B) postures.

ERP studies have also demonstrated attentional shifts of multimodal attention in response to unimodal exogenous spatial cuing and anticipatory multimodal shifts in response to unimodal endogenous cuing. It is suggested that attentional shifts may be hard-wired and mediated by a supramodal attentional control system located in the frontoparietal region (Eimer, 2004). Similarly, PET and fMRI results suggest activation of brain areas traditionally considered to be modality-specific during multimodal processing. Such findings indicate that in addition to the widespread notion of feedforward convergence from modality specific to multimodal brain areas, there may be a supramodal attentional control system responsible for the transfer of spatial information from multimodal areas back to sensory-specific areas (and between unimodal areas themselves) via back-projections (Macaluso & Driver, 2004).

Extinction. Individuals with brain injury also provide neurobiological evidence of modality-based attention. Patients with unilateral brain damage, particularly those with right-hemisphere lesions, often demonstrate a pattern where single stimuli presented to either side are detected with no indication of deficit. However, when stimuli are presented to both sides concurrently, stimulation on the contralesional side is typically missed (Driver & Spence, 1998; Ladavas & Farne, 2004). Extinction is thought to arise from a pathological attentional mechanism that favors input to the unimpaired hemisphere in instances of competition. Extinction has been extended to crossmodal demonstrations using concurrently presented stimuli from two sensory modalities. A tactile stimulus presented to the contralesional (affected) side is missed when visual simultaneous stimulation is delivered near the ipsilesional hand (Ladavas & Farne, 2004). This, again, demonstrates the existence of crossmodal links in attention and perception (Driver & Spence, 1998).

3.6.3 Multisensory Prior Entry

Multisensory prior entry is the idea that attending to one modality speeds up perceptual processing of stimulation delivered in that modality. This is akin to Titchner's assertion that attended stimuli are processed more rapidly than unattended stimuli, or the *law of prior entry* (see section 3). Due to discrepancies in the literature on multisensory prior entry, there remains heated debate regarding its very existence. Interpretation of previous results may have been confounded by response biases, response priming artifacts, spatial attention, or failures to manipulate modal attention successfully.

Spence and colleagues (2001) conducted a series of experiments designed to circumvent many of the confounds inherent in previous research. They utilized a novel methodology in which participant attention was always manipulated in a direction orthogonal to response. For example, participants might be asked to attend to one modality while responding whether the first stimuli appeared on the left or right of midline. Whereas spatial biases might have confounded previous results when auditory stimuli were presented from headphones while visual stimuli were presented from computer screens, the multimodal stimuli used here were always presented from identical spatial locations. Even though potential confounds were addressed, these authors found clear support for multisensory prior entry. When attention was directed to touch, visual stimuli had to lead by a much greater interval in order for the point-of-subjective-simultaneity (PSS) to be achieved (i.e., for participants to judge that the visual and tactile stimuli were presented simultaneously). One unexpected result was that accuracy for detecting which modality was presented first was improved when bimodal stimulation was presented from opposite sides rather than the same location. This result may have occurred due to binding of the audiotactile stimulation when stimuli were presented from a common spatial location.

3.6.4 Crossmodal Links in Attention

Studies of crossmodal links in attention have primarily focused on spatial attention and on links between audition, vision, and touch (Driver & Spence, 1998). Whether cued endogenously or exogenously, touch and vision have been shown to draw each other into spatial register with one another. Endogenous crossmodal shifts are attenuated (approximately 50%) when compared to the corresponding attentional effect for exogenously cued crossmodal shifts. Touch and audition are symmetrically linked such that endogenous or exogenous shifts of attention in one modality induce corresponding attentional shifts in the other. Interestingly, while audition and vision seem to have similar symmetrical links in endogenous attention, the same is not so for exogenous attention. There is an asymmetry in the audiovisual crossmodal link in exogenous spatial attention. The asymmetry is revealed when nonpredictive auditory cues affect visual attention, whereas the spatial distribution of auditory covert attention is unaffected by nonpredictive visual cues (Driver & Spence, 1998). The mechanisms behind such asymmetries are examined in the following section.

4. INTERSENSORY BIAS

Intersensory bias is a potential outcome of multisensory integration where the information from one modality is weighted differently than information from other modalities. In this way, perceptual interpretation of attributes of one modality may be biased towards the value of the more heavily weighted modality (e.g., Soto-Faraco, et al., 2004). Intersensory bias is related to a particular form of modality-based attention, namely exogenous modality-based selective attention. The phenomenon appears to be purely perceptual, and in no way attributable to some strategic or decisional effect (Koppen, Levitan, & Spence, 2009).

4.1 Modality Dominance

4.1.1 Visual Dominance

For a majority of the history of this topic, belief regarding *modality dominance,* or the ability of one modality to 'win out' over another proximal modality, was that vision always dominates multimodal perception (e.g., Shams, Kamitani, & Shimojo, 2002). An archetypical example of visual dominance is demonstrated when participants report that straight surfaces feel curved to the touch when simultaneously viewing the surface through prisms that make straight edges appear curved (Gibson, 1933). During a resurgence several decades after an initial interest in the topic, researchers went on to demonstrate visual dominance in judgments of shape (Rock & Victor, 1964), location (Pick, Warren, & Hay, 1969), orientation (Over, 1966), size (Fishkin, Pishkin, & Stahl, 1975), and memory tasks (Posner, Nissen, & Klein, 1976).

Visual dominance is often induced by the presentation of ambiguous or artificially conflicting stimuli. Perhaps the best-known example of this is the perceptual result generated during the "*Ventriloquist Effect*" (Howard & Templeton, 1966, p. 361). Here, the observer receives conflicting audiovisual input. The auditory modality indicates that the speech is originating from the ventriloquist (veridical), but the visual modality indicates that the dummy produces the speech (illusory). Assuming the ventriloquist is skilled, he or she is able to ‘throw’ his or her voice and the observer compellingly perceives that the dummy is the source of the voice. Empirical research on the ventriloquist effect generally incorporates very basic auditory and visual signals such as point lights and sound bursts. Participants are presented with signals that are moderately separated in space (15-20°) and are asked to perform either discrepancy detection tasks or pointing tasks. In the former, participants are asked to indicate whether sources originate from the ‘same’ or ‘different’ location. In the latter, participants are asked to point in the direction of one of two signals (Colin, et al., 2002). When conditions are right, the ventriloquist effect will occur and participants will underestimate or ignore the spatial separation in favor of the origin of the visual signal. In this case, vision dominates audition.

Crossmodal aftereffects are evident as lingering perceptual modifications occurring subsequent to removal of the conflicting information source. During recalibration, vision has been shown to remain constant while it is the other modality that adapts and later recalibrates to normal (Harris, 1965). Early visual dominance theories generally focused on the reliability and accuracy of visual input (Posner, et al., 1976). Namely, that vision is the dominant sensory modality due to its superior perceptual reliability and accuracy compared with other sensory modalities. Posner et al. further elaborated that visual dominance stems from the characteristic that visual stimuli are inherently less alerting than non-visual stimuli. They suggest that observers must allocate more attention to visual information to compensate for its low alerting capabilities. Thus, fewer attentional resources are left available for processing of non-visual input and these other modalities contribute less to overall perception. However, even Posner et al. noted that visual dominance was not impervious to modification, but rather depended on a number of factors such as relations between stimuli and observer attention allocation. These authors put forth a set of propositions regarding visual dominance that, taken together, imply that while under most circumstances vision will dominate, under some it will not (Posner, et al., 1976).

4.1.2 Auditory Dominance

As reviewed above, there exist many instances of visual dominance where conflicting information is modified by a heavier weighting of visual input. Over time and especially more recently, however, a growing number of reports have demonstrated cases where vision is modified by audition (Shams, et al., 2002). At a point it became clear that any sweeping generalizations claiming vision to be the most important or dominant modality were incorrect (Welch & Warren, 1980). The majority of auditory dominance findings involve modification of *temporal perception* of the visual stimulus. Irrelevant auditory stimuli can capture visual attention and shift the apparent timing of visual events (Binda, Morrone, & Burr, 2009). As a direct analogy to the "ventriloquist effect", the ability of auditory signals to alter the perceived timing of visual signals is often termed *temporal ventriloquism* (Morein-Samir, Soto-Faraco, & Kingstone, 2003). Audition has been found to be weighted more heavily in audiovisual perception of frequency (Gebhard & Mowbray, 1959; Myers, Cotton, & Hilp, 1981), duration (Ben-Artzi & Marks, 1995; Burr, Banks, & Morrone, 2009; Walker & Scott, 1981), temporal resolution (Scheier, Nijwahan, & Shimojo, 1999), and temporal order judgment (Boenke, et al., 2009).

Frequency. Analogous to cases of visual dominance, presenting the observer with ambiguous or conflicting multimodal stimuli can induce auditory dominance. One of the earliest reported cases yielding empirical evidence for auditory dominance came from studies comparing auditory flutter with visual flicker (Gebhard & Mowbray, 1959). Previously, it had been established that observers were accurate in detecting small differences of frequency for both flicker and flutter. The smallest discernable frequency change is dependent on the reference frequency. Over a small range where the two modalities can be compared (about 1-45 Hz), the auditory and visual frequency relative difference limens (ΔF/F) are comparable, varying from approximately .005 to .025. This variation reflects only 0 to 0.5 standard deviations from correctly matching the actual rate (see Figure 9).

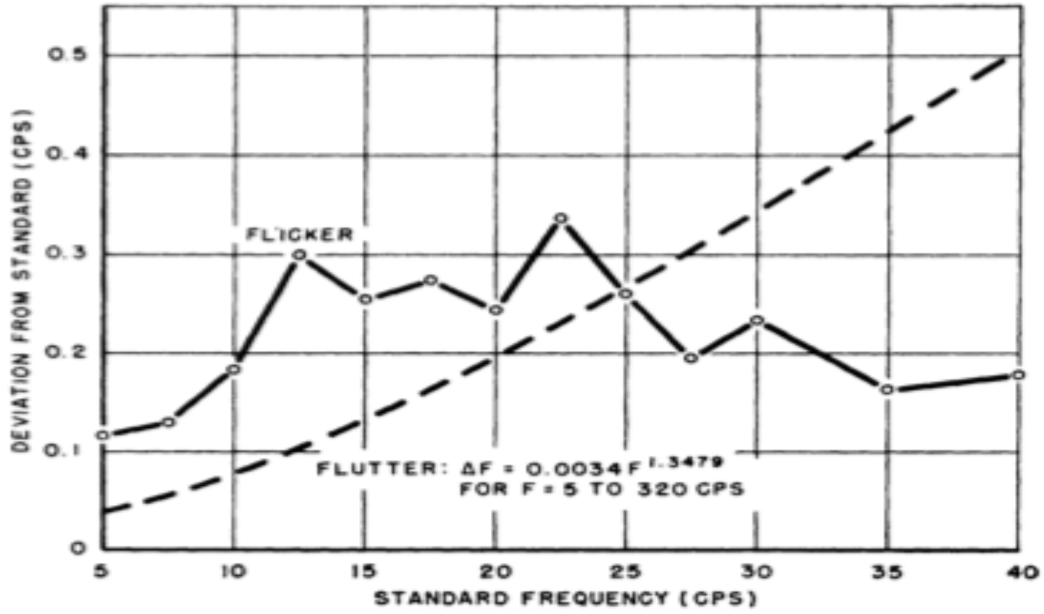


Figure 9. Average observer frequency judgment deviation from actual frequency of auditory flutter and visual flicker.

Given these previous findings with intramodal frequency comparison, Gebhard and Mowbray (1959) hypothesized that an intermodal frequency matching of auditory flutter to visual flicker, or vice versa, would be fairly simple to do and would result in a trend line lying between the two intramodal result patterns. Instead, these authors found that task completion was quite difficult with accuracy being an order of magnitude worse than intramodal matching.

Further, they discovered an effect that they coined *auditory driving.* That is, changes in the physical rate of flutter induced changes in the perceived rate of flicker. Matches are more accurate (i.e., closer to veridical frequency) when flicker is matched to flutter. Auditory flutter seems to drive the perceived frequency of visual flicker. Auditory driving is also evidenced by subjective reports and the effect seems to increase at progressively higher frequencies (Myers, et al., 1981).

Duration. Auditory dominance has also been demonstrated in studies of perceived duration. Duration of a one-second tone is consistently judged to be longer than a separately presented one-second light, given that the intensities of each are moderately matched (Walker & Scott, 1981) . Because the perceived duration of a tone is longer than that of a light of equal veridical length, audio-visual perceptual conflicts arise when a light and tone are presented simultaneously for an interval of about a second. Walker and Scott (1981) demonstrated a form of auditory capture where the perceived duration of an audiovisual stimulus was similar to that of a tone alone, and significantly different from a light alone. Further, they demonstrated that perceived duration of a gap in an otherwise continuous tone is longer than that of a gap in an otherwise continuous light. Again, the perceived duration of a gap in an otherwise continuous bimodal stimulus was perceived similarly to the gap in tone alone and significantly different from a gap in light alone. A trend emerged where auditory dominance over vision increased at a 1500 ms interval (97% bias) versus a 1000 ms interval (67% bias). The magnitudes demonstrated in this auditory bias effect are comparable to corresponding time courses of visual bias demonstrated in previous studies (e.g.,Welch & Warren, 1980).

Temporal resolution. Scheier, Nijhawan, and Shimojo (1999) demonstrated that auditory capture can act to enhance or degrade visual temporal resolution. Temporal resolution is often expressed in terms of the just noticeable difference (JND) between the presentation of two stimuli, or the smallest temporal gap where two stimuli do not become perceptually fused (Shimojo & Shams, 2001). Auditory modulation of visual JNDs were obtained by presenting irrelevant sounds temporally adjacent to the visual target stimuli. Participant performance in determining the order of two visual stimuli was enhanced by presenting auditory stimuli before and after a pair of visual targets (i.e., A-V-V-A). Conversely, temporal resolution was degraded when auditory distractors were presented between the pair of visual targets (i.e., V-A-A-V) (as cited in Shimojo & Shams, 2001). These findings have proven robust to spatial variation (Vroomen & Keetels, 2006). Scheier and colleagues suggest that auditory capture is the mechanism responsible for the modulation of task accuracy. Because auditory temporal resolution was not explicitly tested, it is possible that a more accurate explanation would consider the effect of visual driving as well.

Temporal-order judgment. Temporal-order judgment (TOJ) tasks are those that require participants to indicate which of multiple temporally asynchronous stimuli was presented first. The TOJ task can be utilized to determine the point of subjective simultaneity (PSS), or the relative timing required for presentation of multimodal stimuli to be perceived as occurring concurrently. PSS can be determined using the method of constant stimuli to calculate the point at which the participant is 50% likely to indicate that one stimulus preceded the other. The temporal integration involved in multimodal PSS is affected by both physical transmission and sensory processing time differences between the modalities (e.g., van Eijk, Kohlrausch, Juola, & van de Par, 2009). Interestingly in the case of audiovisual stimuli, although auditory simple RT is quicker than visual simple RT (e.g., Stone, et al., 2001; Walker & Scott, 1981), the physical transmission time for auditory stimuli is slower than that for visual stimuli (i.e., the speed of sound is slower than the speed of light). Similarly, some have reported that PSS is reached only when visual stimuli are presented prior to auditory stimuli presentation (e.g., Zampini, Guest, Shore, & Spence, 2005). Part of the reason for this seeming contradiction between differences in modality transmission time and behavioral data is that the transduction of light on the retina is slower than the transduction of sound on the basilar membrane (Massaro, Cohen, & Smeele, 1996). Still, this cannot fully make-up for the vast difference in physical transmission times (e.g., at a distance of 10 m light will arrive 30 ms before the sound). The remaining difference is likely due to attentional mechanisms such as those involved in Posner's assertion that auditory stimuli have a greater 'alerting' effect.

Contradictory results have also been put forth indicating that visual stimulation must be presented prior to auditory stimuli in order to achieve PSS (though the opposite is more often the case). The values determined for PSS vary greatly because they depend on a number of factors including task demands, cognitive factors, stimulus intensity, stimulus duration, and individual differences. The more intense a stimulus, the more likely it is to be judged as having preceded a less intense stimulus up to a certain SOA. Stimulus intensity can be controlled for using a method to adjust both stimuli to matched intensity thresholds (e.g., 75% threshold for both auditory and visual stimuli). Despite such experimental controls, between-subjects variability has been broadly reported. Boenke, Deliano, and Ohl (2009) demonstrated that even though notable individual differences in PSS values are present, values across participants converge towards a common attractor value as duration of bimodal stimulation increases. These authors suggest that a multisensory integration mechanism stabilizes perceived asynchronies with increasing stimulus duration.

During typical TOJ tasks, participants may be asked either to indicate which modality occurred first (onset or offset) or to indicate the location of the modality which occurred first (Boenke, et al., 2009). Typically the latter variety is used to avoid potential modality bias (i.e., tendency to select one modality over the other). However, this variation may incidentally bias perception nonetheless by placing more emphasis on spatial aspects of the task. By shifting task demands away from a temporal towards a spatial emphasis, visual bias may occur (see section 5.3 for a more thorough discussion).

4.2 Crossmodal Masking

The above examples reflect cases where visual or auditory signals are somewhat ignored in favor of the alternate source. For example, duration judgment was 67% biased by the auditory stimulus lasting 1000 ms in Walker and Scott's (1981) study. Though experimental discordances reported are typically less than 100% bias (Bertelson & Radeau, 1981), there are instances where conflicting modality adds little or nothing to the overall perception (Fendrich & Corballis, 2001). There are cases of seemingly complete modality dominance where participants wholly disregard information from the non-dominant modality. In these cases, participants respond to bimodal stimuli no differently than they do to the dominant modality in isolation. Further, observers often report that they are unaware that any other signals have occurred. One such case can occur when participants are presented with intensity-matched auditory, visual, and audiovisual stimuli and asked to determine the occurrence of each modality independently. For instance, a left-button press might indicate the detection of an auditory cue and a right-button press a visual cue. During bimodal presentations, participants would be asked to press both buttons. Though participants are equally accurate in detecting either stimulus in isolation, they respond to audiovisual stimuli as though only the visual input were present and are unaware that any non-visual signals have occurred (Koppen, et al., 2009; Posner, et al., 1976). This phenomenon is termed the *Colavita*

Visual Dominace Effect (Colavita, 1974). In studying such cases, one may wonder whether a mechanism similar to those in forward or backward masking might be at play.

Masking involves the presentation of a target stimulus and a mask stimulus that suppresses perception of the target stimulus (Raab, 1963). This can occur whenever stimuli with specific parameters occur close together in time or space. In forward masking, the mask stimulus is presented before target onset. Conversely, in backward masking, the target stimulus is presented before mask onset. *Visual persistence* is one factor contributing to visual forward masking where neural activity cessation following visual stimulation is sluggish (Schiffman, 2001a). The percept may linger beyond veridical mask offset and interfere with perception of target onset. A more general masking hypothesis rests on the idea that perception involves two distinct processing channels (Breitmeyer & Ganz, 1976). Structural information is processed along sustained channels, whereas spatial location information is processed along transient channels. Within this model, forward masking arises from intrachannel lateral inhibition via center-surround antagonism of sustained cell receptive fields. Backward masking arises from interchannel inhibition where transient cells laterally inhibit the activity of sustained cells. Masking is therefore explained in terms of a multichannel model where sensory integration of sustained channel information occurs at different levels of visual processing (Breitmeyer & Ganz, 1976). Given what we are now beginning to understand regarding multimodal integration, it is reasonable to question whether information in multiple modalities might affect each other in a manner similar to traditional, unimodal masking.

Koppen, Levitan, and Spence (2009) designed an investigation of the Colavita Visual Dominance Effect in order to determine whether this phenomenon was truly perceptual or merely the result of post-perceptual strategy. They accomplished this by first determining 75%

detection thresholds for each participant using a LED and a 4,000 Hz pure tone, both presented from a distance of 60 cm. Participants were then presented with the standard Colavita signal detection task where they were instructed to press separate visual and auditory response keys. As expected, the Coalvita visual dominance effect manifested. The results showed significantly more visual-only than auditory-only responses. Using signal detection theory (SDT), Koppen and colleagues determined the number of hits, misses, false alarms, and correct rejections made by participants for each sensory modality. They were then able to calculate the sensitivity (d') and the response criterion (c) adopted by participants during the task. Results indicated that sensitivity to auditory stimuli decreased in the presence of concurrent visual stimuli. Further, participants were significantly more liberal in responding to bimodal than a unimodal targets (Koppen, et al., 2009), likely due to an intersensory facilitation effect. Though the stimuli were presented concurrently in the bimodal condition, it is known that the timing of auditory versus visual stimuli varies according to several factors (see section 5.1.2). The perceptual suppression of the auditory stimulus in the presence of the visual stimulus might reflect a form of crossmodal masking where the visual stimulus initiates lateral inhibition of auditory information in the sustained sensory channel.

4.3 Task-Context Dependent Modality Dominance

Welch and Warren (1980) provided what is arguably the most comprehensive view of intersensory bias to date. The “New View” they presented integrates what were the three predominant explanations at that point (Welch & Warren, 1980, p.661). Among these explanations of intersensory bias, both the *Modality Precision Hypothesis* and the *Modality Appropriateness Hypothesis* point toward the idea that audition will dominate temporal tasks while vision will dominate spatial tasks.

Modality precision hypothesis. Under this hypothesis, intersensory bias can be explained by the assumption that when two sensory modalities provide discrepant information, the resulting perceptual outcome will favor the modality with the more precise resolution within the task context. While audition has a higher temporal resolution, vision has a higher spatial resolution (Welch & Warren, 1980). The superior spatial resolution of vision is most pronounced for central, foveal vision and becomes increasingly degraded in the periphery. Further, visual localization is limited to observer field of view. External events occurring outside the line-of-sight can be localized via audition, assuming there are audible signals given off by that event.

Modality appropriateness hypothesis. In the Modality Appropriateness Hypothesis, the assumption is that for simple tasks the different modalities are not differentially effective, but in more complex tasks modalities show superiority due to differential information processing characteristics (Welch & Warren, 1980). For instance, auditory spatial localization is certainly feasible; it is something we are all able to do and that individuals with little-to-no vision rely on (Després, Candas, & Dufour, 2005). However, for individuals with normal functioning in all sensory modalities, vision is the most appropriately designed for spatial judgments. The more suited modality will dominate others in a given task.

Directed-attention hypothesis. The third explanation of intersensory bias integrated into Welch and Warren's view rests on the idea that the bias can be explained by differential allocation of attention to the conflicting modality information. That is, the more-attended modality has a dominating role in the overall perception of multimodal input. There is an important distinction here between what Welch and Warren term "salience", or the ability of a stimulus to elicit attention, and observer bias towards directing attention towards a particular modality. As used here, salience is reminiscent of the "alerting effect" discussed by Posner and colleagues, where salience is measured by the degree to which stimuli serve to alert an organism of their presence (Posner, et al., 1976, p. 160). A more modern interpretation of this same distinction is discussed in the previous section on endogenously (bias) and exogenously (salience) controlled attention.

Optimal cue combination model. Optimal cue combination draws on many of the same concepts pulled together in Welch and Warren's (1980) "New View", along with more modern concepts from Bayesian modeling (Ernst & Banks, 2002). Specifically, this theory posits that for any multimodal task, the integrated perception will result from an optimal weighting of the inputs according to their respective signal-to-noise ratios (Berger, et al., 2003; Ernst & Banks, 2002). This idea is reinforced by findings that when visual input is degraded (e.g., by blurring or by presentation during saccades), auditory inputs become more heavily weighted for perception of visual location (Binda, et al., 2009). Similarly, when auditory signals are degraded, visual frequency can change perception of auditory frequency (i.e., reverse of auditory driving) (Wada, Kitagawa, & Noguchi, 2003). Though there are alternative flexible weighting theories of intersensory bias, optimal cue combination is most successful in explaining the role of auditory stimuli in disambiguating between multiple possible interpretations of ambiguous visual stimuli (Berger, et al., 2003). The Bayesian model also fits some counter-intuitive audiovisual integration behavioral data such as the finding that visual signals (i.e., lip-reading) are most effective in speech perception during cases of moderate, as opposed to high, auditory noise (Ma, Zhou, Ross, Foxe, & Parra, 2009).

4.4 Auditory Dominance Outside the Realm of Temporal Tasks

Though most findings in multimodal integration demonstrate that audition dominates in temporal tasks whereas vision dominates in spatial tasks, some argue that there are exceptions where audition dominates perception in other types of tasks (e.g., Shimojo & Shams, 2001).

4.4.1 Motion Perception

Motion perception involves an important spatial aspect. Therefore, cases where audition alters vision in motion perception have been posited as evidence that contradicts the modality appropriateness hypothesis (i.e., vision should bias audition in a spatial task, but not vice versa). Sekuler, Sekuler, and Lau (1997) presented one such example where auditory information altered perception in a motion perception task with two equally feasible visual interpretations. A visual display depicted two identical circles moving towards one another, coinciding, and then moving apart. This could be interpreted as two objects passing in the middle and continuing along an uninterrupted path, or as a second equally feasible event of two objects colliding in the middle before bouncing back in opposite directions.

When this visual display is presented in isolation, interpretation generally favors passing. There is an increase towards the colliding interpretation when a pause of increasing duration is included at the point of coincidence. The inclusion of a transient sound stimulus (i.e., an audible click) at the point of, prior to, or after coincidence facilitates the collision interpretation significantly. A sound presented just prior to coincidence has almost as much of an enhancing effect on the bouncing interpretation as a sound presented at coincidence. But, a sound presented just after coincidence, while still promoting the bouncing interpretation, has a significantly smaller effect (Sekuler, et al., 1997).

4.4.2 Illusory Flash

Shams, Kamitani, and Shimojo (2000) presented findings in which they established the discovery of a visual illusion induced by sound. Unlike previous demonstrations of auditory influence outside temporal perception such as the alteration in motion perception presented by Sekuler et al. (1997), the illusory flash is not limited to a case of ambiguity or visual degradation. Shams and colleagues found that when presenting participants with a single visual flash, the perceived number of flashes depends on the number of simultaneously presented auditory beeps. For example, a single flash paired with a single beep is perceived as a single flash, where as a single flash paired with a double beep is perceived as a double flash. The illusory flash persists within a 100 ms SOA timeframe, consistent with reports of multimodal integration processes. The perceptual illusion has been found to be robust to manipulations of stimulus variables and to observer discrepancy awareness. That is, even when participants are aware that the visual disc is only flashed one time, they persist in observing multiple flashes in the presence of multiple beeps (see Figure 10). Conversely, altering the number of flashes did not affect the accurate perception of the number of presented beeps (Shams, et al., 2002; Shams, et al., 2004). The visually evoked responses recorded during illusory flash presentation were qualitatively similar to those during veridical flash. Additionally, auditory presentation modulated activity in visual cortical areas as early as V1 (Shams, Kamitani, Thompson, & Shimojo, 2001). Taken together, these findings strongly suggest that crossmodally induced effects mimic unimodal response in the cognitive mechanisms and neural pathways utilized. These are true perceptual effects and not the result of response bias or some similar artifact.

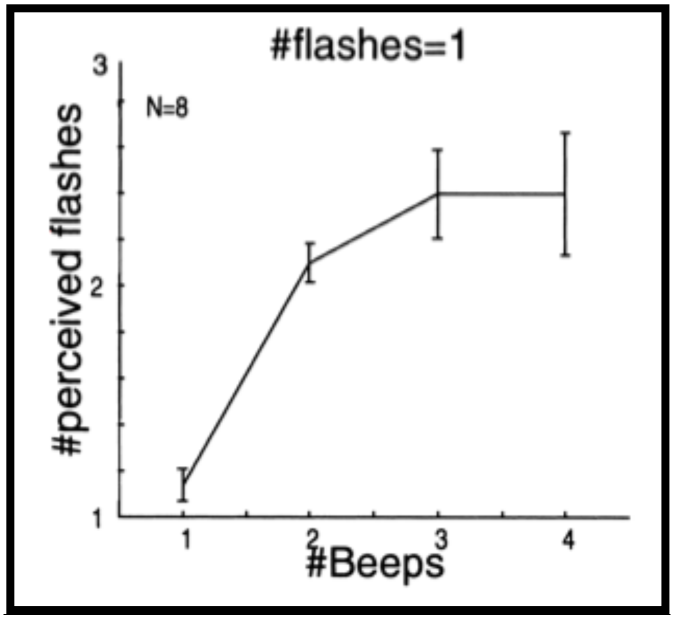


Figure 10. Illusory flash results presented in Shams, Kamitani, and Shimojo (2000).

4.5 Multisensory Fusion: Intersensory Modification without One-way Domination

The concept of illusory *fusion* involves the percept that discordant inputs belong to the same event (Bertelson & Aschersleben, 2003). In terms of neural pooling, illusory fusion is an instance of an invalid co-occurrence, or a case where sensory information arising from two separate events is mistakenly taken as redundancy. Fusion is a form of cross-sensory binding that generally results in a percept that is distinct from either input (Bertelson & de Gelder, 2004). An illustrative example is the McGurk effect. During the McGurk illusion, the observer fuses information about manner of articulation (i.e., stop vs. fricative) and voicing (i.e., voiced vs. voiceless) from the auditory signal with information about the place of articulation (i.e., alveolar vs. bilabial) from the visual signal (Brown, 2005; Massaro, 1999). The resulting percept, ‘da’ is an intermediary interpretation based on a weighting of both auditory and visual input.

Often it is the case that in addition to a moderating effect of one input on a second, the effect of the second on the first can also be observed. Though one could argue that this is always true except in cases of total modality dominance (e.g., crossmodal masking), this possibility has not always been explicitly tested. In a study akin to the auditory driving line of research, Fendrich and Corballis (2001) demonstrated an audiovisual cross-capture effect. First they showed that a sound burst (i.e., an auditory click) has the ability to modify the perceived timing of a visual event. Specifically, the perceived position of a rotating visual marker at the time of a visual flash was modified. This study was designed to disentangle the influence of auditory capture from that of visual capture. These are confounded in traditional auditory driving studies.

Participants were asked to indicate the location of a visual marker in one of twelve positions on a clock face at the time of visual flash presentation (see Figure 11). The visual marker was depicted as an open circle that rotated clockwise around the region within a larger circle. The visual flash was produced by the disappearance of the inner circle and outer ring. An audible click was presented before, on, or after the visual flash. For example, if the visual marker was located at the 12 o'clock position at the time of visual flash presentation, the audible click could occur at the 10, 11, 12, 1, or 2 o'clock position. Findings indicated that when the audible click occurred before the visual flash, reported location of the marker at the time of visual flash was shifted earlier. When the audible click occurred after the visual flash, the reported location of the visual marker was shifted later. In other words, auditory capture occurred. When participants were instead asked to indicate the location of the target marker at the time of the click, similar but much smaller visual capture effects were observed. Fendrich and Corballis (2001) suggest a temporal cross-capture of vision and audition that they term, intersensory temporal locking (ITL). A similar capture asymmetry has been reported in other studies of audiovisual interaction (Shams, et al., 2000; Wada, et al., 2003; Welch & Warren, 1980).

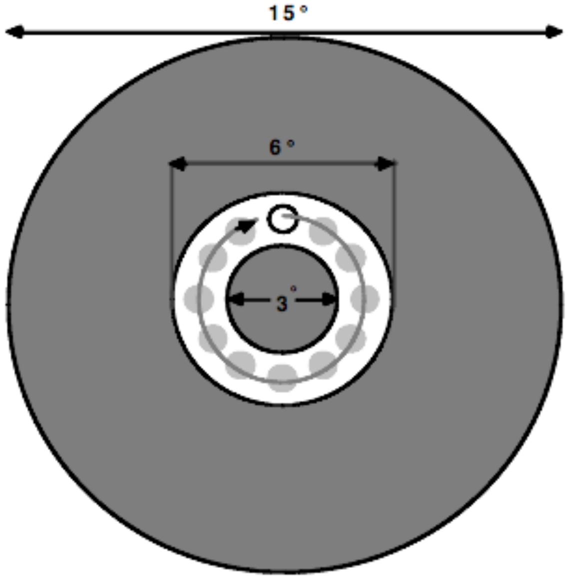


Figure 11. Display used in Fendrich and Corballis's 2001 temporal cross-capture study.

5. CROSSMODAL BINDING

5.1.1 Exceptions to the Rule?

Many researchers have investigated the effect of visual input on auditory perception, but the mechanisms underlying this effect are not well studied or understood (Marks, 2004). Though much of the evidence fits within the framework presented by Welch and Warren (1980), the respective findings of Sekuler et al. (1997) and of Shams et al. (2000) are presented as evidence contradicting this framework (see section 4.4). Whereas Welch and Warren outline a number of reasons why vision will dominate spatial tasks and audition will dominate temporal tasks, these two studies are offered as cases where auditory input is dominating perception in a "spatial task". However, it is argued here that neither of these respective experimental tasks would have meaning if temporal factors were absent (i.e., if static displays were presented).

Sekuler and colleagues' task involves motion perception. The perception of motion itself requires integration of spatial change as a function of time, $s = f(t)$ (Johansson, 1976). In fact, some claim that it is the temporal relation only, as opposed to spatiotemporal integration, that determines the audiovisual interaction observed during motion perception (Allen & Kolers, 1981). On the other hand, there are also many examples of motion perception where vision dominates discrepant auditory input. For example, auditory and visual illusory induced motion presented synchronously, but in opposite directions, will be interpreted in favor of the visual illusory motion only (Soto-Faraco, et al., 2004).

Presenting temporally sequential and spatially adjacent visual or auditory stimuli can induce illusory motion. Typically in the visual case, a point-light stimulus is presented to the left or right of central fixation followed by a second point-light stimulus presented to the opposite side of central fixation. The illusory effect is that of a light stimulus originating at the location of the first stimulus that then travels to and terminates at the location of the second stimulus. When presented unimodally, participants have no difficulty in determining the direction of induced motion in either the visual or auditory modality. However, when discrepant audiovisual signals are presented, observers are unable to correctly identify the direction of the auditory stream while perception of the visual stream is unaffected by the incongruent direction of auditory motion (Soto-Faraco, et al., 2004). This is a phenomenon called crossmodal dynamic capture (Sanabria, Soto-Faraco, Chan, & Spence, 2004). Thus, solely within motion perception tasks, there exist examples of both auditory dominance (Sekuler, et al., 1997) and visual dominance (Soto-Faraco, et al., 2004).

The results of Sekuler and colleagues (1997) can potentially be explained by a flexible weighting theory of intersensory bias (e.g., the optimal cue combination model). Such models suggest that the auditory input is more heavily weighted in this case due to the ambiguity in the visual signal (i.e., higher degree of variance). Shams and colleagues (2000) posit that their 'illusory flash' is resilient to such explanations because the visual stimulus is not ambiguous or degraded in any way. Here again, however, the task is not purely spatial as it involves integration of temporal aspects (i.e., temporal patterning). Temporal patterning is put forth as an amodal property that serves to bind multimodal stimuli by the unity assumption (e.g., Brown, 2005; Colin, et al., 2002; Vatakis & Spence, 2006). It has been demonstrated that auditory manipulations that increase contrast between individual beeps also increase the number of illusory flashes perceived (Wilkie, Stevens, & Dean, 2008). In the case of the illusory flash, perhaps it is other commonalities between the inputs that induce artificial binding (i.e., fusion).

5.1.2 Away from Dichotomy, Toward a Continuum

The studies highlighted in the previous section exemplify an unresolved pattern in the literature that might be attributable to an overemphasis on the dichotomous classification of tasks as either spatial or temporal. Many tasks are more accurately described as spatiotemporal with some falling toward the spatial end, and others falling toward the temporal end of a ‘spatiotemporal continuum’. All other factors being equal, perception during tasks that fall near the spatial end of the continuum are dominated by visual input. Correspondingly, perception during tasks that fall near the temporal end of the continuum are dominated by auditory input. The closer to the midrange of the continuum a task falls, the more equally balanced the audiovisual input will be in the final, integrated perception. Only tasks falling at either extreme will induce complete modality dominance, or crossmodal masking. If the above assumptions hold true, then it should be possible to modulate the respective weighting of auditory and visual information by manipulating task emphasis from spatial to temporal or vice versa. This possibility is elaborated in the program of research conducted for this dissertation, but before delving into these studies, the bounds of intersensory binding are considered.

5.2 Spatiotemporal Bounds

As discussed previously, time and space are considered to be the most important amodal factors contributing to the assumption of unity for multimodal binding (e.g., Zampini, et al., 2005). Indeed, one would be hard pressed to imagine cases where external events did not occur at a particular point in time and space. This does not imply, however, that the unity assumption is restricted only to synchronous or spatially overlapping multimodal signals. Instead, multimodal binding can occur within somewhat flexible temporal and spatial windows. This inherent flexibility is in fact necessary for audiovisual integration given the temporal discrepancies in physical transmission, sensory transduction, and neural processing between auditory and visual signals (Massaro, et al., 1996). These temporal discrepancies can lead to spatial discrepancies (i.e., as is the case for crossmodal dynamic capture). Congruous with behavioral ‘temporal window’ investigation results, neural evidence suggests that slightly discrepant information can be integrated by a temporal extension of activation beyond the point of initial stimulation (Massaro, et al., 1996).

There are, however, limitations to audiovisual spatial and temporal windows. Outside these boundaries lies an inhibitory region where perceptual processes may act to suppress neural response. This inhibition degrades the signal in order to disambiguate separate external events (Stein, et al., 2004). The bounds of the spatial and temporal windows required for audiovisual binding do vary with changing task-demands. This is demonstrated by the degree to which spatial or temporal manipulations disrupt audiovisual binding within different task contexts. For example, the McGurk effect is only temporally flexible up to 180 ms but is not degraded by spatial separation (Colin, Radeau, Delenre, & Morais, 2001; Munhall, Gribble, Sacco, & Ward, 1996). The Ventriloquist effect, on the other hand, has little temporal flexibility (though strict synchrony is not required) and degrades with distance (e.g., effect occurs at 20 but not at 60 degrees of separation) (Colin, et al., 2002). Because speech is a predominantly auditory task, temporal desynchonization of audiovisual components are far more devastating in the perception of speech than is spatial misalignment (Zampini, et al., 2005). The position of a given task on the spatiotemporal continuum is operationally defined by its susceptibility to manipulations of spatial and temporal coincidence. Tasks that are impervious, or at least less susceptible, to temporal manipulations depend more toward spatial aspects and will lie further on the spatial end of the continuum. Conversely, tasks that are less susceptible to spatial manipulations depend more on temporal aspects and will lie further toward the temporal end of the continuum.

5.3 Additional Amodal Commonalities

While commonality in space and time are regarded as the most powerful amodal properties serving to bind signals together, this does not mean that additional amodal properties should be disregarded. Indeed, several amodal properties (e.g., number, shape, size) have been investigated and found to provide an important role in intersensory binding. Though it has not previously been treated as such, it is proposed here that synesthetic correspondence itself can act as an amodal binding property. Just as commonality in time or space does not always result in a bound percept, not all signals sharing a synesthetic commonality will necessarily be bound. For clarity, two or more signals that have been bound by way of synesthetic congruency will be referred to here as having been *SynCed*. Under this assertion, integral variables (i.e., signals whose orthogonal variation affects judgment on another “irrelevant” signal) are those that have been SynCed. Separable variables, on the other hand, are those that have not been SynCed. In this way, it is SynCing (synesthetic congruence facilitated binding) that renders some variables integral while others remain separable. Behavioral consequences such as Garner Interference arise because one cannot truly “orthogonally vary” part of a SynCed whole. Further, just as the spatiotemporal unity assumption requirements are somewhat flexible, the likelihood that signals sharing a synesthetic commonality will be perceived as a whole is flexible. Whereas before the binding windows existed in time and space, here the flexibility stems from the degree of synesthetic dimensional overlap. The greater the dimensional overlap, the stronger the binding and the more likely the observer is to perceive a unified signal. To test the prediction that synesthetic dimensional overlap can moderate the strength of audiovisual binding, the program of research presented here is designed to assess the strength of the influence audiovisual pairs have on each other as synesthetic correspondence is modulated.

6. PROGRAM OF RESEARCH

This series of studies was designed to investigate the role of synesthetic correspondence on the phenomenon of crossmodal binding. Whereas other amodal properties have been considered (e.g., direction, number, shape, size), the role of synesthetic correspondence, and specifically the degree of dimensional overlap, has largely been neglected. The following hypotheses were tested:

> H1: Synesthetic correspondence is an amodal property that serves to bind intersensory signals and manipulating this correspondence between pairs of audiovisual signals will affect performance on a TOJ task.
>
> H2: Manipulating emphasis during a TOJ task from spatial to temporal aspects will strengthen the influence of task-irrelevant auditory signals.
>
> H3: The degree of dimensional overlap between audiovisual pairs will moderate the effect of synesthetic correspondence on the TOJ task.
>
> H4: Synesthetic correspondence is a confound that has affected empirical investigations of multimodal perception.

Studies 1-3 in this series were intended to demonstrate that synesthetic correspondence affects performance during a task that has previously been shown to elicit temporal ventriloquism, a form of auditory capture. Once having demonstrated that synesthetic correspondence does in fact facilitate crossmodal binding, the final study in this series, Study 4, was intended to illustrate that synesthetic correspondence presents a confound that perception researchers do not, but should, consider when designing empirical studies. The studies included in this program of research are as follows:

Study 1A: Manipulating Synesthetic Correspondence During Temporal Order Judgment

Study 1B: A Shift from Spatial towards Temporal Emphasis

Study 2: Manipulating Synesthetic Dimensional Overlap

Study 3: Investigating the Temporal Ventriloquist Effect

Study 4: A Meaningful Difference in Empirical Outcomes

6.1 Studies 1-3: Effects of Synesthetic Correspondence on Crossmodal Binding

Previous researchers have investigated auditory capture utilizing the temporal ventriloquist effect (e.g., Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Such research has demonstrated that presenting task-irrelevant auditory signals before and after a pair of visual target signals improves TOJs by perceptually pulling the visual signals further apart temporally (i.e., further in perceptual time, not veridical time). For example, when given the task of identifying the position (e.g., higher or lower) of the first of two asynchronous visual stimuli, participants perform better when a task-irrelevant pair of auditory stimuli are presented before and after the pair of target visual stimuli (i.e., A1-V1-V2-A2) than when the same pair of irrelevant auditory stimuli are presented between the pair of visual stimuli (i.e., V1-A1-A2-V2). Note that an improvement in TOJ performance corresponds with a shorter visual point of subjective simultaneity (PSS) value. As illustrated in Figure 12, Studies 1-3 extended this method by including a manipulation of the synesthetic correspondence between the audiovisual pairs (i.e., A1-V1 and A2-V2).

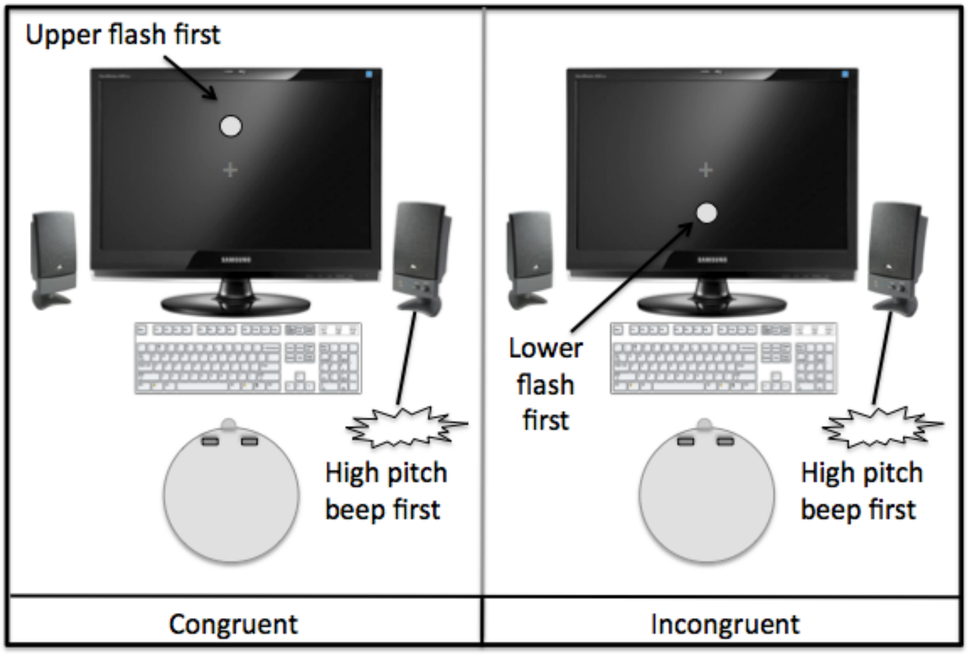


Figure 12. Illustration of the approximate experimental set-up for Studies 1-3. The left panel depicts a Congruent trial while the right panel depicts an Incongruent trial.

Previous research gives some indication of what was expected for baseline visual PSS, baseline AV PSS, and the ideal temporal distance for temporal ventriloquism to occur. The smallest visual PSS value under optimal conditions has been reported at approximately 20ms, where optimal conditions include highly trained participants with maximally visible stimuli (Grossberg & Grundewald, 1997). The current research was intentionally designed to use sub-optimal conditions (e.g., intensity level at 75% threshold). Therefore, baseline visual PSS was expected be longer than 20ms. Published results for baseline AV PSS are not as straightforward. As was discussed in section 4.1, some have reported that in order for an audiovisual pairing to be perceived as synchronous, the auditory stimulus must be presented slightly before the visual stimulus, whereas the majority have reported that the visual stimulus must be presented first (e.g., Zampini, et al., 2005). Boenke, et al. (2009) published average AV PSS values in their "high intensity condition" similar to those reported for visual PSS in optimal conditions (i.e., ~20 ms) with visual leading. However in their "low intensity condition" average AV PSS values were closer to 50 ms with visual leading. Boenke and colleagues show that while a number of factors greatly affect AV PSS (e.g., task demands, stimulus duration), individual differences play a large role in this value. Similarly, Stone and colleagues (2001) reported a range of individual differences spanning from -21ms (auditory leading) to +150 ms (visual leading) with a mean AV PSS of approximately +50 ms across their participants. It was expected that the current results would fall into this reported range. The maximal temporal ventriloquist effect has been reported to occur with an audiovisual separation of approximately 100 ms (Vroomen & Keetels, 2011). These previous findings were used to guide the SOA values chosen for the current series of studies (see Figure 13).

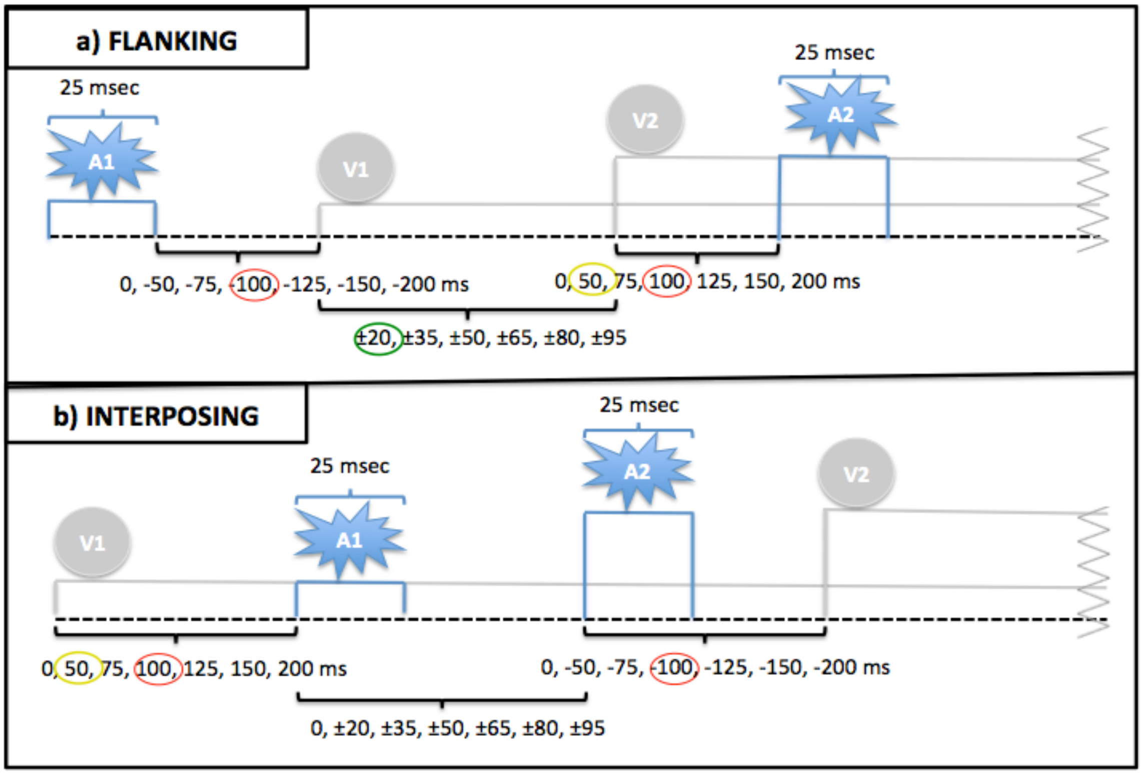


Figure 13. *Stimulus presentation order, SOAs, and duration for Flanking (panel a) versus Interposing (panel b) trials. Ideal AV-SOA temporal ventriloquism is ±100 ms. Average reported AV PSS is at +50 ms (visual leading). Shortest reported visual PSS under ideal circumstances is at ±20 ms.*

6.1.1 Procedure: Studies 1-3

The methods for studies 1-3 are similar except for subtle manipulations that will be discussed in each section, respectively. Participants were seated in a sound attenuated room 21 inches in front of a 22-inch Dell monitor. A chinrest was used to ensure viewing distance was maintained throughout the computer task. Once consent was obtained, participants were asked to complete a hearing test followed by a computer task and, finally, a post-task questionnaire. The consent form was approved by the Georgia Institute of Technology Institutional Review Board under standard protocol (see Appendix A). The hearing test was conducted using the standard procedure for threshold audiometry recommended by the American National Standards Institute (ANSI) (2004). The computer task was comprised of several subtasks. These first measured individual participant 75% threshold levels and baseline PSSs before moving into the main task. Finally, the post-session survey was used to collect participant demographic information and to assess participant strategy during the computer task (see Appendix A).

Hearing test. The hearing test was used to confirm participant self-report of normal auditory functioning. A Micro Audiometrics Corporation DSP Pure Tone Audiometer ® was used following the ANSI standards. Once the headphones were properly adjusted for fit, participants were asked to raise a hand in order to non-verbally communicate tone detection. The experimenter stood behind the participant (i.e., out of the participant's view) for the duration of the test. Each session began with a 1000 Hz tone at 40 dB delivered monaurally. This tone is easily audible by individuals with normally functioning hearing (ANSI, 2004). Once the participant indicated detection, the amplitude was lowered in 10 dB steps until the participant no longer reported detection. The tone was then increased in 10 dB steps until the participant reported detection, at which point the tone was again decreased, and so forth, until the 50%

detection point was determined (i.e., the amplitude at which the participant reported being able to detect the tone 50% of the time). These steps were then repeated starting with a 30 dB tone at frequencies of 2000 Hz, 4000 Hz, 8000 Hz, 500 Hz, and 250 Hz, in that order (as per ANSI standards), for each ear.

Determination of 75% threshold. Threshold values were used to ensure that high and low frequency tones were of equal perceptual loudness as well as to equate the detectability across modality (i.e., crossmodal matching where visual signals seem as ‘bright’ as auditory signals are ‘loud’). A 75% detection threshold was specifically chosen because it allows the bimodal performance advantages of *inverse effectiveness* (see Section 2.1.4) while minimizing the risk of a complete dominance of one modality over the other.

Participant individual 75% visual and auditory thresholds were determined using a detection task. The detection task utilized the QUEST (quick estimate by sequential testing) procedure included in Psychophysics Toolbox (PST), a downloadable collection of procedures that utilize the Matlab programming environment (Brainard, 1997; Watson & Pelli, 1983). For the visual threshold procedure, participants were instructed to maintain their gaze on a central cross while visual stimuli were displayed 4.5° above or below fixation (in Study 2 only, the stimuli were instead presented to the left or right of fixation). Visual stimuli were small, gray disks subtending 0.5° displayed on a black background that remained on-screen until a response was made or until a predefined time-out. Participants were given 2.5 seconds to respond by key-press to a displayed stimulus. Whether the participant detected a given stimulus or not was used to calculate the contrast of the following stimulus until the 75% threshold level was converged upon. For the auditory threshold procedure, participants were again asked to maintain central fixation while high (1000 Hz) and low (330 Hz) frequency auditory signals of 25 ms duration were presented via Altec Lansing ACS90 computer speakers. Again, participants were given 2.5 seconds to respond by key-press and responses determined the amplitude of subsequent stimuli until the 75% threshold level was converged upon. The calculated 75% threshold values for the upper visual, lower visual, high-frequency auditory, and low-frequency auditory stimuli were used during the remainder of the computer task.

Determination of baseline PSS subtasks. After completing the threshold determination subtasks, participants completed the baseline PSS subtasks. As previously discussed, the PSS is defined as the duration of time between the presentation of two stimuli (i.e., the SOA) required for a participant to be equally likely to judge that one versus the other occurred first. Both the visual and audiovisual baseline PSS were determined using a TOJ procedure. For the visual baseline PSS subtask, participants were asked to indicate which of the two visual stimuli (upper or lower) appeared first by responding with a key press. The Quest procedure was used to converge upon the individual's baseline PSS and these values were stored for later comparison with main task results. For the audiovisual baseline PSS subtask, participants were asked to indicate whether an auditory or visual stimulus appeared first. The asynchrony at which the participant was equally likely to indicate one modality versus the other was stored as his or her baseline audiovisual PSS.

Main task. Each trial of the main task included two visual stimuli (upper/lower or left/right) and two auditory stimuli (high/low frequency). As was the case in the earlier measurement subtasks, main task visual stimuli were grey discs (RGB value determined during 75% threshold determination) subtending 0.5° presented on a black background 4.5° from central fixation. Likewise, main task auditory stimuli were puretones (amplitude determined during 75% threshold determination) of 25 ms duration presented at either 1000 or 330 Hz. Participants were asked to make a temporal order judgment about the visual stimuli only. Each trial of the main task could be 'Flanking' or 'Interposing' as well as 'Congruent' or 'Incongruent' as described below:

- ***Flanking trials*** were those where the two auditory signals were presented before and after the pair of visual trials (A1, V1, V2, A2).
- ***Interposing trials*** were those where the two auditory signals were presented between the pair of visual signals (V1, A1, A2, V2).
- ***Congruent trials*** were those where the target visual stimulus was presented adjacent to a synesthetically corresponding auditory stimulus (i.e., high position with high frequency or low position with low frequency).
- ***Incongruent trials*** were those where the visual target and adjacent auditory signals did not synesthetically correspond (i.e., high position with low frequency or low position with high frequency).

6.2 Study 1A: Manipulating Synesthetic Correspondence During TOJ

Study 1A was designed to demonstrate that synesthetic correspondence is an amodal property that serves to bind intersensory signals. In order to accomplish this, performance on a TOJ task was compared across a manipulation of synesthetic correspondence. Synesthetic correspondence was manipulated between task-irrelevant auditory stimuli and adjacent target visual stimuli.

6.2.1 Method: Study 1A

trials. Participants followed the procedure as outlined in section 6.1.1. In Flanking trials, the separation between visual stimuli (i.e., the visual SOA) was 0, ±20, ±35, ±50, ±65, ±80, or ±95 ms (a negative SOA indicates the lower visual stimulus appeared first while a positive SOA indicates that the upper visual stimulus appeared first). In Interposing trials, the separation between auditory stimuli (i.e., the auditory SOA) was 0, ±20, ±35, ±50, ±65, ±80, or ±95 ms (a negative SOA indicates the 330 Hz was presented first). In all trials, the audiovisual SOA was 0, ±50, ±75, ±100, ±125, ±150, or ±200 ms (a negative SOA indicates an auditory-leading pair while a positive SOA indicates a visual-leading pair). For an illustration of Study 1A stimulus timing, see Figure 13. Each possible SOA combination was presented four times at random for a total of 700 trials.

Instructions. Participants in Study 1A were asked to press one key if the upper flash appeared first and to press another key if the bottom flash appeared first.

Participants. Forty-four individuals participated in Study 1A (M = 20, F= 24). Participants were recruited from the Georgia Tech undergraduate participant pool and ranged from 17-25 years old (mean = 2.0 years). Nine participants were excluded (M=5, F=3) in cases when the participant reported being left handed (2), being unable to detect both visual and both auditory stimuli at equal intensity during the main task (5), or being unwilling to complete the study due to fatigue (2).

6.2.2 Results: Study 1A

Baseline PSS. The average baseline PSS for Study 1A participants to perceive audiovisual fusion was 90.2 ms (SD = 24.2 ms). The average baseline SOA required for Study 1A participants to perceive visual fusion was 79.3 ms (SD = 10.1 ms).

Main task. An initial inspection of RT performance data revealed a right-skewed distribution as is common for reaction time data. To correct for this, a log transformation was performed on RT. The distribution of LogRT is a closer approximation of the normal curve (see Figure 14). The transformed data was used for all further Study 1A analyses.

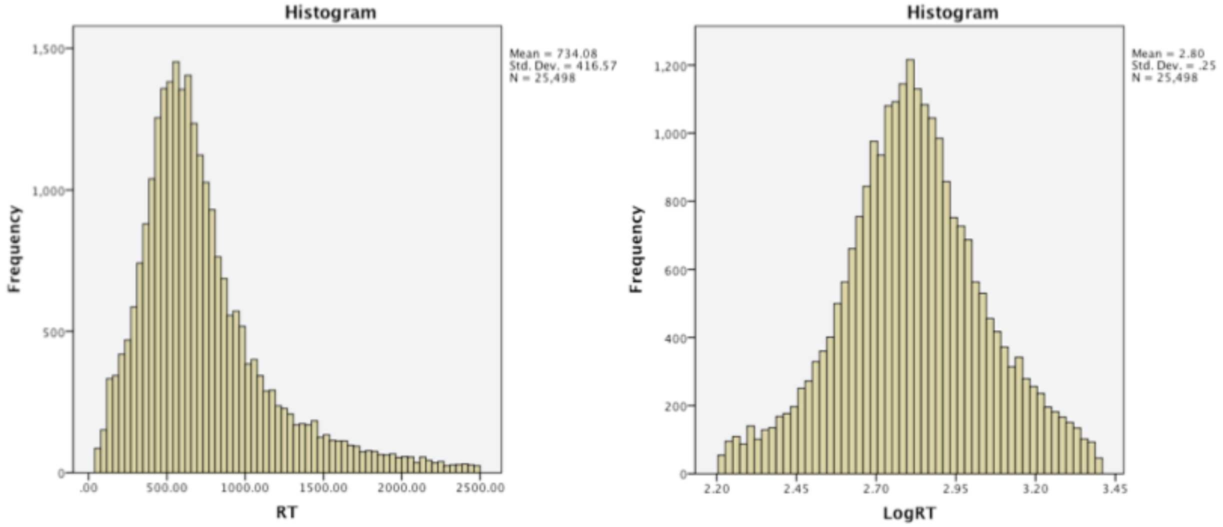
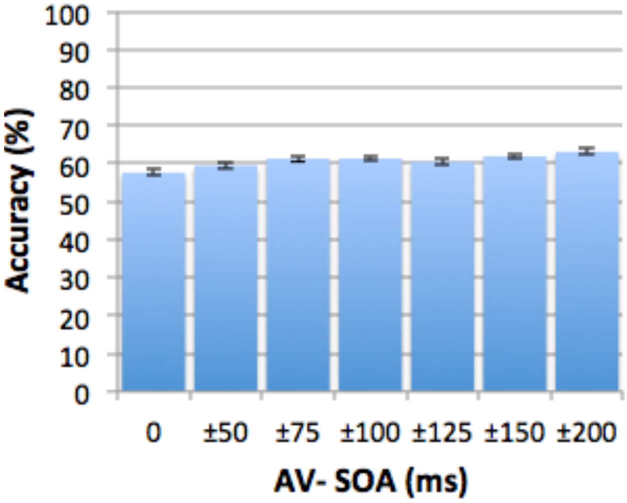


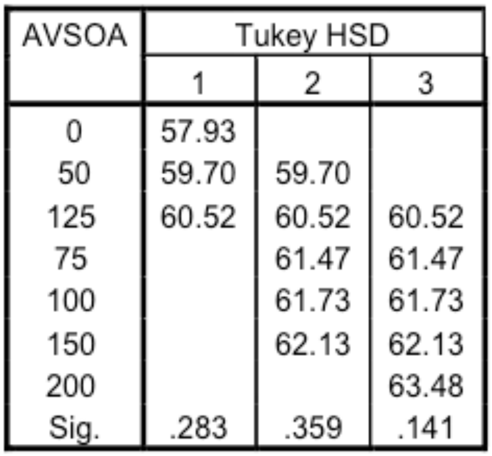
Figure 14 A log transform is performed on Study 1A RT data in order to correct for an initial right skew of the RT distribution.

A MANOVA was run using Accuracy and LogRT as the independent factors with Congruency (Congruent, Incongruent), AV-SOA (0, ±50, ±75, ±100, ±125, ±150, ±200 ms), and Unimodal SOA (0, ±20, ±35, ±50, ±65, ±80, ±95 ms) as the dependent factors. Significant main effects were revealed for all dependent variables: Congruency Wilks' Λ = .998, F(2, 25387) = 29.330, $p < .001$; Wilks' Λ = .993, AV-SOA F(12, 50776) = 12.869, $p < .001$, and Unimodal SOA Wilks' Λ = .996, F(10, 50776) = 7.959, $p < .001$. There were also significant interactions of Congruency x Unimodal SOA Wilks' Λ = .999, F(10, 50776) = 2.153, $p = .011$ and AV-SOA x Unimodal SOA Wilks' Λ = .995, F(82, 50776) = 1.549, $p = .020$ (see Appendix B).

Congruency. Tests of between-subjects effects show that accuracy performance was significantly different across Congruency. In Congruent trials, participant accuracy averaged 63.7% compared to 57.9% accuracy in Incongruent trials F(1, 25388) = 54.297, p < .001. Congruency did not affect RT performance.

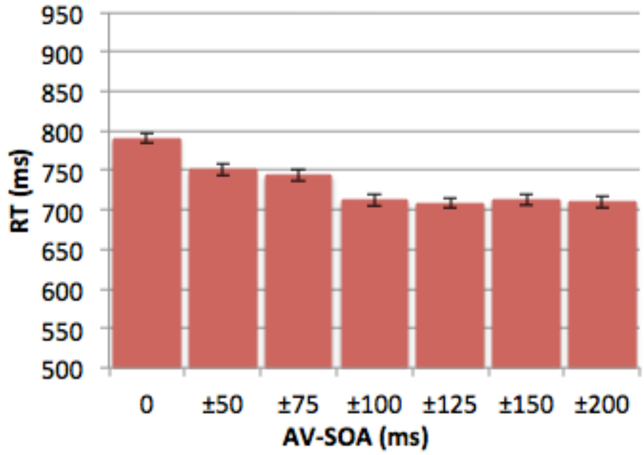
AV-SOA. Tests of between-subjects effects show that accuracy and reaction times did vary significantly across the levels of audiovisual separation. Post-hoc analyses reveal that accuracy improved with longer SOAs up to an audiovisual separation of 100 ms F(6, 25388) = 2.448, *p* = .017 (see Figure 15). Also, RTs were faster with longer AV-SOAs up to an audiovisual separation of 100 ms F(6, 25388) = 24.094, *p* < .001 (see Figure 16).





AVSOA	Tukey HSD		
	1	2	3
0	57.93		
50	59.70	59.70	
125	60.52	60.52	60.52
75		61.47	61.47
100		61.73	61.73
150		62.13	62.13
200			63.48
Sig.	.283	.359	.141

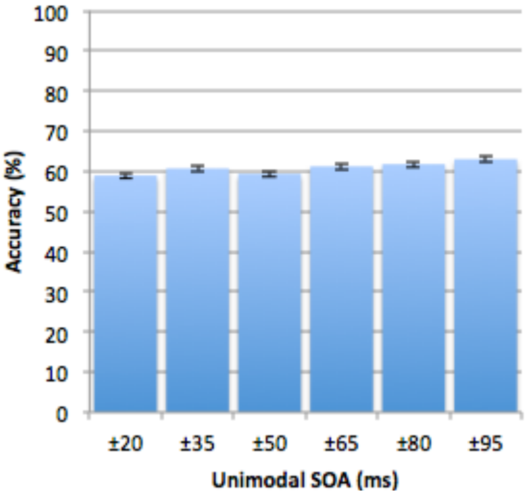
Figure 15. Study 1A Main effect of Unimodal SOA. These are the post-hoc results for accuracy performance across AV-SOAs.



AVSOA	Tukey HSD		
	1	2	3
200	2.769		
150	2.775		
125	2.779		
100	2.785		
75		2.810	
50		2.818	
0			2.848
Sig.	.130	.785	1.000

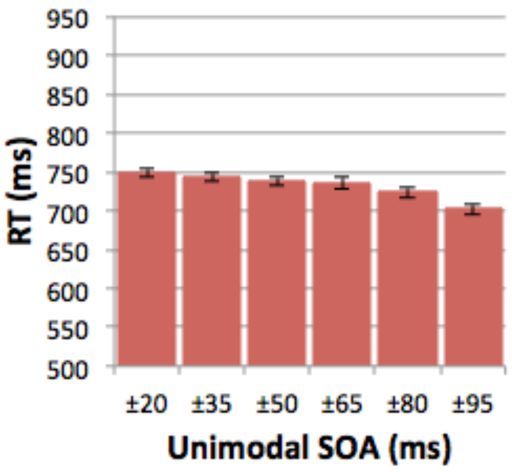
Figure 16. Study 1A main effect of AV-SOA, These are the post-hoc results for RT performance across AV-SOAs.

Unimodal SOA. The asynchrony between the inner-pair of unimodal signals (i.e., visual signals in flanking trials or auditory signals in interposing trials) did significantly affect both accuracy and RT performance. Post-hoc analyses reveal that accuracy tended to increase with increasing Unimodal SOAs F(5, 25388) = 3.979, $p < .001$ (see Figure 17). Also, RT performance tended to be best at longer Unimodal SOAs F(5, 25388) = 14.377, $p < .001$ (see Figure 18).



Unimodal SOA	Tukey HSD	
	1	2
20	59.22	
50	59.65	
35	60.87	60.87
65	61.50	61.50
80	62.06	62.06
95		63.24
Sig.	.092	.242

Figure 17 Study 1A main effect of Unimodal SOA. These are the post-hoc results for accuracy performance across Unimodal SOAs.



Unimodal SOA	Tukey HSD		
	1	2	3
95	2.77		
80	2.79	2.79	
65		2.80	2.80
50		2.80	2.80
20			2.81
35			2.81
Sig.	.081	.273	.142

Figure 18. Study 1A main effect of Unimodal SOA. These are the post-hoc results for RT performance across Unimodal SOAs.

Congruency x Unimodal SOA. Tests of between-subjects effects reveal that this interaction was driven by accuracy performance F(5, 25388) = 3.733, $p < .001$. Accuracy performance was significantly better during Congruent than Incongruent trials at all Unimodal SOAs except for at 65 ms (see Figure 19).

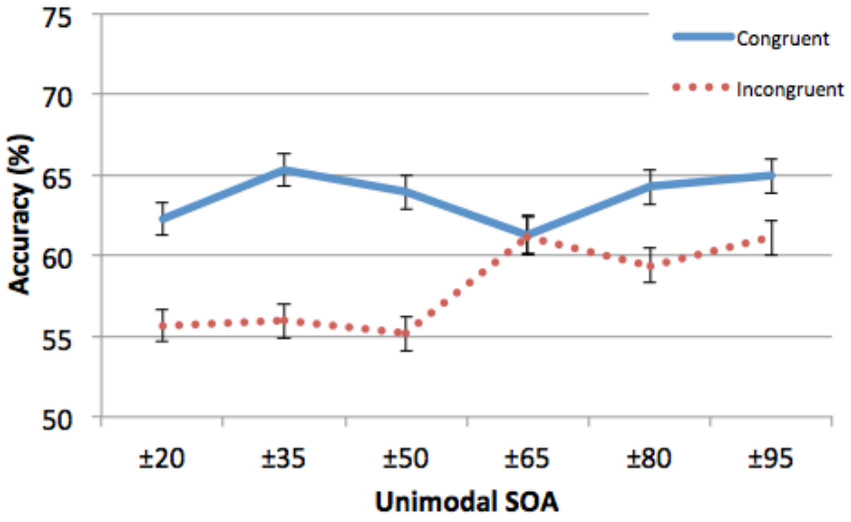


Figure 19. Study 1A interaction between Congruency and Unimodal SOA.

AV-SOA x Unimodal SOA. Tests of between-subjects effects reveal that this interaction was driven by RT performance only. F(41, 25388) = 2.136, $p = .004$ (see Figure 20).

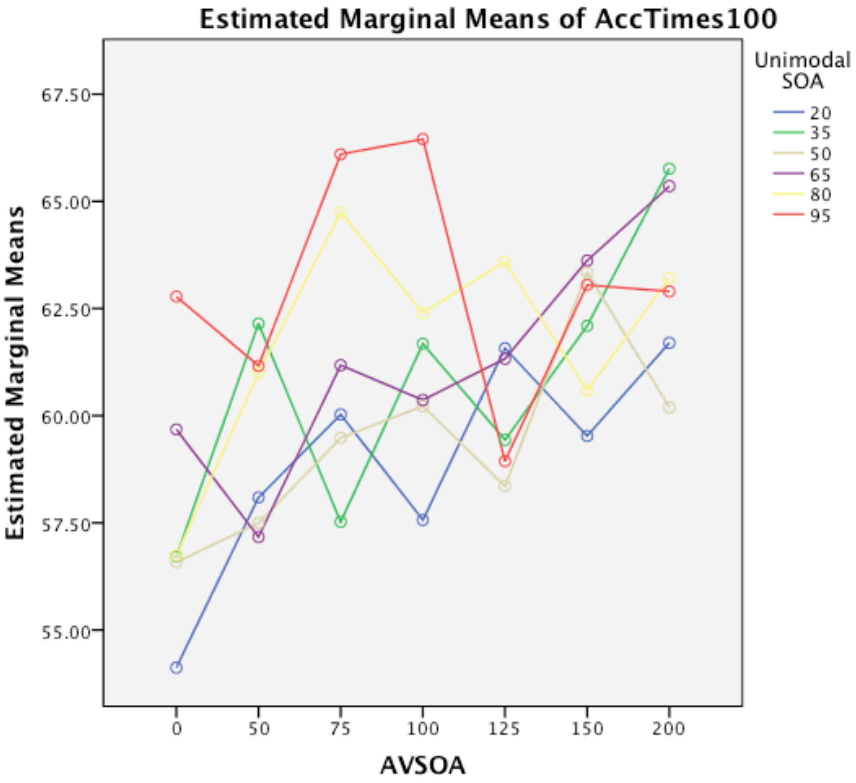


Figure 20. Study 1A interaction between Unimodal SOA and AV-SOA.

6.2.3 Discussion: Study 1A

Synesthetic correspondence did affect accuracy in the predicted direction with performance being significantly higher when auditory signals synesthetically matched visual targets. Congruency did not affect RT performance, indicating that the accuracy performance enhancement in Congruent trials cannot be explained by a speed-accuracy tradeoff. Overall, performance increased with increasing separation between audiovisual pairs up to an AV-SOA of 100 ms. Recall that an audiovisual asynchrony of approximately 100 ms is the reported SOA for maximal temporal ventriloquism to occur (see section 6.6 for further discussion). Performance also tended to increase with increasing Unimodal SOAs. This finding was expected because the TOJ task difficulty should decrease with increasing separation between the target signals. The interaction between Congruency and Unimodal SOA was unexpected. However, upon closer inspection it is apparent that performance for Congruent trials was significantly better than for Incongruent trials in all but one case (i.e., a separation of 65 ms). The interaction between Unimodal SOA and AV-SOA was also unexpected. Instead, it was expected that accuracy would improve uniformly with longer Unimodal SOAs and with longer AV-SOAs. Indeed, the data generally show the expected pattern up to an AV-SOA of 100 ms (see Figure 20. However, after that point, the pattern shows that some of the Unimodal SOAs jump out of the expected rank order. Most notably, the longest Unimodal SOA of 95 ms switches order with Unimodal SOAs of 35 ms and 65 ms. Whether these unexpected results were spurious or systematically driven will be brought to light by the results of subsequent studies reported here.

6.3 Study 1B: A Shift from Spatial towards Temporal Emphasis

The TOJ task in Study 1A is representative of a typical TOJ procedure. In a variation of the TOJ task, participants might instead be asked to always indicate the position of the second stimuli rather than the first, or to make judgments based on offset rather than onset (Boenke, et al., 2009). In each of these TOJ variations, the participant's task is to indicate the position of one target or the other. Though this task is labeled as a 'temporal' one, perhaps by asking participants to indicate the *location* of the target stimulus, there is more emphasis paced on the 'spatial' aspects of the task. Study 1B was designed to shift the TOJ task toward the temporal end of the spatiotemporal continuum. In doing so, it was hypothesized that the influence of task-irrelevant auditory signals found in Study 1A would not only be replicated, but also enhanced. Study 1B accomplished this shift from a spatial towards a temporal emphasis by a manipulation of the task instructions.

6.3.1 Method: Study 1B

Trials. Participants followed the procedure as outlined in section 6.1.1. As in Study 1A, there were 700 trials whose order was randomized. See Figure 13 for stimulus presentation order and timing.

Instructions. Rather than ask participants to indicate the location of a target stimulus, Study 1B instructions specified a target location (always upper or always lower) and asked participants to indicate whether the target appeared first or second (i.e., to indicate the temporal order of the target). The location of the target visual stimulus was counterbalanced across participants. For instance, a participant might be asked to press one button when the upper stimulus appeared first and to press another button when the upper stimulus appeared second. Other than this instruction manipulation, Study 1A and Study 1B methods were identical.

Participants. Forty-three people participated in Study 1B (M = 23, F =21). Participants were recruited from the Georgia Tech undergraduate participant pool and ranged from 17-24 years old (mean = 19.14 years). Eleven participants were excluded (M = 6, F = 5) in cases where the participant reported being left handed (3), being unable to detect both visual and both auditory stimuli at equal intensity in the main task (4), described the main task incorrectly in the post-study survey (3), or being unable to complete the study due to fatigue (1).

6.3.2 Results: Study 1B

Baseline PSS. The average baseline SOA required for Study 1B participants to perceive audiovisual fusion was 95.6 ms (SD = 12.0 ms). The average baseline SOA required for Study 1B participants to perceive visual fusion was 72.9 ms (SD = 23.4 ms).

Main task. An initial inspection of RT performance data revealed a right-skewed distribution as is common for RT data. To correct for this, a log transformation was performed on RT data. The distribution of LogRT is a closer approximation of the normal curve (see Figure 21). The transformed data was used for all further Study 1B analyses.

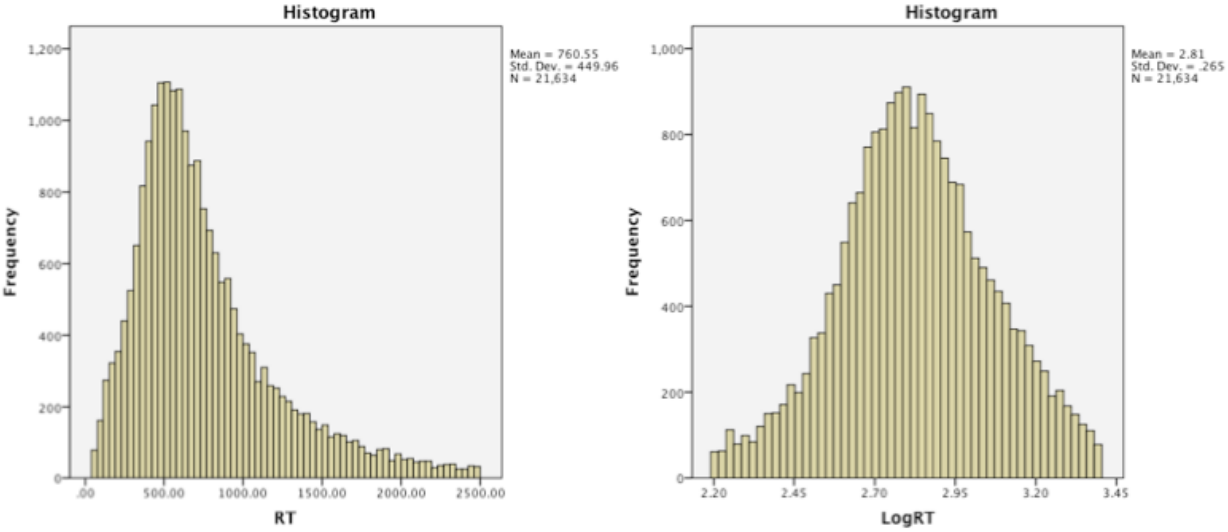
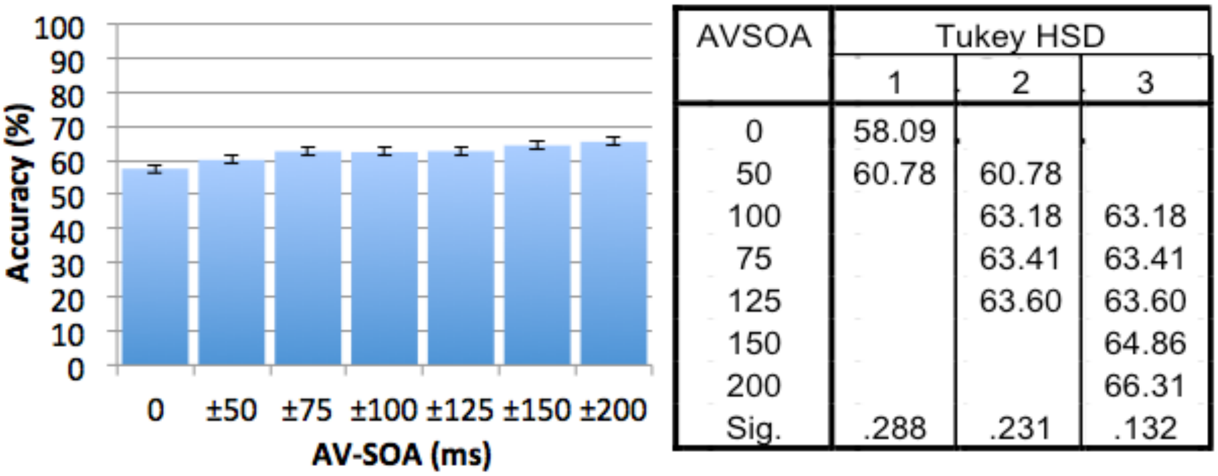


Figure 21. A log transform was performed on Study 1B RT data in order to correct for an initial right skew of the RT distribution.

A MANOVA was run using Accuracy and LogRT as the independent factors with Congruency (Congruent, Incongruent), AV SOA (0, ±50, ±75, ±100, ±125, ±150, ±200 ms), and Unimodal SOA (0, ±20, ±35, ±50, ±65, ±80, ±95 ms) as the dependent factors. Significant main effects were revealed for all dependent variables: Wilks' Λ = .979, Congruency F(2, 20704) = 227.190, *p* < .001; AV-SOA Wilks' Λ = .994, F(12, 41410) = 10.652, *p* < .001, and Unimodal SOA Wilks' Λ = .996, F(10, 41410) = 9.138, *p* < .001 (see Appendix C).

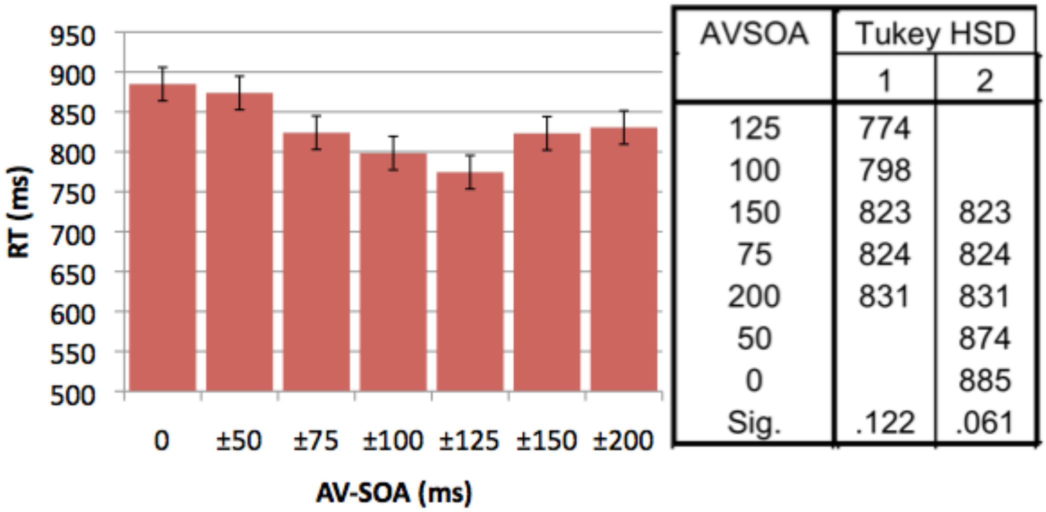
Congruency. Tests of between-subjects effects reveal that synesthetic congruency did affect accuracy in the predicted direction. In Congruent trials, participant accuracy averaged 69.1% compared to 55.3% accuracy in Incongruent trials F(1, 20705) = 451.474, *p* < .001. Synesthetic congruency did not affect RTs, indicating that the accuracy performance enhancement in Congruent trials cannot be explained by a mere speed-accuracy tradeoff.

AV-SOA. The asynchrony between auditory and visual signals (i.e., AV-SOA) did affect both accuracy and RT performance. Post-hoc analyses reveal that accuracy improved with longer SOAs (see Figure 22) and RTs were quickest at an audiovisual separation between 100 and 125 ms (see Figure 23).



AVSOA	Tukey HSD		
	1	2	3
0	58.09		
50	60.78	60.78	
100		63.18	63.18
75		63.41	63.41
125		63.60	63.60
150			64.86
200			66.31
Sig.	.288	.231	.132

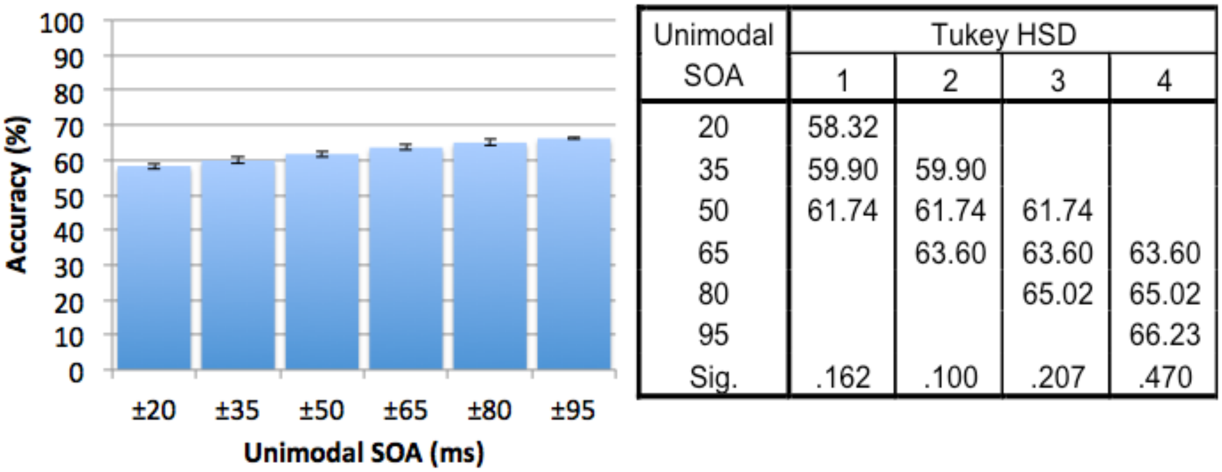
Figure 22. Study 1B Main Effect of AV-SOA. These are the post-hoc results for accuracy performance across AV-SOAs.



AVSOA	Tukey HSD	
	1	2
125	774	
100	798	
150	823	823
75	824	824
200	831	831
50		874
0		885
Sig.	.122	.061

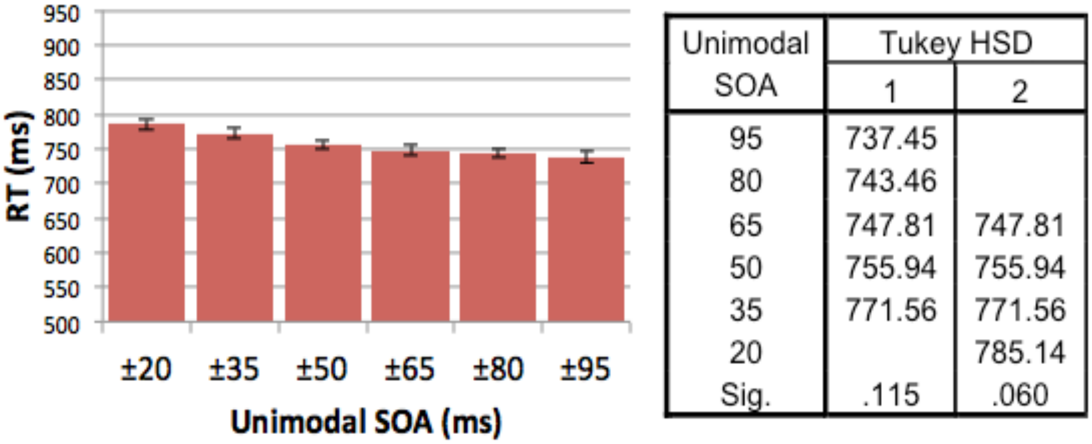
Figure 23. Study 1B Main Effect of AV-SOA. These are the post-hoc results for reaction time across AV-SOAs.

Unimodal SOA. The asynchrony between the inner-pair of unimodal signals (i.e., visual signals in flanking trials or auditory signals in interposing trials) affected both accuracy and RT performance at the $p < .001$ level. Post-hoc analyses reveal that accuracy (see Figure 24) and RT (see Figure 25) performance were best at longer unimodal SOAs. This finding is expected because the TOJ task difficulty decreases with increasing separation between the target signals.



Unimodal SOA	Tukey HSD 1	2	3	4
20	58.32			
35	59.90	59.90		
50	61.74	61.74	61.74	
65		63.60	63.60	63.60
80			65.02	65.02
95				66.23
Sig.	.162	.100	.207	.470

Figure 24. Study 1B Main effect of Unimodal SOA. These are the post-hoc results for accuracy across Unimodal SOAs.



Unimodal SOA	Tukey HSD 1	2
95	737.45	
80	743.46	
65	747.81	747.81
50	755.94	755.94
35	771.56	771.56
20		785.14
Sig.	.115	.060

Figure 25. Study 1B Main effect of Unimodal SOA. These are the post-hoc results for reaction time across Unimodal SOAs.

6.3.3 Instruction Manipulation Results: Study 1B compared with Study 1A

A MANOVA was run to compare studies 1A and 1B using Accuracy and LogRT as the independent factors with Instruction Emphasis (Spatial, Temporal) and Congruency as the dependent factors. This analysis revealed a significant main effect for Instruction Emphasis Wilks' Λ = .999, F(2, 46282) = 9.082, *p* < .001; a significant main effect for Wilks' Λ = .989, Congruency F(2, 46282) = 259.583, *p* < .001; and a significant interaction for Instruction Emphasis x Congruency Wilks' Λ = .998, F(2, 46282) = 38.025, *p* < .001 (see Appendix D).

Instruction Emphasis x Congruency. Tests of between-subjects effects show that for Congruent trials, Study 1B accuracy performance was significantly higher (70%) than Study 1A performance accuracy (64%). For Incongruent trials however, Study 1B accuracy performance was significantly lower (55%) than Study 1A accuracy performance (58%). See Figure 26. There was no significant Instruction x Congruency interaction for RT performance.

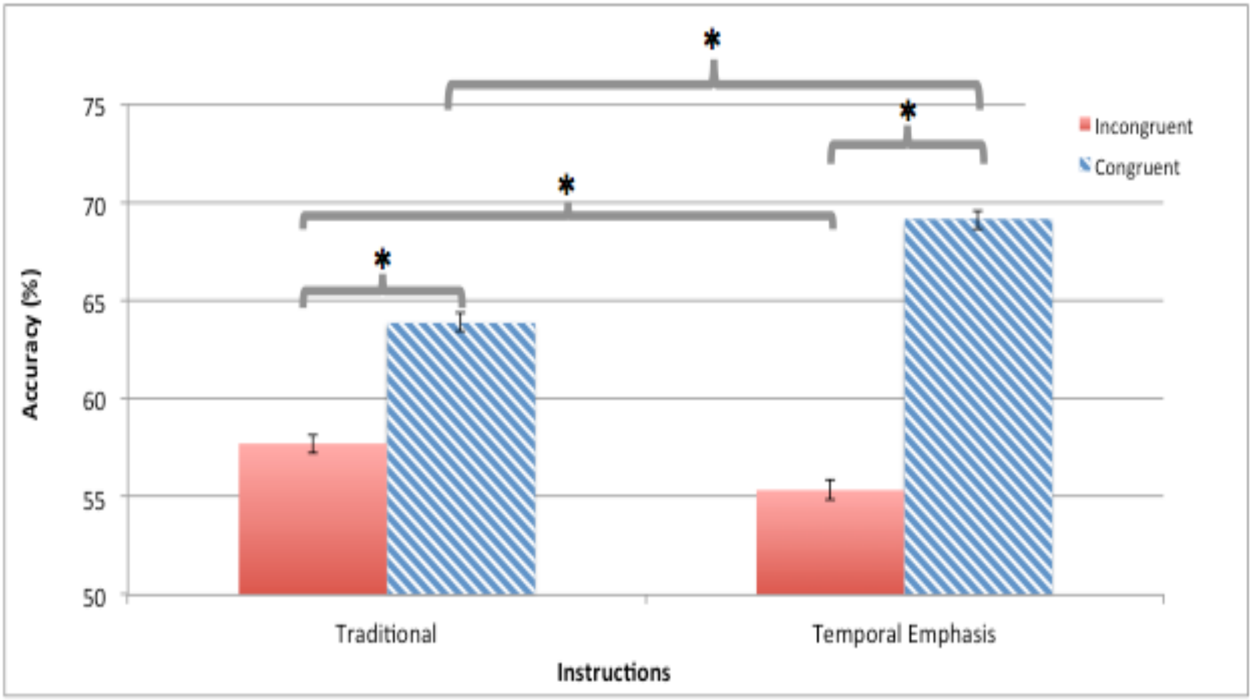


Figure 26. Instruction by congruency interaction. Asterisk (*) marks significant differences at a level of p < .001.

6.3.4 Discussion: Study 1B

Study 1B results upheld the main hypothesis by supporting that synesthetic correspondence does affect audiovisual perceptual binding. Congruency affected accuracy in the predicted direction with performance being higher when auditory signals synesthetically matched visual targets. Congruency did not affect RT performance, indicating that the accuracy performance enhancement in Congruent trials cannot be explained by a mere speed-accuracy tradeoff. Performance tended to increase with increasing separation between audiovisual pairs. Reaction time was fastest at an audiovisual separation of 100 - 125 ms. This pattern of results supports the possibility that temporal ventriloquism, which has been reported to be maximally effective at an audiovisual separation of about 100 ms, is at play here. Performance also tended to increase with increasing Unimodal SOAs. This finding was expected as the TOJ task difficulty should decrease with increasing separation between the target signals.

The interaction between Instruction Emphasis and Congruency demonstrates that slightly shifting the instructions from a more spatial emphasis (study 1A) to a more temporal emphasis (study 1B) moderated the effect of synesthetic correpsondence (see Figure 26). The RT results demonstrate that this performance pattern cannot be explained by a speed-accuracy tradeoff. Specifically, it is suggested here that, because there is a greater influence placed on the temporal demands in Study 1B, the auditory signal was more heavily weighted in the fused audiovisual percept than it was in Study 1A. The pattern of results revealed by the significant interaction between Instruction Emphasis and Congruency supports the idea that auditory capture was enhanced when temporal aspects of the main task were emphasized. In terms of the spatiotemporal continuum model first discussed in section 5.1.2, though TOJ tasks in both studies 1A and 1B fall towards the temporal of the spectrum, the instruction emphasis in Study 1B push this task closer to the temporal extreme (see Figure 27).

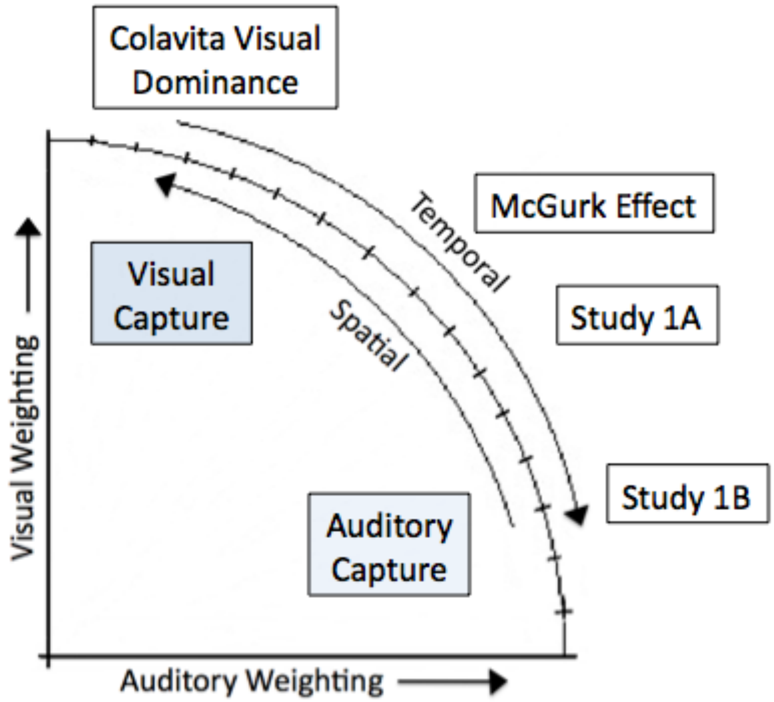


Figure 27. Illustration of the Spatiotemporal Continuum model. Colavita Visual Dominance is presented as an example of a task that is heavily spatial and utilizes a heavier visual weighting. The McGurk Effect is an example of audiovisual fusion and falls in the middle of the range. While studies 1A and 1B fall near the temporal end of the spectrum, Study 1B falls closer to the temporal extreme.

6.4 Study 2: Manipulating Synesthetic Dimensional Overlap

Study 2 was a horizontal extension of the vertical findings from studies 1A and 1B. In addition to the pitch-vertical position mapping, there is also evidence supporting a pitch-lateral position mapping albeit the association is not as strong. The goal of this extension was two-fold: 1) to examine and compare a mapping with less synesthetic dimensional overlap than the mapping considered thus far, and 2) to provide evidence that the pitch-vertical position mapping is not merely a linguistic association, but a truly perceptual phenomenon.

Some researchers have provided evidence to support that low pitches are associated with leftward space while high pitches are associated with rightward space (e.g., Mudd, 1963; Rusconi, Kwan, Giordano, Umilta, & Butterworth, 2006). In the literature, the lateral position mapping has proven to be less robust than the vertical position mapping. It is argued here that this weaker association is due to a lesser degree of dimensional overlap in the lateral than in the vertical case. Some researchers have suggested that the pitch-vertical position association is purely a linguistic one (i.e., ‘low’ and ‘high’ can be used to describe both the pitch of the auditory stimulus and the location of the visual stimulus) (e.g., Gallace & Spence, 2006). If a linguistic mechanism were solely responsible for pitch-position mappings, then significant findings would be expected in the vertical, but not the lateral case. Contrary to this, the aim of Study 2 was to find a significant pitch-lateral position mapping thus providing evidence for an underlying synesthetically driven perceptual mechanism.

6.4.1 Method: Study 2

Trials. Participants followed the procedure as outlined in section 6.1.1. As in studies 1A and 1B, there were 700 trials whose order was randomized. See Figure 13 for stimulus presentation order and timing. Due to the significant and enhanced effect of auditory driving supported in Study 1B, the same, temporally emphasized instructions were used in Study 2. In fact, the procedure used for Study 2 did not differ from Study 1B with the only exception being that visual stimuli were presented 4.5° to the left and right of central fixation.

Participants. Fifty-two people participated in Study 2 (M =30, F =22). Participants were recruited from the Georgia Tech undergraduate participant pool and ranged from 17-24 years old (mean = 19.7 years). Eighteen participants were excluded (M = 10, F = 8) in cases when the participant reported being left handed (3), being unable to detect both visual and both auditory stimuli at equal intensity in the main task (6), or described the main task incorrectly in the post-study survey (9).

6.4.2 Results: Study 2

Baseline PSS. The average baseline asynchrony required for Study 2 participants to perceive audiovisual fusion was 83.9 ms (SD = 28.9 ms). The average baseline SOA required for Study 2 participants to perceive visual fusion was 71.5 ms (SD = 10.1 ms).

Main task. An initial inspection of RT performance data revealed a right-skewed distribution as is common for RT data. To correct for this, a log transformation was performed on RT data. The distribution of LogRT is a closer approximation of the normal curve (see Figure 28). The transformed data was used for all further Study 2 analyses.

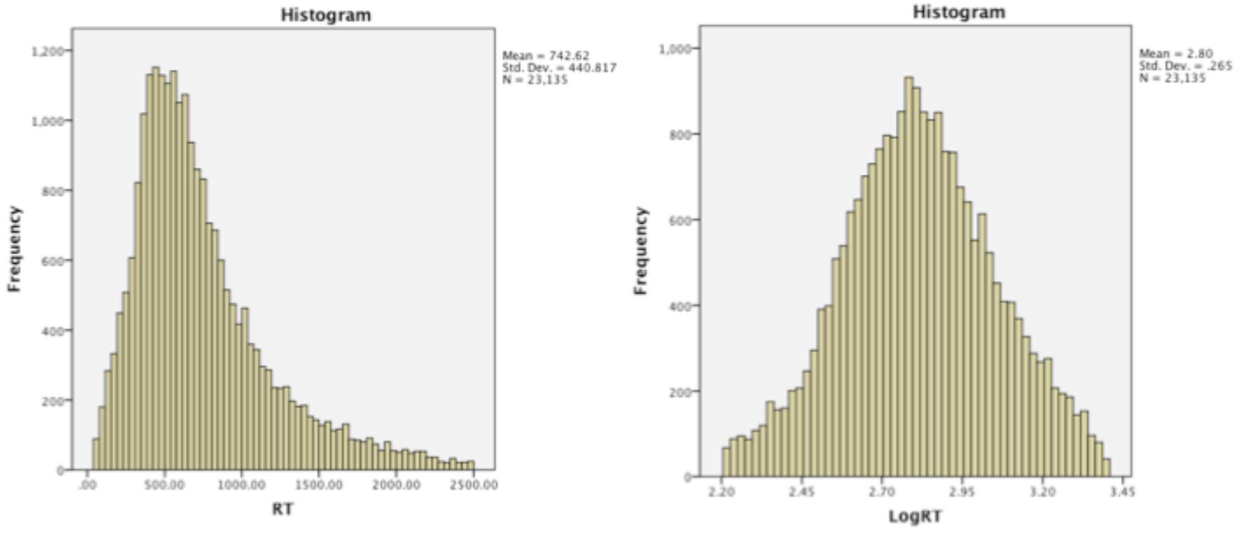


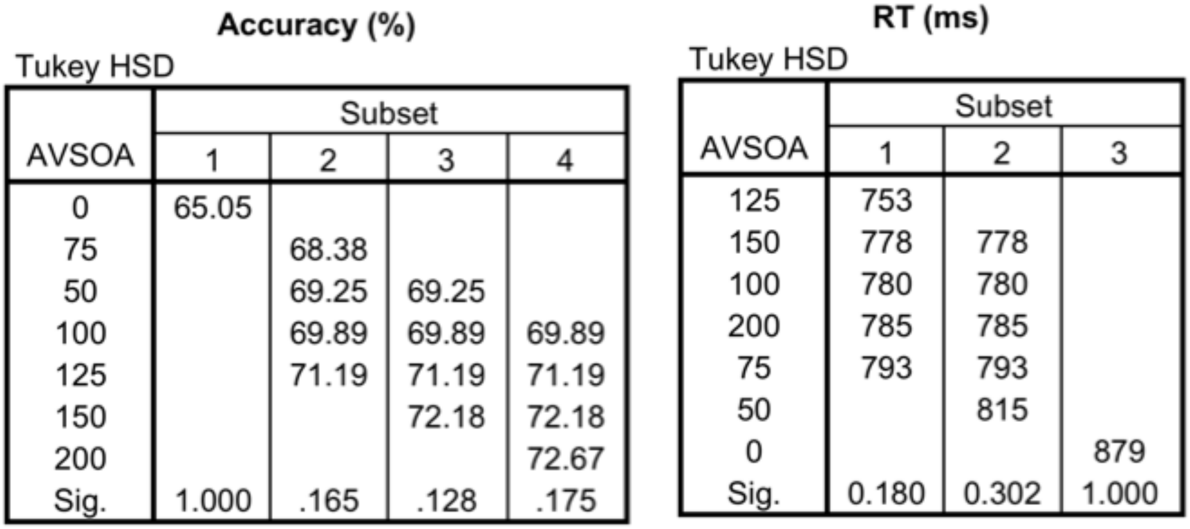
Figure 28. A log transform was performed on Study 2 RT data in order to correct for an initial right skew of the RT distribution.

A MANOVA was performed using Accuracy and LogRT as the independent factors with Congruency (Congruent, Incongruent), AV SOA (0, ±50, ±75, ±100, ±125, ±150, ±200 ms), and Unimodal SOA (0, ±20, ±35, ±50, ±65, ±80, ±95 ms) as the dependent factors. Significant main effects were revealed for all dependent variables: Congruency Wilks' Λ = .998, F(2, 22125) = 26.306, *p* < .001; AV-SOA Wilks' Λ = .986, F(12, 44252) = 26.306, *p* < .001, and Unimodal SOA Wilks' Λ = .989, F(10, 44252) = 23.887, *p* < .001 (see Appendix E).

Congruency. Recall that in Study 2, trials were considered congruent when left flashes were paired with low-frequency sounds or when right flashes were paired with high-frequency sounds. Conversely, trials were considered incongruent when left flashes were paired with high-frequency sounds or when right flashes were paired with low-frequency sounds. Tests of between-subjects effects reveal that synesthetic correspondence affected accuracy in the predicted direction. For Congruent trials, participant accuracy averaged 71.5% compared to 68.1% accuracy for Incongruent trials F(1, 22210) = 34.081, *p* < .001. Synesthetic congruency did not affect RTs, suggesting there was no speed-accuracy tradeoff.

AV-SOA. The asynchrony between auditory and visual signals (i.e., AV-SOA) affected both accuracy and RT performance at the *p* < .001 level. Post-hoc analyses reveal that accuracy improved with longer AV-SOAs and RTs were quickest at an audiovisual separation of 125 ms (see Table 1).

Table 1. Study 2 AV-SOA homogenous subgroups for accuracy and RT performance.



Accuracy (%)

Tukey HSD

AVSOA	Subset			
	1	2	3	4
0	65.05			
75		68.38		
50		69.25	69.25	
100		69.89	69.89	69.89
125		71.19	71.19	71.19
150			72.18	72.18
200				72.67
Sig.	1.000	.165	.128	.175

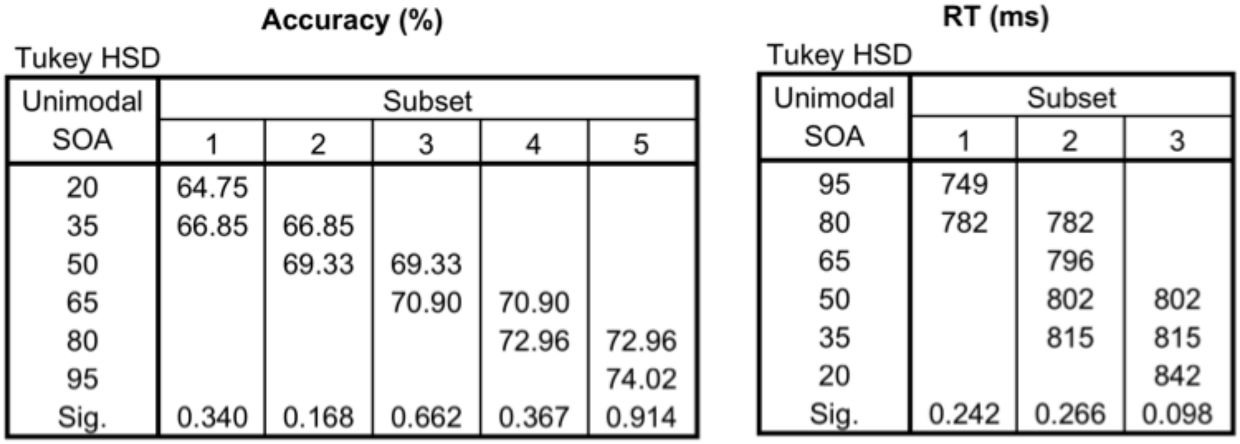
RT (ms)

Tukey HSD

AVSOA	Subset		
	1	2	3
125	753		
150	778	778	
100	780	780	
200	785	785	
75	793	793	
50		815	
0			879
Sig.	0.180	0.302	1.000

Unimodal SOA. The asynchrony between the inner-pair of unimodal signals (i.e., visual signals in flanking trials or auditory signals in interposing trials) affected both accuracy and RT performance at the *p* < .001 level. Post-hoc analyses reveal that accuracy and RT performance were best at longer unimodal SOAs (see Table 2). This finding is expected as the TOJ task difficulty decreases with increasing separation between the target signals.

Table 2. Study 2 Unimodal SOA homogeneous subgroups for accuracy and RT performance.



Accuracy (%)

Tukey HSD

Unimodal SOA	Subset 1	2	3	4	5
20	64.75				
35	66.85	66.85			
50		69.33	69.33		
65			70.90	70.90	
80				72.96	72.96
95					74.02
Sig.	0.340	0.168	0.662	0.367	0.914

RT (ms)

Tukey HSD

Unimodal SOA	Subset 1	2	3
95	749		
80	782	782	
65		796	
50		802	802
35		815	815
20			842
Sig.	0.242	0.266	0.098

Target orientation. In order to compare the horizontal-pitch association with the vertical-pitch association, Study 2 was compared with Study 1B. Though congruency did affect accuracy performance at the $p < .001$ level in both cases, the effect size is larger in the vertical case ($\eta^2 = 0.144$) than the horizontal case ($\eta^2 = .047$). This is considered a large effect size for Study 1B (i.e., $\eta^2 \approx 0.14$) and a small effect size for Study 2 (i.e., $\eta^2 \approx .01$). Though synesthetic correspondence significantly affected performance in both the vertical and horizontal cases presented here, there was a larger effect size for the vertical case.

6.4.3 Discussion: Study 2

Despite that the lateral-pitch association has been shown to be less robust in the literature than the vertical-pitch association, the pattern of results for Study 2 were quite similar to those from studies 1A and 1B. Study 2 results upheld the main hypothesis by supporting that synesthetic correspondence affected audiovisual perceptual binding. Congruency affected accuracy in the predicted direction with performance being higher when auditory signals synesthetically matched visual targets. Synesthetic congruency did not affect RT performance, indicating that the accuracy performance enhancement in Congruent trials cannot be explained by a mere speed-accuracy tradeoff. Performance tended to increase with increasing separation between audiovisual pairs. Reaction time was fastest at an audiovisual separation of 125 ms. Again; this pattern of results supports the possibility that temporal ventriloquism is at play here. However, this possibility is not explicitly tested until Study 3. Performance also tended to increase with increasing Unimodal SOAs. This finding is expected because the TOJ task difficulty decreases with increasing separation between the target signals.

Though the visual degree of separation between the central fixation point and the visual targets remained constant between studies 1 and 2, attention allocation along the vertical versus horizontal meridian of the visual fields are not equal. (As would be expected, participant performance was generally higher in Study 2 than in 1A or 1B). For this reason, the main task in Study 1 is not perceptually comparable to the main task in Study 2. This was the resason the meta-analytic approach of comparing effect sizes was used to compare studies 1 and 2. Study 2 was specifically compared with Study 1B because they both utilize the more temporally emphasized instructions. The finding of a larger effect size for Study 1B (vertical case) than Study 2 (horizontal case) fits well with the literature given previous findings showing the lateral-pitch association to be less robust than the vertical-pitch association.

A comparatively smaller effect size aside, the significant main effect of Study 2 provides evidence against a ‘purely linguistic’ account of the broader position-pitch association. Further, it is possible that the linguistic argument is putting the cart before the horse altogether. Consider the influence that inherent synesthetic correspondences have had in the development of language (discussed in section 2.3.1) (Werner & Wapner, 1952). Likely, it is no coincidence that many languages (e.g., English, Chinese, German, Spanish, etc.) classify pitch in terms of having a ‘spatial’ relationship (Rusconi, et al., 2006). Pratt states, “the (tone) of greater frequency is called higher, not because of any extraneous associations with altitude, but simply because it is perceived as occupying a higher position in *phenomenological* space" (1930, p. 283).

6.6 Interim Discussion: Studies 1-2

In each of the studies presented thus far, synesthetic correspondence significantly affected performance in the predicted direction. Accuracy for Congruent trials was significantly higher than for Incongruent trials in each case. Taken together, these results demonstrate that the presence of irrelevant, auditory stimuli affect participants' temporal order judgments for visual target stimuli when the synesthetic congruency between audiovisual pairs is orthogonally varied. This suggests that perceived temporal occurrence and synesthetic congruency are integral (as opposed to separable) variables. While it is widely accepted that commonalities in space and time are amodal properties serving to bind signals together, these results suggest that synesthetic correspondence also acts as a significant binding property.

Performance also tended to improve with increasing separation between the inner two stimuli presented. Given that the main task was a temporal order judgment task, it was anticipated that performance would improve with increasing separation of the target stimuli. This consistent pattern of results is taken as assurance that participants remained on-task and attempted to complete main task trials as instructed. Additional checks that the participant understood the instructions were included in the post-session survey, which specifically asked the participant to explain the main task and to describe the strategy he or she used during main task trials (see Appendix A). Recall that individuals who explained the main task incorrectly, or who described a strategy of relying on auditory signals, were excluded from analysis.

Performance was also significantly affected in each case by the separation between pairs of audiovisual signals (i.e., AV-SOA). The combined results show that performance tended to improve across the range of AV-SOAs included here up to a separation around 100-125 ms. A general alerting effect is one interpretation of this data. Though the auditory signals are task-irrelevant, in cases where the first stimulus presented was auditory, it may have acted to cue participant attention. However, a general alerting effect interpretation does not explain this pattern of results for cases where the first stimulus presented was visual and it does not explain why performance tended to peak around an audiovisual separation of 100-125 ms. Interestingly, this range aligns with the reported audiovisual separation for the maximal temporal ventriloquist effect to occur (Vroomen & Keetels, 2011). It is suggested that the audiovisual pairs presented here are bound and treated as perceptual wholes. Study 3 was designed to explicitly test this possibility.

6.5 Study 3: Investigating the Temporal Ventriloquist Effect

Study 3 was designed to determine whether the temporal ventriloquist effect cited in the literature was being reproduced in the current set-up. This was accomplished by setting Visual SOAs to be equal across both Flanking and Interposing trials such that they could be directly compared. In order to accomplish this, the SOA values were adjusted. As in studies 1-2, participant performance was a dependent variable analyzed across the independent factors. It is expected that the pattern of results for Study 3 should resemble those from studies 1-2. Additionally, the psychometric function relating the physical stimulus to the psychological response is examined. Psychometric functions are often used to describe the relationship between stimulus intensity and participant detection. Here, the analogous relationship examined was between the separation of the visual signals (i.e., the Visual SOA) and the participant's determination of temporal order. The psychometric function is illustrated by examining the proportion of responses where the participant indicated that the upper visual stimulus occurred first across the range of Visual SOA values. This proportion should approach a value of one with increasing positive Visual SOAs (i.e., where the upper signal actually occurs first) and approach a value of zero with decreasing negative Visual SOAs (i.e., where the lower signal occurs first). As such, the psychometric function was expected to approximate a sigmoid function (an S-curve). The point where the response proportion equals .5 is the point of subjective simultaneity (PSS). Recall that the visual PSS is the temporal separation between the target visual signals where the participant is equally likely to indicate that one signal occurred first as the other. Because the performance results reported thus far have suggested that the auditory signals act to pull the visual signals into temporal register, it was expected that the PSS values would be larger for Flanking trials than for Interposing trials. This finding would indicate that temporal ventriloquism does occur in the current set-up.

6.5.1 Method: Study 3

Trials. Participants followed the procedure as outlined in section 6.1.1. Stimulus order and timing differed in Study 3 compared to studies 1 and 2. Study 2 was designed to directly compare Flanking with Interposing trials during data analyses (see Figure 29). In both Flanking and Interposing trials, the separation between visual stimuli (i.e., the visual SOA) was ±25, ±50, ±100, ±150, ±200, ±250, or ±300ms (a negative SOA indicates the lower visual stimulus appeared first while a positive SOA indicates that the upper visual stimulus appeared first). In all trials, the audiovisual SOA was ±20, ±35, ±50, ±65, ±80, ±95, or ±125 (a negative SOA indicates an auditory-leading pair while a positive SOA indicates a visual-leading pair).

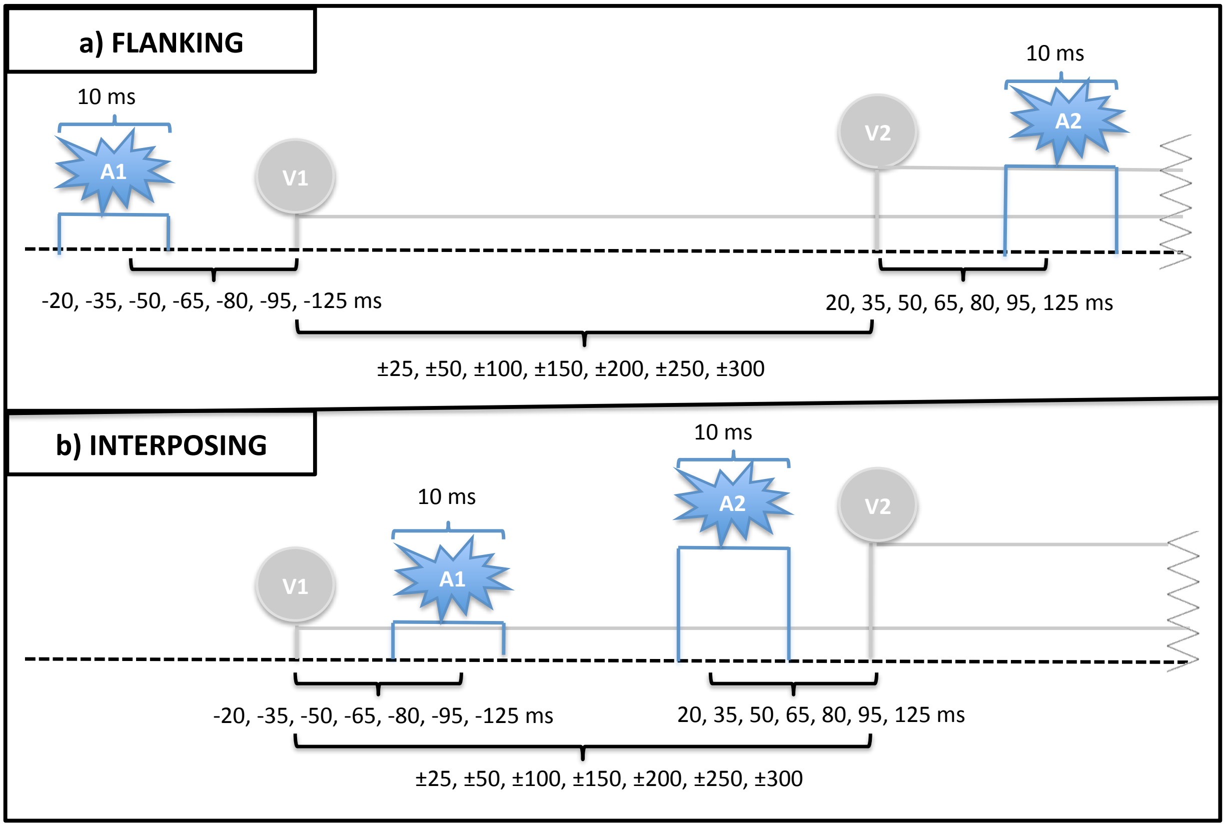


Figure 29. Stimulus timings for Study 3.

Participants. Forty-one people participated in Study 3 (M = 19, F =22). Participants were recruited from the Georgia Tech undergraduate participant pool and ranged from 18-25 years old (mean = 19. 9 years). Eight participants were excluded (M = 3, F = 5) in cases when the participant reported being left handed (1), being unable to detect both visual and both auditory stimuli at equal intensity (4), or described the main task incorrectly in the post-study survey (3).

6.5.2 Performance Results: Study 3

Baseline PSS. The average baseline SOA required for Study 3 participants to perceive audiovisual fusion was 91.1 ms (SD = 18.2 ms). The average baseline SOA required for Study 3 participants to perceive visual fusion was 71.2 ms (SD = 14.2 ms).

Main task. An initial inspection of RT performance data revealed a right-skewed distribution as is common for RT data. To correct for this, a log transformation was performed on RT data. The distribution of LogRT is a closer approximation of the normal curve (see Figure 30). The transformed data was used for all further analyses.

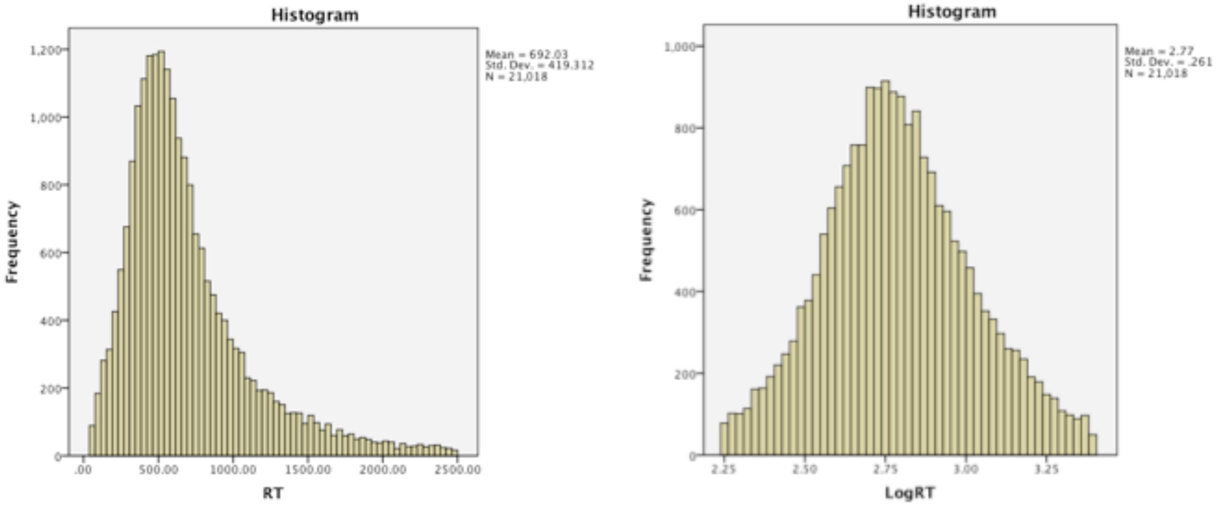


Figure 30. A log transform was performed on Study 3 RT data in order to correct for an initial right skew of the RT distribution.

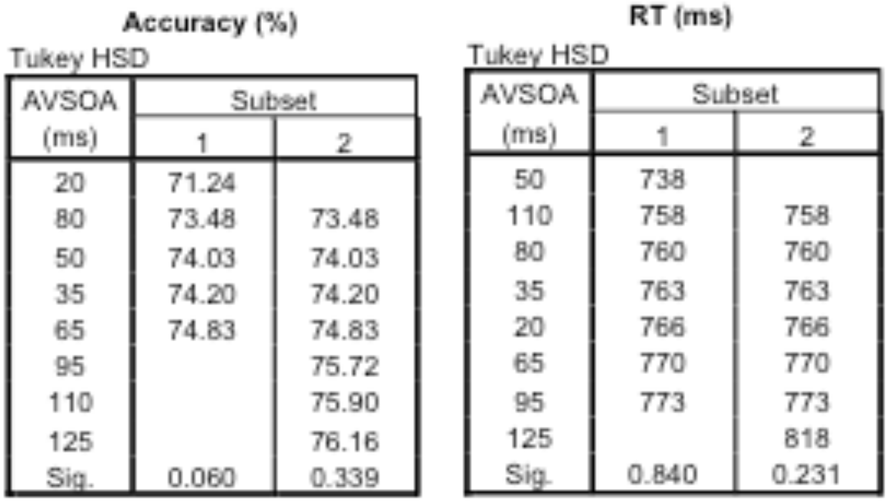
A MANOVA was performed using Accuracy and LogRT as the independent factors with Trial Type (Flanking, Interposing), Congruency (Congruent, Incongruent), AV-SOA (±20, ±35, ±50, ±65, ±80, ±95, or ±125), and Visual SOA (±25, ±50, ±100, ±150, ±200, ±250, or ±300ms) as the dependent factors. Significant main effects were revealed for all dependent variables: Trial Type Wilks' Λ = .998, F(2, 19903) = 18.902, *p* < .001; Congruency Wilks' Λ =.997, F(2, 19903) = 30.433, *p* < .001; AV-SOA Wilks' Λ = .998, F(14, 39808) = 3.467, *p* < .001; and Visual SOA Wilks' Λ =.969 F(12, 39808) =51.891, *p* < .001. There were also two significant interactions: Trial Type x Visual SOA Wilks' Λ = .998, F(10, 39808) = 3.601, *p* < .001 and Congruency x

Visual SOA Wilks' Λ = .998, F(12, 39808) = 3.940, $p < .001$. Results from the tests of between-subjects effects are reported below (see Appendix F).

Congruency. As was the case in studies 1-2, synesthetic congruency affected accuracy in the predicted direction in Study 3. In Congruent trials, participant accuracy averaged 76.8% compared to 71.6% accuracy in Incongruent trials F(1, 19904) = 57.066, $p < .001$. There was also a significant difference in RT for Congruent (685 ms) versus Incongruent (699 ms) trials F(1, 19904) = 9.760, $p = .002$.

AV-SOA. The asynchrony between auditory and visual signals (i.e., AV-SOA) did not significantly affect accuracy performance, however, tests of between-subjects effects do reveal a significant difference for RT performance across AV-SOA. Post-hoc analyses reveal that the shortest asynchrony (20 ms) resulted in significantly lower performance (71.2%) than the three longest asynchronies (95, 110, and 125 ms; 75.7%, 75.9, and 76.2%, respectively) at the $p = .002$ level for each case. The AV-SOA affected RT performance F(7, 19904) = 4.932, $p < .001$. Reaction times were significantly faster at an AV-SOA of 50 ms (738 ms) than at an AV-SOA of 125ms (818 ms) (see Table 3). While there is no significant interaction of Trial Type x AV-SOA, the RT results for Flanking and Interposing trials across AV-SOA are presented in Figure 31. RT results for Flanking and Interposing trials across AV-SOA.

Table 3. Study 3 AV-SOA homogeneous subgroups.



Accuracy (%)

Tukey HSD

AVSOA (ms)	Subset 1	Subset 2
20	71.24	
80	73.48	73.48
50	74.03	74.03
35	74.20	74.20
65	74.83	74.83
95		75.72
110		75.90
125		76.16
Sig.	0.060	0.339

RT (ms)

Tukey HSD

AVSOA (ms)	Subset 1	Subset 2
50	738	
110	758	758
80	760	760
35	763	763
20	766	766
65	770	770
95	773	773
125		818
Sig.	0.840	0.231

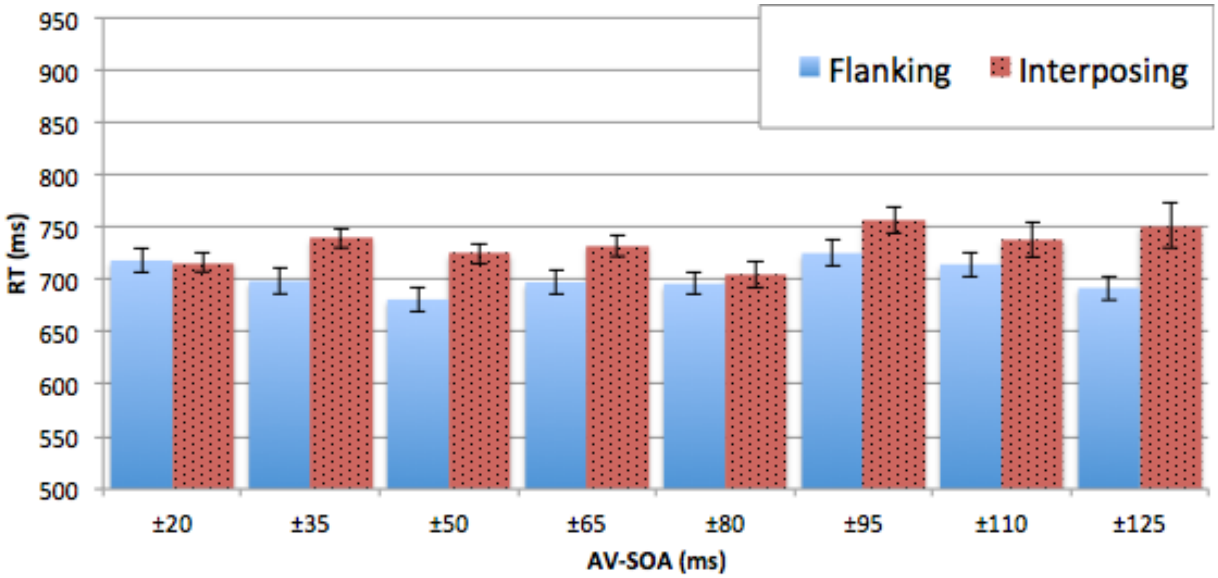
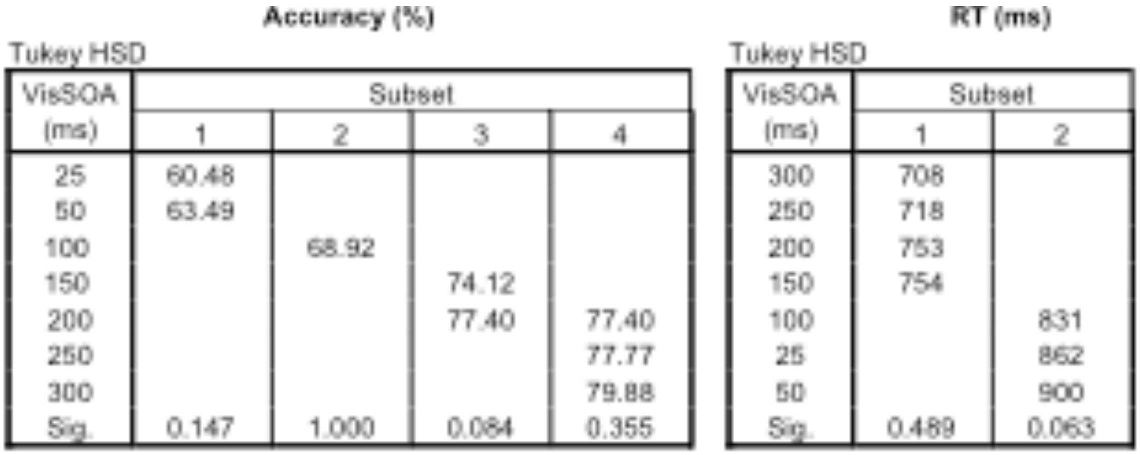


Figure 31. RT results for Flanking and Interposing trials across AV-SOA.

Visual SOA. The asynchrony between the pair of visual target signals did affected accuracy and RT performance at the $p < .001$ level. Post-hoc analyses reveal that accuracy and RT performance were best at longer Visual SOAs (see Table 4).

Table 4. Study 3 Visual SOA homogenous subgroups.



Accuracy (%)

Tukey HSD

VisSOA (ms)	Subset 1	Subset 2	Subset 3	Subset 4
25	60.48			
50	63.49			
100		68.92		
150			74.12	
200			77.40	77.40
250				77.77
300				79.88
Sig.	0.147	1.000	0.084	0.355

RT (ms)

Tukey HSD

VisSOA (ms)	Subset 1	Subset 2
300	708	
250	718	
200	753	
150	754	
100		831
25		862
50		900
Sig.	0.489	0.063

Trial type. Stimulus order did make a significant difference in performance in the predicted direction. In trials where auditory stimuli flanked the pair of visual targets, TOJ accuracy was 74.4% compared with 73.9% in trials where the auditory stimuli interposed the pair of visual targets F(1, 19904) = 36.455, p < .001. RT data showed the same pattern of improved performance for Flanking (678 ms) versus Interposing (706 ms) trials F(1, 19904) = 4.386, p = .036.

Trial Type x Visual SOA. This interaction was driven by RT performance F(5, 19904) = 6.871, p < .001 (see Figure 32). Reaction time was significantly faster in Flanking trials at Visual SOAs of 50, 100, 200, and 250ms, but significantly slower at a Visual SOA of 300ms.

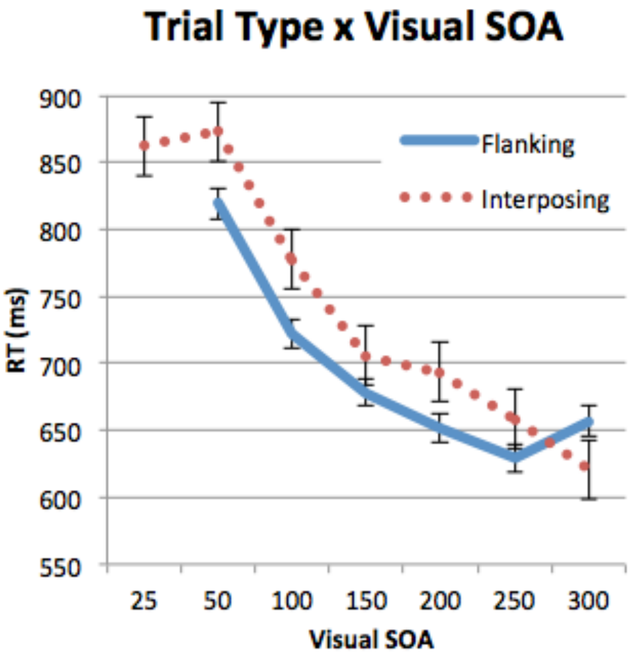


Figure 32. Interaction between Trial Type and Visual SOA for RT performance.

Congruency x Visual SOA. This interaction was driven by Accuracy performance F(6, 19904) = 5.971, *p* < .001 (see Figure 33). Accuracy performance was significantly higher at Visual SOAs of 25, 50, 200, 250, and 300 ms.

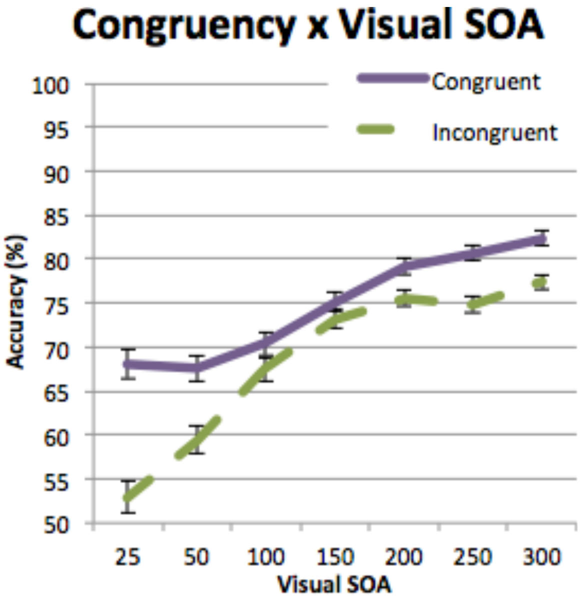


Figure 33. Interaction between Congruency and Visual SOA.

6.5.3 Psychometric Results: Study 3

In order to investigate whether the presence of the task-irrelevant auditory signals shifted the point of subjective simultaneity for the visual signals (Visual PSS), the proportion of responses indicating that the upper stimulus appeared first (“top first”) was calculated across all levels of Visual SOA. From these data, psychometric functions were calculated and compared across Trial Type. Psychometric functions were calculated using Palamedes, a downloadable collection of Matlab routines developed specifically for psychophysical data analysis (Prins & Kingdom, 2009). Cumulative Gaussian functions were fitted for for Flanking and Interposing trials, respectively (see Figure 34). The Visual PSS is calculated as the threshold of the psychometric function, or the asynchrony where 50% of “top first” responses were made (i.e., by taking the inverse of the Gaussian function at a value of 0.5). For Flanking trials, the Visual PSS was 19.06 ms compared with 52.79 ms for Interposing trials (see Figure 34). A comparison between the models by way of the likelihood ratio test indicates that there was a significant difference between the Flanking and Interposing psychometric functions $p(\chi^2) < .001$.

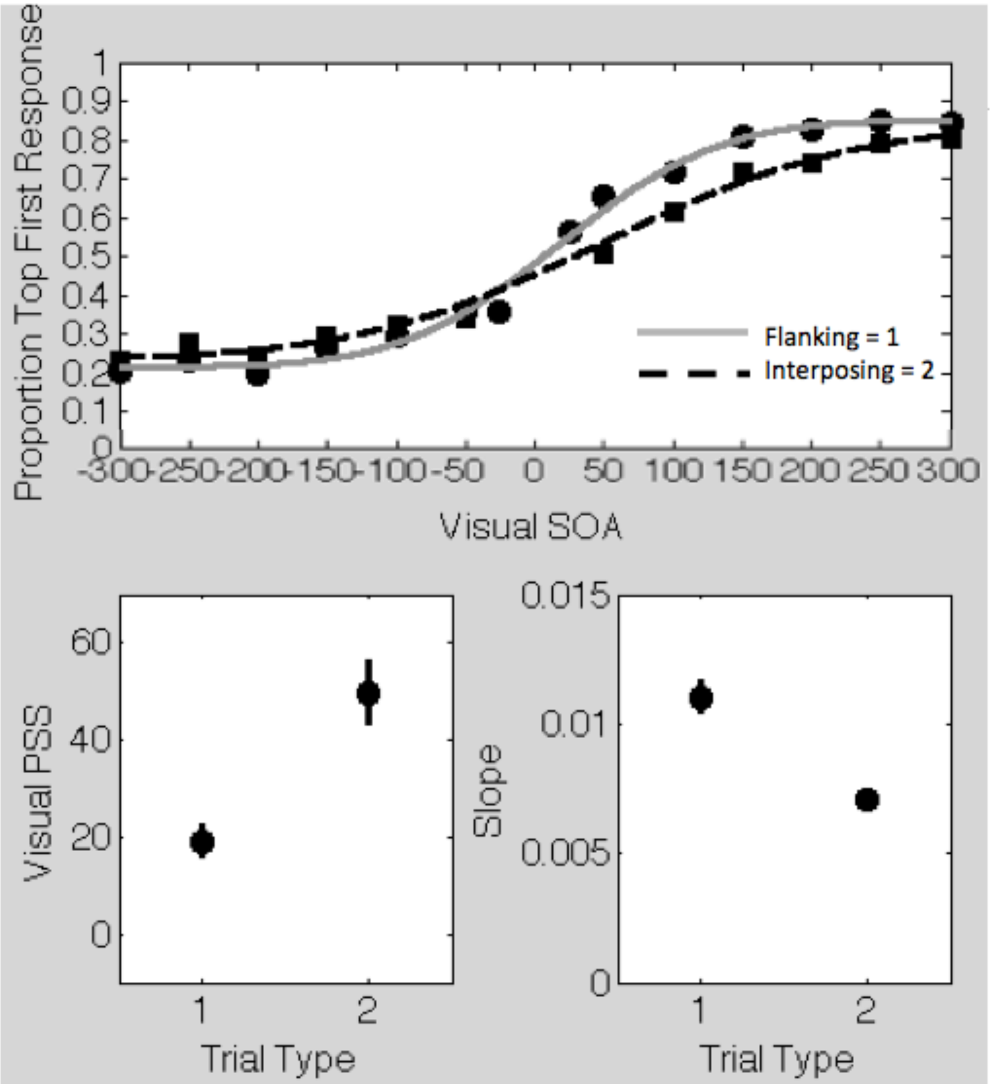


Figure 34. Psychometric functions fitted with cumulative Gaussian functions for Flanking and Interposing trials, respectively. The lower plots compare the parameter values for the fitted functions.

6.5.4 Discussion: Study 3

The results for Congruency and Visual SOA were similar for Study 3 to the results from studies 1-2. Here, synesthetic correspondence significantly affected both accuracy and RT performance in the predicted direction (i.e., performance improves when the auditory frequency was congruent with visual position). Once again and not surprisingly, Visual SOA results show increased performance with increasing separation between the two visual signals. Given that the main task is a TOJ task, it is expected that performance should improve with increasing temporal separation of the target stimuli. This pattern of results is interpreted as assurance that participants remained on-task and attempted to complete main task trials as instructed.

Though there were significant differences in Accuracy across AV-SOA in studies 1-2, it is not surprising that Accuracy remained constant across AV-SOA in Study 3 because the range of AV-SOAs is condensed and the intervals between AV-SOAs are much smaller compared to those used in studies 1-2. Also, the post-hoc results show a significant difference between the shortest and three longest AV-SOAs, which demonstrates a similar trend to the results from studies 1-2. The RT results are unexpected because they indicate that performance is significantly faster at an audiovisual separation of 50 ms versus 125 ms. Therefore, the pattern of results across AV-SOA may reflect a speed-accuracy tradeoff in this case. However, though there is no significant interaction of Trial Type x AV-SOA, splitting the results across trial type (see Figure 31) illustrates that the main effect of AV-SOA for RT is driven by quicker responses at an audiovisual separation of 50 ms for Flanking trials compared with slower responses at an audiovisual separation of 125 ms for Interposing trials. Given that performance for Interposing trials is expected to be worse than for Flanking trials, this specific pattern of results is not problematic with the overall findings for the current program of research.

As previous investigations of temporal ventriloquism have found, performance for Flanking trials exceeded performance for Interposing trials in the current study. The interpretation is that the task-irrelevant auditory signals aided in the visual TOJ task by pulling the perceived temporal occurrence of the visual signals further apart during Flanking trials. Conversely, in Interposing trials, the irrelevant auditory signals acted to pull the perceived temporal occurrence of the visual signals closer together, thus making the visual TOJ task more difficult. The psychometric data support this interpretation by revealing a shorter Visual PSS for Flanking thank for Interposing trials.

6.7 Study 4: A Meaningful Difference in Empirical Outcomes

Study 4 was designed to demonstrate that failing to consider synesthetic correspondence as a potential confound is an oversight that makes a meaningful difference in empirical outcomes. This is an important point that has implications for 1) the interpretation of existing data, 2) the design of future studies, and 3) a more complete development of perceptual theory. For illustrative purposes, one study published by an influential researcher in the field of multimodal perception was selected for examination under the lens of the current dissertation.

Study 4 replicates and extends a study conducted by Ho and Spence (2006). The original study aimed to investigate whether auditory direction cues influence the distribution of covert visual attention. Ho and Spence utilized an orthogonal variation of the standard cuing paradigm with non-predictive central cues indicating the direction of a subsequent, lateral target. Participants were asked to complete a 'digit discrimination' task by indicating whether a laterally presented visual target was either a '6' or a '9' without shifting gaze from central fixation (i.e., by shifting covert attention only). The researchers emphasize that the task was orthogonal because their direction cues were independent of the digit discrimination. The implication here is that the stimulus and response are separable variables (i.e., that the variation of one does not affect classification of the other). Contrary to this, it is suggested here that the lateral position is not separable from the variable of numeric magnitude, due to synesthetic correspondence between the two. Evidence has shown that individuals demonstrate a mental model where lower magnitudes correspond with the left region of space while higher magnitudes correspond with the right region of space. These numerical-spatial associations are expounded in investigations of the ‘mental number line’ (e.g., Gut, Szunska, Wasilewska, & Jaskowski, 2012) and the ‘spatial numerical associations of response codes’ (SNARC) effect (Dehaene, Bossini, & Giraux, 1993). The former relates to the underlying mental representation and the latter to consequent response behaviors. Study 4 is designed to demonstrate that numerical-spatial associations act as confounds at both the cognitive and behavioral level in the widely utilized digit-discrimination task.

6.7.1 Method: Study 4

Participants. Twenty-two participants participated in Study 4 (M = 13, F = 11). Participants were recruited from the Georgia Tech undergraduate participant pool and ranged from 18-34 years old (mean = 20.5 years). Two participants’ data were excluded (M = 1, F = 1) from analysis because they reported being left-handed.

Materials. Study 4 included two paper forms and a computerized task. The consent form was approved by the Georgia Institute of Technology Institutional Review Board under standard protocol (see Appendix G). Participans also completed a post-task survey, which was used to collect participant background information (e.g., age, handedness) and self-assessment of computer-task performance (see Appendix G). The computer task was built using the E-Prime 2.0 application suite (Schneider, Eschman, & Zuccolotto, 2002) and was presented on a 22-inch Dell monitor.

Design. Study 4 included four conditions: Digit (6 and 9), Masked Digit, Cross (x and +), and Masked Cross. The Digit conditions replicated the procedure employed by Ho and Spence. The Cross conditions were included as a control designed to demonstrate the performance bias arising from a numerical-spatial confound in the original study. Each trial began with the presentation of a central fixation point (0.76°) lasting for 760 ms followed by a 400 ms direction cue. Visual direction cues were presented as the words, ‘LEFT’ or ‘RIGHT’ presented centrally. Visual cues (2.3°) and target symbols (1.15°) were presented in Times New Roman font at a font size of 48. The target symbols followed cue onset at an asynchrony of either 50, 100, 200, 400, or 600 ms located 142 mm to the left or right of the central fixation point. The visual masking stimulus consisted of a cosine grating directly overlaid on the target digit (see Figure 35). In all cases the direction cue was non-predictive, meaning cues validly indicated the subsequent target location in half of the trials.

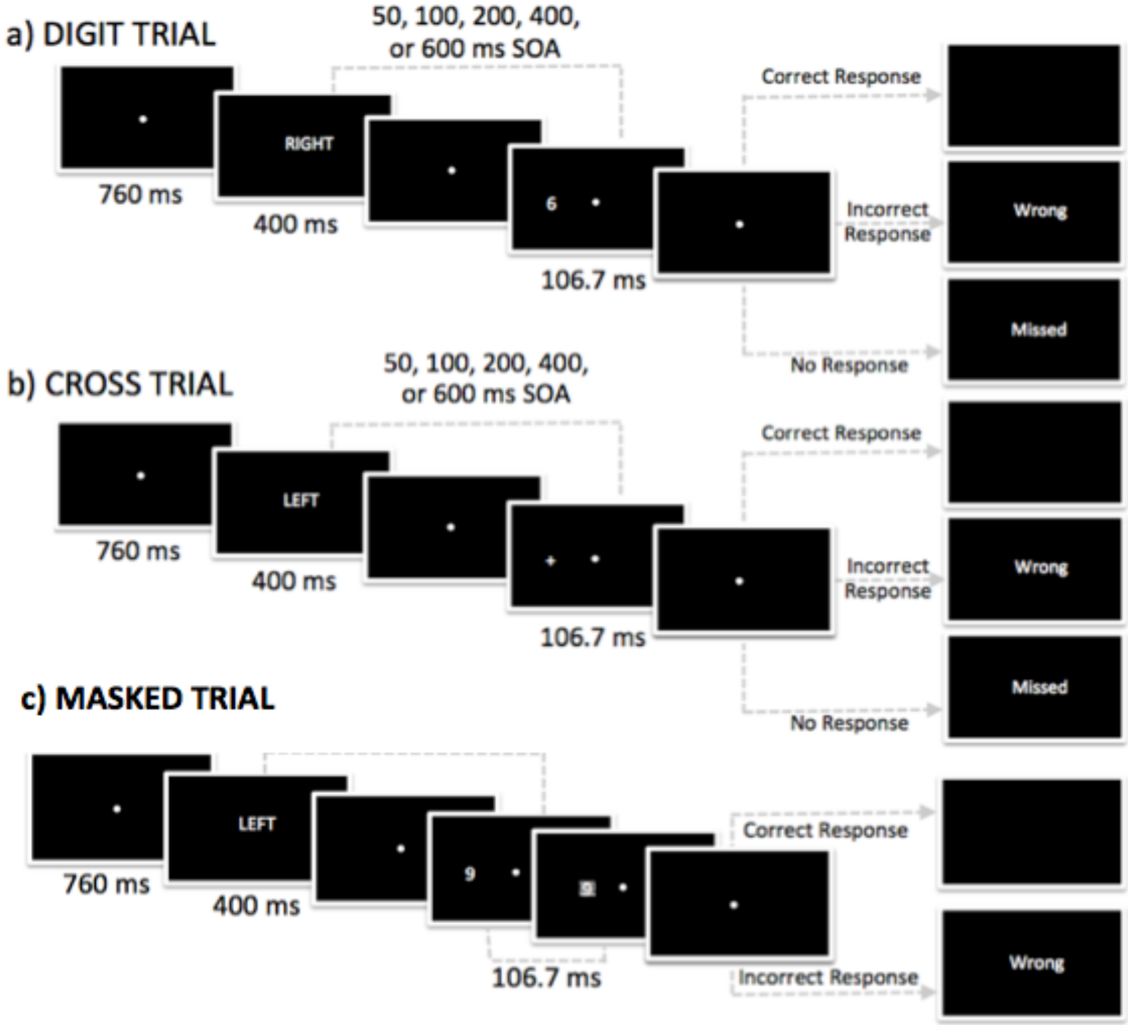
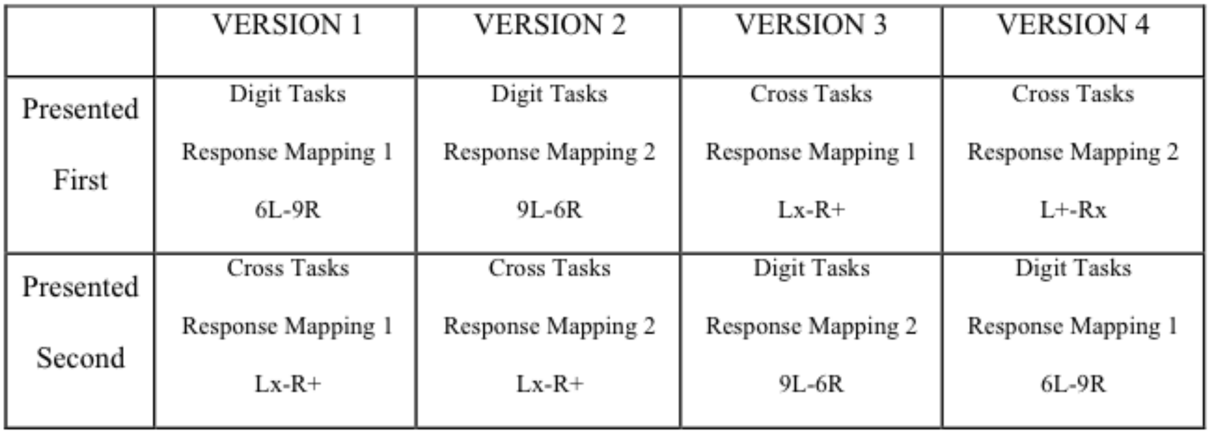


Figure 35. Study 4 Procedure for a) Digit trials, b) Cross trials, and c) Masked trials. Note that panel c happens to depict a Masked Digit trial, but Masked Cross trials followed the same stimulus order and timings.

There were two dependent factors included in Study 4 that allowed for investigation of synesthetic correspondence. These were Response Mapping and Target Presentation. Ho and Spence asked their participants to either make a left key press to indicate that a 6 was detected and a right key press to indicate a 9 was detected or vice versa. In the current study, the former (L6-R9) is labeled as a congruent Response Mapping and the latter (L9-R6) is labeled as an incongruent Response Mapping. If synesthetic congruency affected behavioral outcomes in this experimental set-up, then it is expected that Response Mapping affected performance during Digit trials. Regardless of the assigned Response Mapping, synesthetic correspondence can seperately affect perception of targets. Target Presentation was labeled as congruent in Digit trials where a ‘6’ was presented on the left side of the screen or when a ‘9’ was presented on the right side of the screen (6L-9R). Target Presentation is labeled as incongruent in Digit trials where a ‘9’ was presented on the left side of the screen or when a ‘6’ was presented on the right side of the screen (9L-6R). If synesthetic congruency affected perceptual outcomes in this experimental set-up, then it is expected that Target Presentation affected performance during Digit trials. In Cross trials, Response Mapping (Lx-R+, L+,Rx) and Target Presentation (xL-+R, +L-xR) were also examined, but not labeled by congruency. This is because in Cross trials, it is expected that there would be no association between targets and position because there is no known synesthetic mapping between these dimensions. Four versions of Study 4 were created to ensure that the assigned Reesponse Mapping for both Digit and Cross trials were counter-balanced and to ensure that the order of symbol presentation was counter-balanced (see Table 5).

Table 5. Four versions of Study 4 counter-balanced across Response Mapping and Symbol Order (Digit trials first versus Cross trials first).



	VERSION 1	VERSION 2	VERSION 3	VERSION 4
Presented First	Digit Tasks Response Mapping 1 6L-9R	Digit Tasks Response Mapping 2 9L-6R	Cross Tasks Response Mapping 1 Lx-R+	Cross Tasks Response Mapping 2 L+-Rx
Presented Second	Cross Tasks Response Mapping 1 Lx-R+	Cross Tasks Response Mapping 2 Lx-R+	Digit Tasks Response Mapping 2 9L-6R	Digit Tasks Response Mapping 1 6L-9R

Procedure. Participants were seated in a sound attenuated room 21 inches in front of a 22-inch Dell monitor. A chin-rest was used to ensure viewing distance was maintained throughout the computer task. Consent was obtained prior to beginning the computer task. Participants then began the computer task with a block of Unmasked trials (either Digits or Crosses, depending on the assigned study version, see Table 5). The block of Unmasked trials instructed the participant regarding which key to press to indicate which response (i.e., the assigned ‘Response Mapping’) and to focus on responding accurately. Each block began with ten practice trials followed by 125 experimental trials. Feedback was given as previously illustrated in Figure 35. At the end of each block, participants were encouraged to take a break if needed. The second block was always a set of Masked trials. Participants were reminded of their assigned Response Mapping and instructed to focus on responding quickly. The third and fourth blocks followed the same procedure, but presented whichever symbol was not already presented in the first two blocks along with its respective Response Mapping. After completing the computer task, participants were asked to complete the post-task survey. Finally, participants were debriefed and thanked for their participation.

6.7.2 Results and Discussion: Study 4

An initial inspection of RT performance data revealed a right-skewed distribution as is common for RT data. To correct for this, a square root transformation was performed on RT data. The distribution of LogRT is a closer approximation of the normal curve (see Figure 36). The transformed data was used for all further analyses.

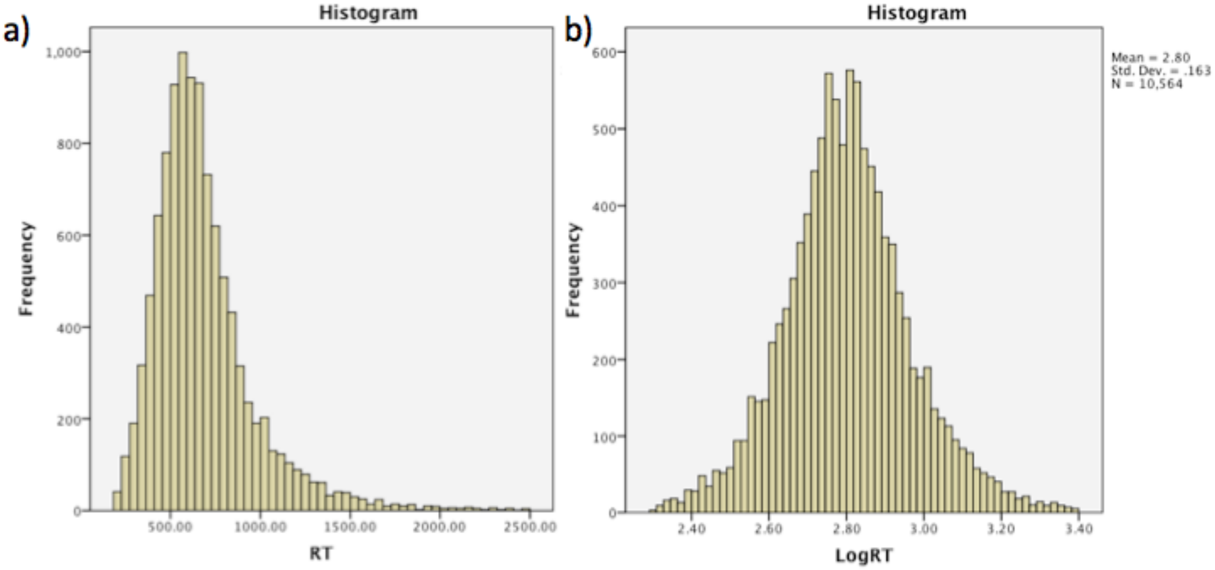


Figure 36. RT versus log transform of RT distributions. A log transform (panel b) was performed in order to correct an initial right skew (panel a).

All trials. A MANOVA was performed using Accuracy and LogRT as the independent factors with Cue Validity (Valid, Invalid), Target Mask (Masked, Unmasked), Cue-Target SOA (50, 100, 200, 400, 600), Symbol (Digits, Crosses), and Symbol Order (Digit Trials First, Cross Trials First) as the dependent factors. Significant main effects were revealed for Target Mask, Cue-Target SOA, and Symbol Order. Significant interactions were revealed for and Cue-Target SOA x Target Mask and Symbol x Symbol Order. Target Mask Wilks' Λ =.956, F(2,10483) = 240.277, p < .001; Cue-Target SOA Wilks' Λ = .969, F(8, 10483) = 41.122, *p* < .001; Symbol

Order Wilks' Λ = .983, F(2,10483) = 93.034, $p < .001$; Symbol x Symbol Order Wilks' Λ = .987, F(2, 10483) = 70.533, $p < .001$; and Cue-Target SOA x Target Mask Wilks' Λ = .938, F(8, 20968) = 83.293, $p < .001$ (see Appendix H).

Target Mask. Tests of between-subjects effects reveal that the main effect of Target Mask was significant for both Accuracy and RT performance. Accuracy was 87.8% for Unmasked trials versus 73.2% for Masked trials F(1, 10484) = 449.188, $p < .001$. RT was 651 ms for Unmasked trials and 624 ms for Masked trials F(1, 10484) = 20.396, $p < .001$. This speed-accuracy tradeoff was expected and interpreted as assurance that participants followed the instructions, which emphasized accuracy for Unmasked trials and speed for Masked trials.

Cue-target SOA. Test of between-subjects effects reveal that RT drove this significant main effect. The main effect of Cue-Target SOA shows a trend where RT increases as the separation between the cue and target increases F(4, 10484) = 80.960, $p < .001$. This initially seemed to counter the well-established alerting effect where reaction time decreases as Cue-Target SOA increases. However, the significant interaction between Cue-Target SOA and Target Mask sheds some light on this initially puzzling finding. Unmasked trials did follow the expected pattern with decreasing RT values at increasing Cue-Target SOAs (see Figure 37). Masked trials, on the other hand, drove the overall opposite trend. This pattern might have been caused by an inhibition of return in masked trials where attention was inhibited from the target location during initial unmasked presentation, only to return after the onset of the visual mask. Post-hoc analyses reveal that for Unmasked trials, RT for a Cue-Target SOA of 50 ms is significantly slower than for every other Cue-Target SOA. For masked trials, results fall into three homogenous subgroups: 1) 50 and 100 ms, 2) 200 ms, and 3) 400 and 600 ms (see Figure 37).

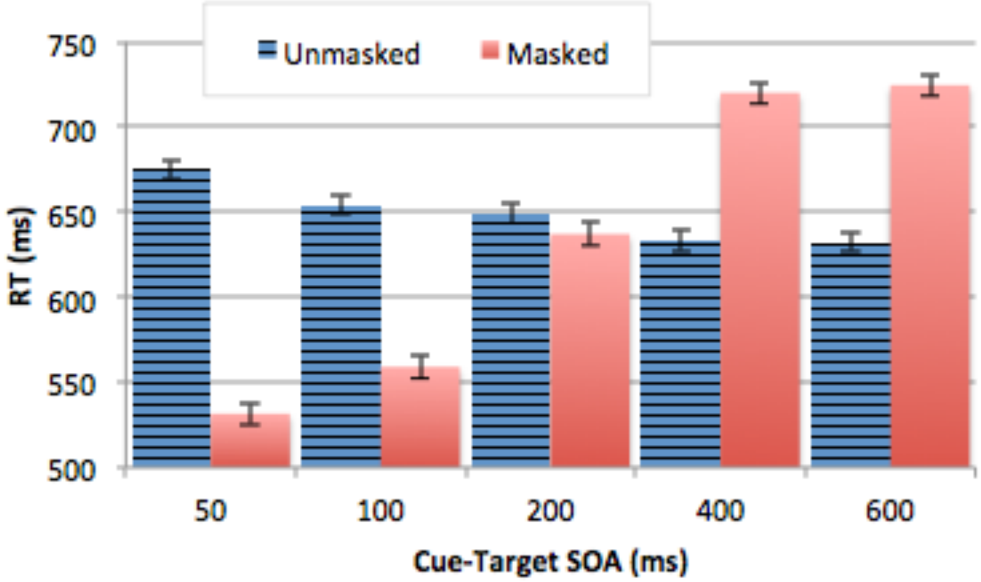


Figure 37. Interaction between Cue-Target SOA and Target Mask. Unmasked trials reveal the expected alerting effect while Masked trials do not.

Symbol order. Overall, participants responded more quickly, but less accurately in versions of the study where Digit trials were presented prior to Cross trials (see Table 5). Accuracy for participants who completed Digit trials first was 77.84% compared to 83.18% accuracy for participants who completed Cross trials first: F(1, 10484) = 6.735, $p = .009$ (see Figure 38). Corresponding RTs are 511.15 ms and 702.76 ms, respectively: F(1, 10484) = 175.359, $p < .001$. There appears to be a speed-accuracy tradeoff dependent on the order of condition presentation. However, there is also a significant interaction between Symbol and Symbol Order driven by RT performance: F(1, 10484) = 136.505, $p < .001$ (see Figure 39). In versions of the study where Digit trials were presented first, participants responded significantly faster to target crosses (598 ms) than to target digits (624 ms). Conversely, in versions of the study where Cross trials were presented first, participants responded significantly faster to target digits (662 ms) than to target crosses (743 ms). In addition to the speed-accuracy tradeoff revealed by the significant main effect of Symbol Order, the interaction between Symbol and Symbol Order demonstrates what appears to be a learning effect where participants responded more quickly (though no more accurately) to the second trial type presented (see Figure 39).

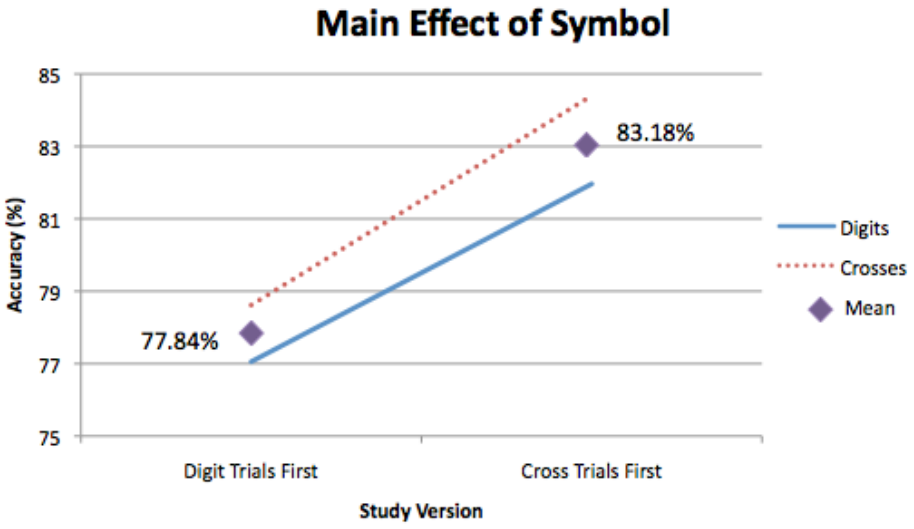


Figure 38. *Main Effect of Symbol. Participants responded more accurately in study versions when Cross trials were presented first than when Digit trials were presented first.*

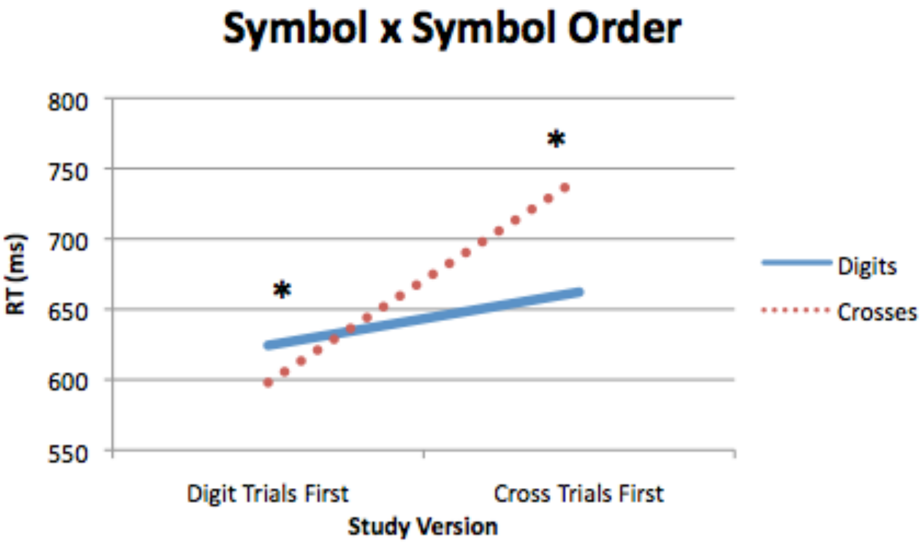


Figure 39. Interaction between Symbol and Symbol Order. Overall, RT was faster when Digit trials were presented first than when Cross trials were presented first (main effect of Symbol). However, there was also an interaction caused by a learning effect where participants became faster on the second trial type presented.

Cross trials. A MANOVA was performed on Cross trials using Accuracy and LogRT as the independent factors with Response Mapping (Lx-R+, L+-Rx) and Target Presentation (xL-+R, +L-xR), as the dependent factors. The analysis revealed no significant effects of either Response Mapping, Target Presentation, or any interaction between them.

Digit trials. A MANOVA was performed on Digit trials using Accuracy and LogRT as the independent factors with Response Mapping (L6-R9 = Congruent, L9-R6 = Incongruent) and Target Presentation (6L-9R = Congruent, 9L-6R = Incongruent) as the dependent factors. Significant main effects were revealed for both dependent variables: Response Mapping F(2, 5215) = 146.452, *p* < .001; and Target Presentation F(2, 5215) = 22.153, *p* < .001. There was also a significant interaction between the independent factors: Response Mapping x Target Presentation F(2, 5215) = 26.266, *p* = .034 (see Appendix I).

The main effect of Response Mapping reveals that performance was significantly better when the digit response side mapping was congruent (left key response to indicate 6, right key response to indicated 9) than when incongruent (right key response to indicate 9, left key response to indicate 6). When Response Mapping was congruent, accuracy was 82.1% and RT was 592 ms. When Response Mapping was incongruent, accuracy was 77.3% and RT was 686 ms. This finding is consistent with literature reporting behavioral consequences of the SNARC effect. Namely, that lower magnitudes correspond with left-hand responses while higher magnitudes correspond with right-hand responses (Dehaene, et al., 1993).

The main effect of Target Presentation reveals that performance was significantly better when the presentation side of the target was congruent (6 presented on the left, 9 presented on the right) than when incongruent (9 presented on left, 6 presented on right). When Target Presentation was congruent, accuracy was 81.5% and RT was 627 ms. When Target Presentation was incongruent, accuracy was 77.5% and RT is 660 ms). This finding is consistent with literature on the MNL.

There was also an interaction between Response Mapping and Target Presentation. The interaction reveals that the performance improvement from incongruent to congruent cases was emphasized when both Response Mapping and Target Presentation were congruent (see Figure 40).

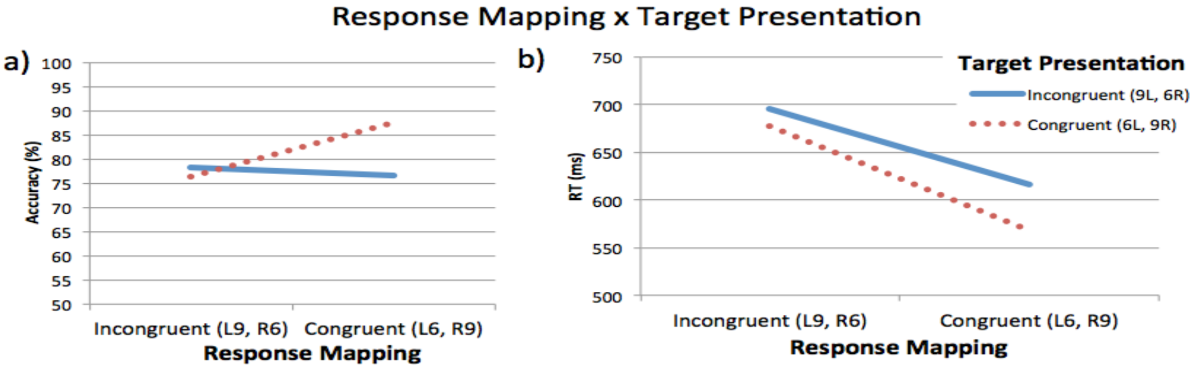


Figure 40. *Interaction between Response Mapping and Target Presentation congruency. Performance is most enhanced when both factors are congruent, respectively.*

6.7.4 Implications: Study 4

Study 4 demonstrated that the general lack of understanding that synesthetic correspondence serves as an amodal binding agent can lead to an incomplete design of empirical studies of multimodal perception. Though Study 4 highlights one study from one prominent researcher for illustrative purposes, this study is far from an isolated instance of failure to consider synesthetic correspondence in empirical design. The digit-discrimination task itself is a prototypical task for investigations of attention and it is one of many with the potential for synesthetically driven confounds. Such confounds have implications for 1) the interpretation of existing data, 2) the design of future studies, and 3) a more complete development of perceptual theory.

7. CONCLUDING REMARKS

The program of research presented here supports the hypothesis that synesthetic correspondence is an amodal binding property that serves to bind audiovisual signals, even when one of those signals is task-irrelevant. The results also support that auditory and visual inputs are integrated by a flexible weighting mechanism that can be manipulated by changes in task demands. Here, a slight instruction shift from a more spatial to a more temporal emphasis resulted in a larger effect of auditory driving (i.e., a heavier weighting of auditory inputs). These results fit well with the proposed spatiotemporal continuum model. Because it is a somewhat abstract concept, the degree of dimensional overlap can be difficult to manipulate. However, this was accomplished in this program of research by comparing the stronger pitch-vertical association to the comparatively weaker pitch-lateral association. The results support that synesthetically driven binding was stronger when there was more dimensional overlap. Finally, the program of research demonstrates that synesthetic correspondence presents a confound that has affected multimodal perception research. In the following section, potential applications for these findings are discussed.

Discoveries involving audiovisual integration such as those presented in the current program of research have implications for the design of multimodal interfaces. Modern interfaces have become more complex and are increasingly needed in dynamic environments. Operators may be faced with a vast array of displays and controls that must be monitored in order for critical information to be extracted and acted upon. Though traditionally such interfaces have been largely visual, there is great interest in incorporating more auditory components. Sounds can be used in stand-alone auditory displays, to augment visual displays, or dynamically depending on task-demands. There are instances when sound is the preferred display modality due to the superiority of audition for specific tasks. For example, in tasks which involve monitoring one or multiple streams of data with quickly changing temporal characteristics, auditory displays are likely more appropriate (Kramer, Walker, Bonebright, Cook, & Flowers, 2010). Additionally, auditory stimuli are inherently more alerting (Posner, et al., 1976), and are therefore less likely to be missed. This combined with the fact that auditory displays do not require the operator to be oriented in a particular direction make them appropriate for monitoring and alarm applications, particularly when the eyes are already engaged in another task (Kramer, et al., 2010).

Currently, applied research of auditory cue augmentation is being investigated across a number of settings from ground combat vehicles, to aircraft cockpits, to healthcare facilities. These may use traditional or more novel interfaces such as speech input, heads-up-displays, or multichannel communication systems. Multimodal displays can be used to deliver redundant information, which is particularly desirable in cases where the consequences of a miss far outweigh those of a false alarm. Oskarsson and colleagues (2012) provide evidence to suggest that redundant information leads to particularly higher performance in high-stress, time-critical, dynamic situations. A number of studies have demonstrated that the presence of an auditory cue at the location of a visual target significantly decreases the time required to locate that target. This is the known as the auditory spatial facilitation (ASF) of visual search. Vu, Strybel, & Proctor (2006) emphasize that at any given moment, there is a large amount of information falling outside of the central visual field in a number of modern displays. They demonstrated that the auditory cues do facilitate visual search times even when the cues are somewhat spatially displaced. Though both vertical and horizontal displacement was better than an uninformative central auditory cue, vertical displacement was more disruptive than horizontal displacement when compared to a non-displaced auditory cue. Similarly, McIntire et al. (2010) compared spatially predictive auditory cues in both static and dynamic environments and found that these were most beneficial when displacement was along the horizontal plane.

The performance benefits related to data redundancy and ASF occur because the operator perceives the auditory and visual inputs as valid co-occurrences. Broadly speaking, similar performance enhancement could be realized by way of synesthetically driven audiovisual binding. The specific pitch-position synesthetic correspondence investigated in the current program of research could be applied to the design of both spatially predicative and non-predictive auditory displays. Spatially predictive auditory displays (i.e., displays utilizing 3D or spatialized audio) consistently yield the greatest performance enhancement. In this case, synesthetic correspondence could provide additional redundant information regarding where the operator should attend. Such redundancies have proven particularly beneficial in high workload situations (Oskarsson, et al., 2012). Unfortunately, 3D audio is not always feasible due to any of a number of constraints (e.g., financial, logistical). The performance cost of non-predictive auditory cues is exacerbated in when the operator is mobile, such as for head-mounted display wearers (Harrison, Thompson, & Sanderson, 2010). In cases where auditory cues are spatially displaced, synesthetic correspondence could be applied to compensate for shortcomings in facilitatory performance. For instance, vertically displaced auditory cues have been shown to be the less beneficial than horizontally displaced ones. Meanwhile, the pitch-vertical position mapping is shown to be a robust effect. Applying this synesthetic correspondence to the design of an auditory display with vertical displacement could enhance performance. The next phase of this work would be to replicate the current findings in an applied setting.

REFERENCES

Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J., & Bolton, P. (1996). Synaesthesia: Prevalence and familiarity. *Perception, 25*(9), 1073-1079.

Baron-Cohen, S., Harrison, J., Goldstein, L. H., & Wyke, M. (1993). Coloured speech perception: Is synaesthesia what happens when modularity breaks down? . *Perception, 22*(4), 419-426.

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics, 57*(8), 1151-1162.

Berger, T. D., Martelli, M., & Pelli, D. G. (2003). Flicker flutter: Is an illusory event as good as the real thing? *Journal of Vision, 3*(6), 406-412.

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology, 87*(2), 241-247.

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

Bertelson, P., & de Gelder, B. (2004). **The psychology of multimodal perception**. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141-177). New York: Oxford University Press.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics, 29*(6), 578-584.

Binda, P., Morrone, M. C., & Burr, D. C. (2009). Temporal auditory capture does not affect the time course of saccadic mislocalization of visual stimuli. *Journal of Vision, 10*(3), 1-13.

Boenke, L. T., Deliano, M., & Ohl, F. W. (2009). Stimulus duration influences peceived simultaneity in audiovisual temporal-order judgment. *Experimental Brain Research, 198*, 233-244.

Bower, T. G. (1971). The object in the world of the infant. *Scientific American, 225*, 30-38.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theorys of visual pattern amsking, saccadic suprression, and information processing. *Psychological Review, 83*(1), 1-36.

Brewster, D. (1839). *Letters on Natural Magic*. New York: Harper.

Brown, J. (2005). *Audiovisual asymmetries in speech and nonspeech audio.* University of Connecticut, Storrs, CT.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research, 198*(1), 49-57.

Canon, L. K. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology, 84*(1), 141-147.

Colavita, F. B. (1974). Human sensory dominance *Perception & Psychophysics, 16*(2), 409-412.

Colin, C., Radeau, M., Soquet, A., Dachy, B., & Deltenre, P. (2002). Electrophysiology of spatial scene analysis: The mismatch negativity (MMN) is sensitive to the ventriloquism illusion. *Clincal Neurophysiology, 113*, 207-215.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion, 14*(3), 289-311.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant behavior and Development, 9*, 133-150.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Experimental Psychology: General, 122*(3), 371-396.

Després, O., Candas, V., & Dufour, A. (2005). Auditory compensation in myopic humans: Involvement of bimaural, monaural, or echo cues? *Brain Research, 1041*, 56-65.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology, 92*, 53-78.

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

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Crossmodal spatial attention, 2*(7), 254-262.

Driver, J., & Spence, C. (2004). **Crossmodal spatial attention: Evidence from human performance**. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 179-220). New York: Oxford University Press.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*(2), 161-177.

Eimer, M. (2004). Electrophysiology of human corssmodal spatial attention. In C. Spence & J. Driver (Eds.), *Crossmodal Space and Crossmodal Attention*. New York: Oxford University Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433.

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

Fishkin, S. M., Pishkin, V., & Stahl, M. L. (1975). Factors involved in visual capture. *Perceptual and Motor Skills, 40*, 427-434.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attetnional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18*(4), 1030-1044.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interaction in the speeded classification of visual size. *Perception & Psychophysics, 68*(7), 1191-1203.

Galotti, K. M. (2004). **Paying attention.** *Cognitive Psychology: In and Out of the Laboratory* (3rd ed., pp. 86-128). Belmont, CA: Wadswoth Thomson Learning.

Garner, W. R. (1977). The effect of absolute size on the separability of the dimensions of size and brightness. *Bulletin of the Psychonomic Society, 9*(5), 380 - 382.

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *American Journal of Psychology, 72*(4), 521-528.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual intergation during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitve Neuroscience, 11*(5), 473-490.

Gibson, J. J. (1933). Adaptation, after-effect, and contrast in the perception of curved lines *Journal of Experimental Psychology, 16*(1), 2-31.

Goldsmith, M., & Yeari, M. (2003). Modulation of object-based attention by spatial fouc under endgoenous and exogenous orienting. *Journal of Experimental Psychology: Human Perception and Performance, 29*(5), 897-918.

Gondan, M., Vorberg, D., & Greenlee, M. W. (2007). Modality shift effects mimic multisensory interations: An event-related potential study. *Experimental Brain Research, 182*, 199-214.

Grossenbacher, P. G., & Lovelace, C. T. (2001). Mechanisms of synesthesia: Cognitive and physiological constraints. *Trends in Cognitive Sciences, 5*(1), 36-41.

Gut, M., Szunska, I., Wasilewska, M., & Jaskowski, P. (2012). Are low and high number magnitudes processed differently while resolving the conflict evoked by the SNARC effect? *International Journal of Psychophysiology, 85*, 7-16.

Harris, C. S. (1965). Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review, 72*(6), 419-444.

Harrison, W. J., Thompson, M. B., & Sanderson, P. M. (2010). Multisensory integration with a head-mounted display: Background visual motion and sound motion. *Human Factors, 52*(1), 78-91.

Held, R. (1965). Plasticity in sensory-motor systems. *Scientific American, 213*(5), 84-94.

Ho, C., & Spence, C. (2006). Verbal interface design: Do verbal directional cues automatically orinet visual spatial attention? *Computer in Human Behavior, 22*, 733-748.

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

Institute, A. N. S. (2004). Methods for manual pure-tone threshold audiometry (Vol. ANSI S2.21-2004).

James, W. (1983/1890). *The principles of psychology.* Cambridge, MA: Harvard University Press.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology, 37*, 43-75.

Jousmäki, V., & Hari, R. (1998). Parchment-skin illusion: Sound-biased touch. *Current Biology, 8*(6), R190.

King, A. J. (2004). **Development of multisensory spatial integration**. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 1-24). New York: Oxford University Press.

Kinsbourne, M. (1993). Orientational bias model of unilateral neglect: Evidence from attentional gradients within hemispace. In I. H. Robertson & J. C. Marshall (Eds.), *Unilateral Neglect: Clinical and Experimental Studies* (pp. 63-86). East Sussex, UK: Lawrence Erlbaum Associates Ltd.

Klein, R. M. (1994). Perceptual-motor expectancies interact with covert visual orienting under conditions of endogenous but not expgenous control. *Canadain Journal of Experimental Psychology, 48*(2), 167-181.

Klein, R. M., & Shore, D. I. (2000). Relations among modes of visual orienting. In S. Monsell & J. Driver (Eds.), *Attention and performance: XVIII. Control of cognitive processes,* (pp. 195-208). Cambridge, MA: MIT Press.

Koppen, C., Levitan, C. A., & Spence, C. (2009). A signal detection study of the Colavita visual dominacne effect. *Experimental Brain Research, 196*, 353-360.

Koppen, C., & Spence, C. (2007). Audiovisial asynchony modulates the Colavita visual dominance effect. *Brain Research, 1186*, 224-232.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility - A model and taxonomy. *Psychological Review, 97*(2), 253-270.

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., & Flowers, J. H. (2010). Sonification report: Status of the field and research agenda. from University of Nebraska-Lincoln: http://digitalcommons.unl.edu/psychfacpub/444

Ladavas, E., & Farne, A. (2004). **Neuropsychological evidence for multimodal representations of space near specific body parts.** In C. Spence & J. Driver (Eds.), *Crossmodal Space and Crossmodal Attention*. New York: Oxford University Press.

Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., & Parra, L. C. (2009). Lip-reading aids word recognition in moderate noise: A bayesian explanation using high-dimensional feature space. *PLoS, 4*(3), 1-14. Retrieved from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0004638

Macaluso, E., & Driver, J. (2004). Functional imaging of crossmodal spatial represtnations and crossmodal spatial attention. In C. Spence & J. Driver (Eds.), *Crossmodal Space and Crossmodal Attention.* New York: Oxford University Press.

Marks, L. E. (1982). Bright sneezes and dark coughs: Loud sunlight and soft moonlight. *Journalof Experimental Psychology, 8*(2), 177-193.

Marks, L. E. (2004). **Cross-modal interactions in speeded classification**. In G. Calvert, C. Spence & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 85-106). Cambridge, MA: The MIT Press.

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science, 10*(2), 61-65.

Massaro, D. W. (1999). Speechreading: Illusion or window into pattern recognition. *Trends in Cognitive Sciences, 3*(8), 310-317.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. T. (1996). Perception of asynchronous and conflicting visual and auditory speech. (Journal of the Acoustical Society of America), 1777-1786.

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

McIntire, J. P., Havig, P. R., Watamaniuk, S. N. J., & Gilkey, R. H. (2010). Visual search performance with 3-D auditory cues: Effects of motion, target location, and practice. *Human Factors, 52*(1), 41-53.

Melara, R. D. (1989). Similarity relations among synesthetic stimuli and their attributes. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 212-231.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General, 116*(4), 323-336.

Melera, R. D., & Marks, L. E. (1990). Dimensional interactions in language processing: Investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 539-554.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389-391.

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

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

Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tones. *Journal of Experimental Psychology, 66*(4), 347-352.

Myers, A. K., Cotton, B., & Hilp, H. A. (1981). Matching the rate of concurrent tone bursts and light flashes as a function of flash surround luminance. *Perception & Psychophysics, 90*(1), 33-38.

Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology, 7*(4), 480-494.

Oskarsson, P.-A., Eriksson, L., & Carlander, O. (2012). Enhanced perception and performance by multimodal threat cueing in simulated combat vehicle. *Human Factors, 54*(1), 122-137.

Over, R. (1966). An experimentally induced conflict between vision and proprioception. *British Journal of Psychology, 57*(3), 335-341.

Pasalar, S., Ro, T., & Beauchamp, M. S. (2010). TMS of posterior parietal cortex disrupts visual tactile multisensory integration. *European Journal of Neuroscience, 31*, 1783-1790.

Pashler, H. E. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Peelen, M. V., Heslenfeld, D. J., & Theeuwes, J. (2004). Endogenous and exogenous attention shifts are mediated by the same large-scale neural network. *NeuroImage, 22*, 822-830.

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception and Psychophysics, 6*, 203-205.

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

Posner, M. I., & Cohen, Y. (1984). **Components of visual orienting**. In H. Bourma & D. Bouwhuis (Eds.), *Attention and Performance* (Vol. 10). Hillsdale, NJ: Erlbaum.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of it origins and significance. *Psychological Review, 83*(2), 157-171.

Pratt, C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology, 13*, 278-285.

Prins, N., & Kingdom, F. A. A. (Producer). (2009) Palamedes: Matlab routines for analyzing psychophysical data. retrieved from http://www.palamedestoolbox.org

Raab, D. H. (1963). Backward masking. *Psychological Bulletin, 60*(2), 118-129.

Ramachandran, V. S., & Hubbard, E. M. (2001a). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings to The Royal Society of London, 268*(979-983).

Ramachandran, V. S., & Hubbard, E. M. (2001b). Synaesthesia - A window into perception, thought and language. *Journal of Consciousness, 8*(12), 3-34.

Rock, I., & Gutman, D. (1981). The effect of inattention of form perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*(2), 275-185.

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science, 143*(3606), 594-596.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition, 99*, 113-129.

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

Sanabria, D., Soto-Faraco, S., Chan, J. S., & Spence, C. (2004). When does visual perceptual grouping affect multisensory integration? . *Cognitive, Affective, & Behavioral Neuroscience, 4*, 218-229.

Scheier, C. R., Nijwahan, R., & Shimojo, S. (1999). Sound alters visual temporal perception. *Investigative Ophthalmology & Visual Science, 40*, 4169.

Schiffman, H. R. (2001a). **Fundamental visual functions and phenomena.** In H. R. Schiffman (Ed.), *Sensation and Perception: An Integrated Approach* (5th ed., pp. 89-115). New York: John Wiley & Sons, Inc.

Schiffman, H. R. (2001b). **Psychophysics**. In H. R. Schiffman (Ed.), *Sensation and Perception: An Integrated Approach* (5th ed., pp. 23-45). New York: John Wiley & Sons, Inc.

Schiller, P. (1935). Interrelation of different senses in perception. *The British Journal of Psychology: General selection, 25*, 265-469.

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science, 27*, 525-559.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime Reference Guide. Pittsburgh: Psychology Software Tools, Inc.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

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

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

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

Shams, L., Kamitani, Y., & Shimojo, S. (2004). Modulations of visual perception by sound. In G. Calvert, C. Spence & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 27-34). Cambridge, MA: The MIT Press.

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

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: Plasticity and interactions. *Current Opinion on Neurobiology, 11*, 505-509.

Smits, S. (2009). *The role of prior knowledge in audio-visual integration.* University Medisch Centrum St. Radboud, Nijmegen, Netherlands.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 330-345.

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

Spence, C., McDonald, J., & Driver, J. (2004). Exogenous spatial-cuing studies of human cross-modal attention and multisensory integration. In C. Spence & J. Driver (Eds.), *Crossmodal Space and Crossmodal Attention*. New York: Oxford University Press.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory Prior Entry. *Journal of Experimental Psychology: General, 130*(4), 799-832.

Stein, B. E., Stanford, T. R., Wallace, M. T., Vaughan, J. W., & Jiang, W. (2004). **Crossmodal spatial interactions in subcortical and cortical circuits** In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention*. New York: Oxford University Press.

Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: Loudness. *Psychological Review, 43*(5), 405-416.

Stone, J. V., Hunkin, N. V., Porrill, J., Wood, R., Keeler, V., Beanland, M., et al. (2001). *When is now? Perception of simultaneity*. Paper presented at the The Royal Society of London.

Teder-Salejarvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integtration. *Journal of Cognitve Neuroscience, 17*(9), 1396-1409.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science, 9*(5), 379-385.

Titchner, E. B. (1908). Lectures on the elementary psychology of feeling and attention Available from http://www.jstor.org.www.library.gatech.edu:2048/stable/2177879?seq=2

Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade intiiation based on competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitve Neuroscience, 13*(2), 256-271.

van Campen, C. (1997). Synesthesia and artistic experimentation. *Psyche, 3*(6). Retrieved from http://psyche.cs.monash.edu.au/v3/psyche-3-06-vancampen.html

van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2009). Temporal interval discrimination thresholds depend on peceived synchrony for audio-visual stimulus pairs. *Journal of Experimental Psychology: Human Perception and Performance, 35*(4), 1254-1263.

Vatakis, A., & Spence, C. (2006). Evaluating the inflence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica, 127*, 12-23.

Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: No role in temporal ventrilloqusim. *Journal of Experimental Psychology: Human Perception and Performance, 32*(4), 1063-1071.

Vu, K.-P. L., Strybel, T. Z., & Proctor, R. W. (2006). Effects of displacement magnitude and direction of auditory cues on auditory spatial facilitation of vsual search. . *Human Factors, 48*(3), 587-599.

Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio-visual intergration in temporal perception. *International Journal of Psychophysiology, 50*, 117-124.

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance, 7*(6), 1327-1339.

Wallace, M. T. (2004). **The development of multisensory integration**. In G. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 625-642). Cambridge, MA: The MIT Press.

Wallach, H., Bacon, J., & Schulman, P. (1978). Adaptation in motion perception: Alteration of induced motion. *Perception & Psychophysics, 24*(6), 509-514.

Watson, A., & Pelli, D. G. (1983). QUEST - A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*, 113-120.

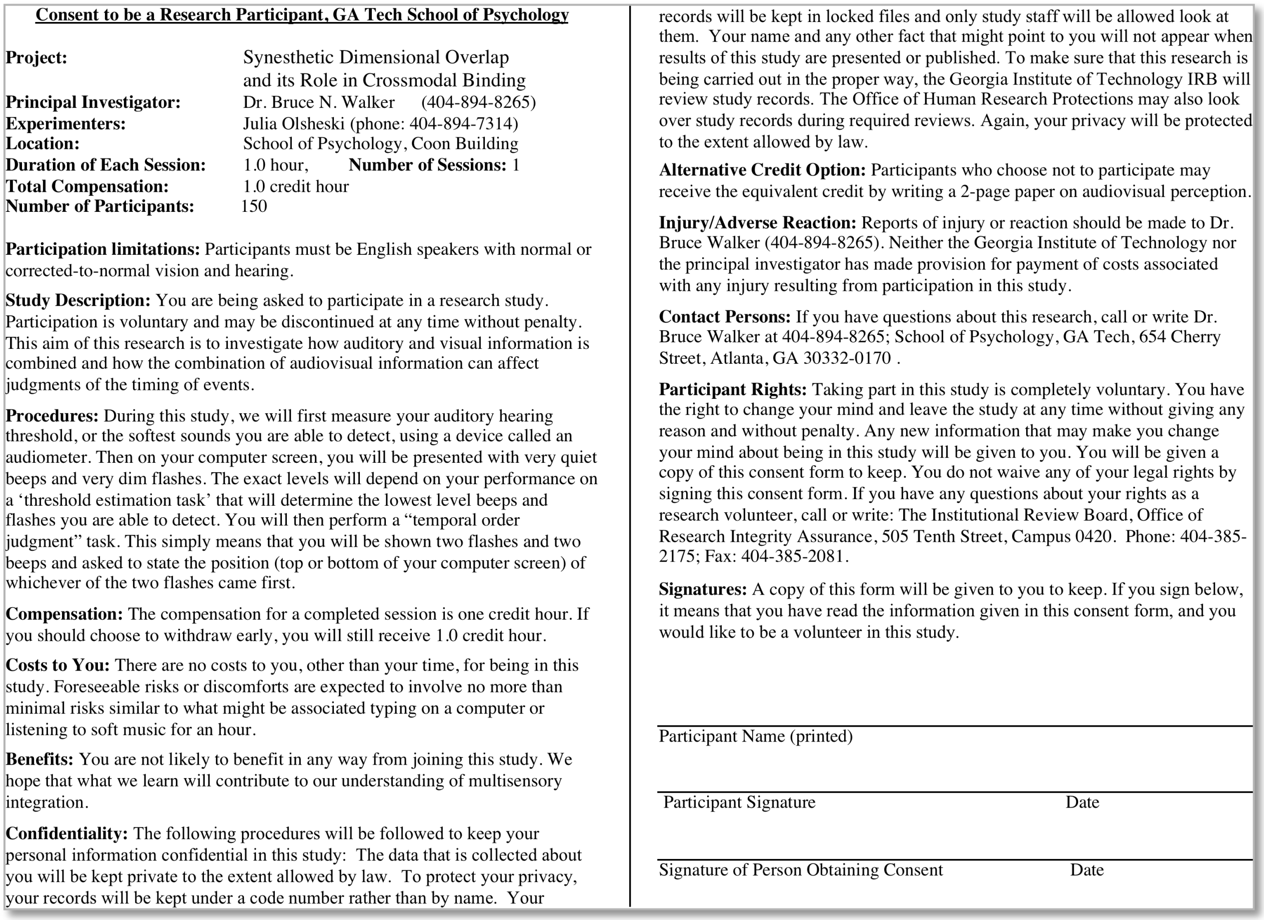
Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638-667.

Werner, H., & Wapner, S. (1952). Toward a general theory of perception. *Psychological Review, 59*(4), 324-338.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics, 67*(3), 531-544.

APPENDIX A

Participant Materials for Studies 1-3



Consent to be a Research Participant, GA Tech School of Psychology

Project:	Synesthetic Dimensional Overlap and its Role in Crossmodal Binding
Principal Investigator:	Dr. Bruce N. Walker (404-894-8265)
Experimenters:	Julia Olsheski (phone: 404-894-7314)
Location:	School of Psychology, Coon Building
Duration of Each Session:	1.0 hour, **Number of Sessions:** 1
Total Compensation:	1.0 credit hour
Number of Participants:	150

Participation limitations: Participants must be English speakers with normal or corrected-to-normal vision and hearing.

Study Description: You are being asked to participate in a research study. Participation is voluntary and may be discontinued at any time without penalty. This aim of this research is to investigate how auditory and visual information is combined and how the combination of audiovisual information can affect judgments of the timing of events.

Procedures: During this study, we will first measure your auditory hearing threshold, or the softest sounds you are able to detect, using a device called an audiometer. Then on your computer screen, you will be presented with very quiet beeps and very dim flashes. The exact levels will depend on your performance on a 'threshold estimation task' that will determine the lowest level beeps and flashes you are able to detect. You will then perform a "temporal order judgment" task. This simply means that you will be shown two flashes and two beeps and asked to state the position (top or bottom of your computer screen) of whichever of the two flashes came first.

Compensation: The compensation for a completed session is one credit hour. If you should choose to withdraw early, you will still receive 1.0 credit hour.

Costs to You: There are no costs to you, other than your time, for being in this study. Foreseeable risks or discomforts are expected to involve no more than minimal risks similar to what might be associated typing on a computer or listening to soft music for an hour.

Benefits: You are not likely to benefit in any way from joining this study. We hope that what we learn will contribute to our understanding of multisensory integration.

Confidentiality: The following procedures will be followed to keep your personal information confidential in this study: The data that is collected about you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code number rather than by name. Your records will be kept in locked files and only study staff will be allowed look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB will review study records. The Office of Human Research Protections may also look over study records during required reviews. Again, your privacy will be protected to the extent allowed by law.

Alternative Credit Option: Participants who choose not to participate may receive the equivalent credit by writing a 2-page paper on audiovisual perception.

Injury/Adverse Reaction: Reports of injury or reaction should be made to Dr. Bruce Walker (404-894-8265). Neither the Georgia Institute of Technology nor the principal investigator has made provision for payment of costs associated with any injury resulting from participation in this study.

Contact Persons: If you have questions about this research, call or write Dr. Bruce Walker at 404-894-8265; School of Psychology, GA Tech, 654 Cherry Street, Atlanta, GA 30332-0170 .

Participant Rights: Taking part in this study is completely voluntary. You have the right to change your mind and leave the study at any time without giving any reason and without penalty. Any new information that may make you change your mind about being in this study will be given to you. You will be given a copy of this consent form to keep. You do not waive any of your legal rights by signing this consent form. If you have any questions about your rights as a research volunteer, call or write: The Institutional Review Board, Office of Research Integrity Assurance, 505 Tenth Street, Campus 0420. Phone: 404-385-2175; Fax: 404-385-2081.

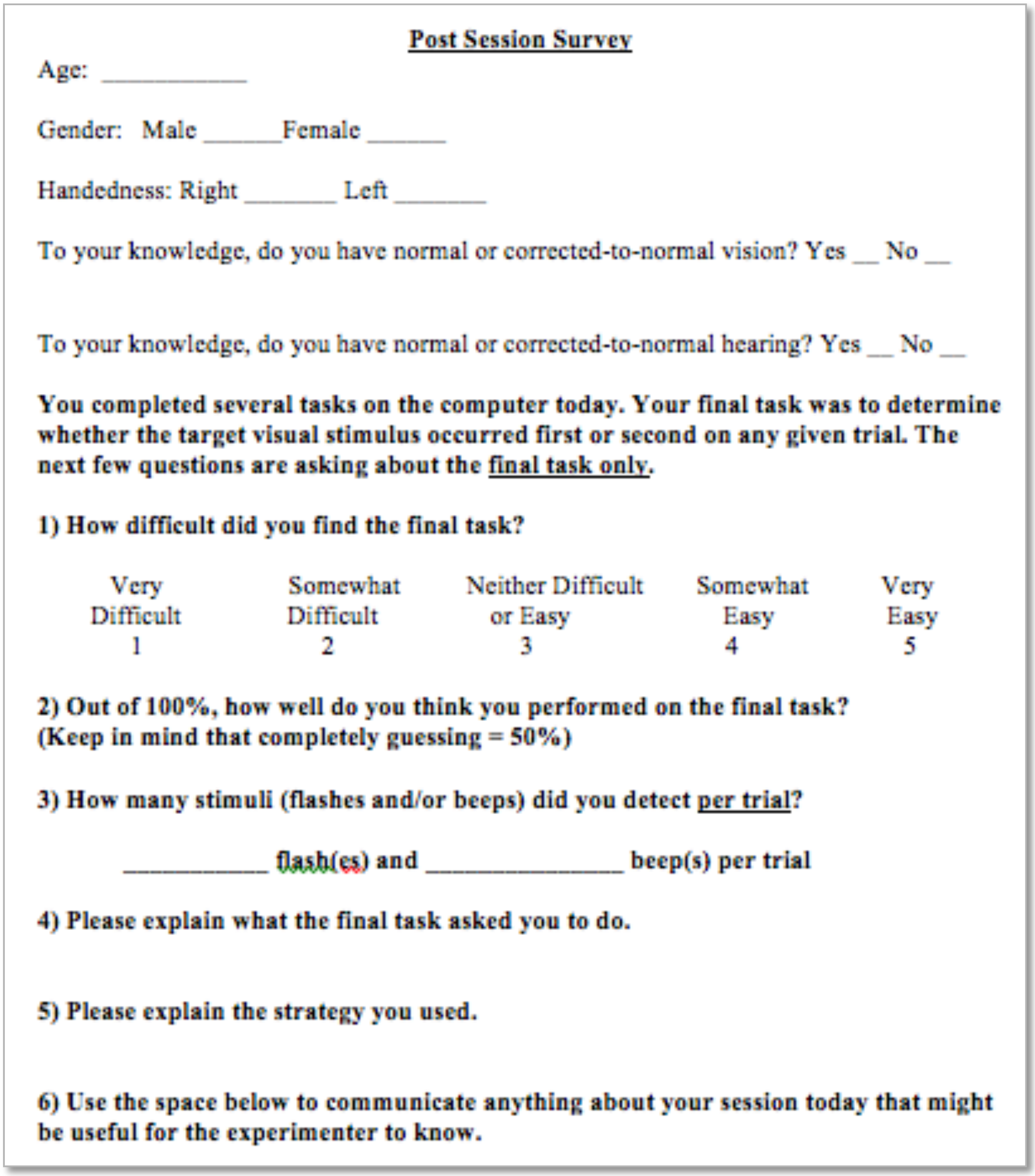
Signatures: A copy of this form will be given to you to keep. If you sign below, it means that you have read the information given in this consent form, and you would like to be a volunteer in this study.

Participant Name (printed)

Participant Signature Date

Signature of Person Obtaining Consent Date

Figure A1. Consent form for participants in studies 1-3.



Post Session Survey

Age: ____________

Gender: Male ______Female ______

Handedness: Right _______ Left _______

To your knowledge, do you have normal or corrected-to-normal vision? Yes __ No __

To your knowledge, do you have normal or corrected-to-normal hearing? Yes __ No __

You completed several tasks on the computer today. Your final task was to determine whether the target visual stimulus occurred first or second on any given trial. The next few questions are asking about the final task only.

1) How difficult did you find the final task?

Very Difficult	Somewhat Difficult	Neither Difficult or Easy	Somewhat Easy	Very Easy
1	2	3	4	5

2) Out of 100%, how well do you think you performed on the final task? (Keep in mind that completely guessing = 50%)

3) How many stimuli (flashes and/or beeps) did you detect per trial?

____________ flash(es) and ________________ beep(s) per trial

4) Please explain what the final task asked you to do.

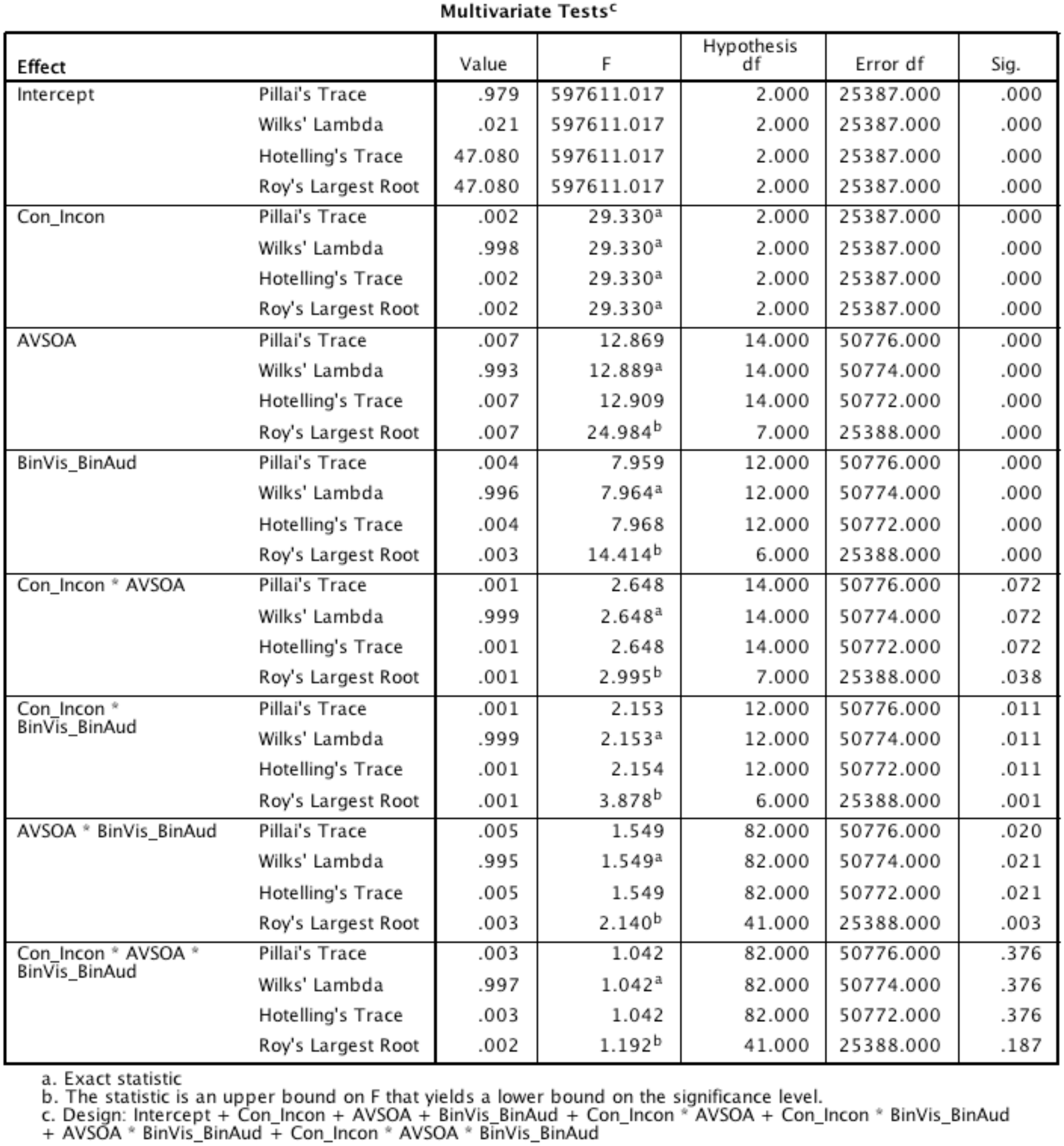
5) Please explain the strategy you used.

6) Use the space below to communicate anything about your session today that might be useful for the experimenter to know.

Figure A2. Post-Session Survey for studies 1-3

APPENDIX B

Study 1A Analysis Tables



Multivariate Tests[c]

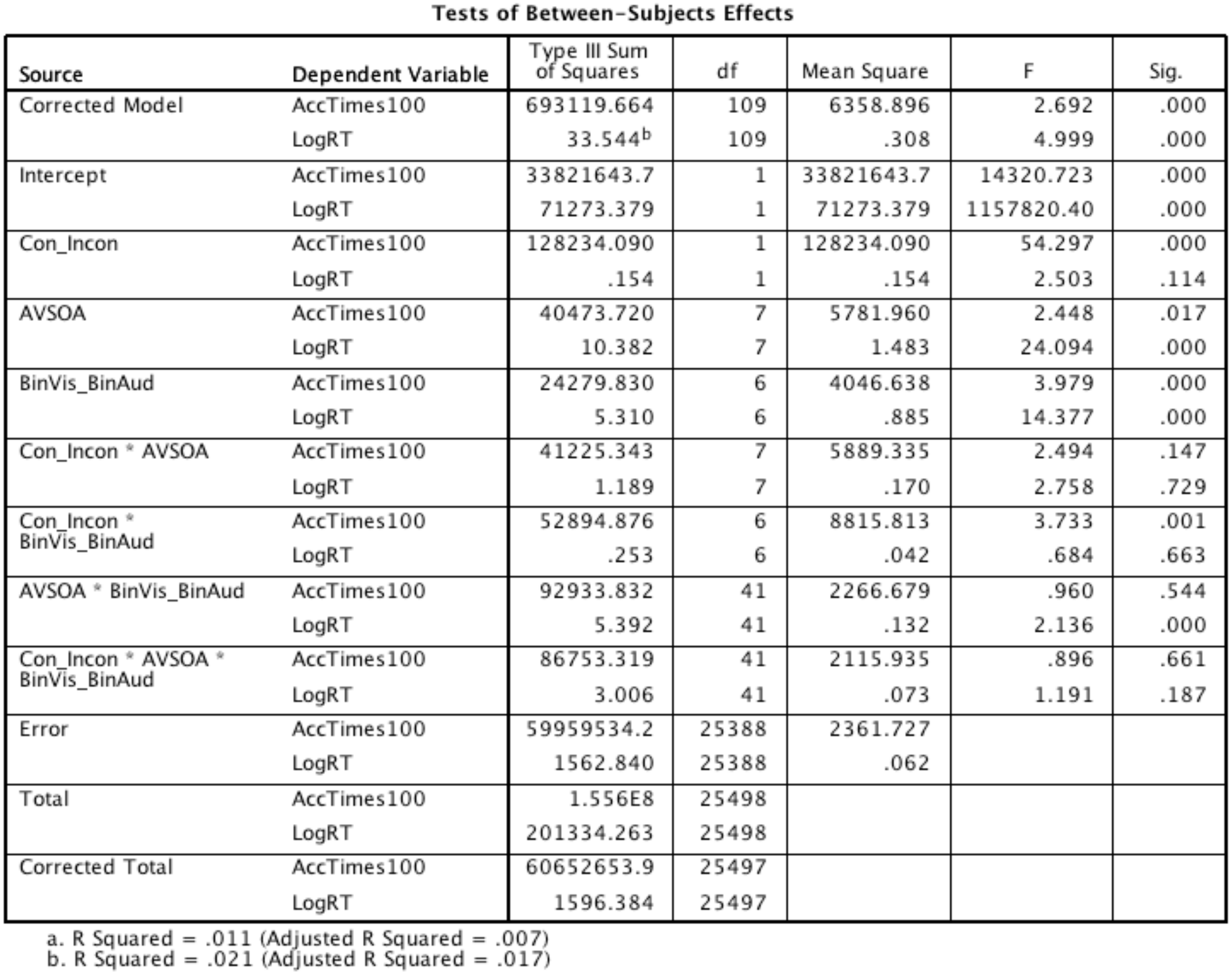
Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.979	597611.017	2.000	25387.000	.000
	Wilks' Lambda	.021	597611.017	2.000	25387.000	.000
	Hotelling's Trace	47.080	597611.017	2.000	25387.000	.000
	Roy's Largest Root	47.080	597611.017	2.000	25387.000	.000
Con_Incon	Pillai's Trace	.002	29.330[a]	2.000	25387.000	.000
	Wilks' Lambda	.998	29.330[a]	2.000	25387.000	.000
	Hotelling's Trace	.002	29.330[a]	2.000	25387.000	.000
	Roy's Largest Root	.002	29.330[a]	2.000	25387.000	.000
AVSOA	Pillai's Trace	.007	12.869	14.000	50776.000	.000
	Wilks' Lambda	.993	12.889[a]	14.000	50774.000	.000
	Hotelling's Trace	.007	12.909	14.000	50772.000	.000
	Roy's Largest Root	.007	24.984[b]	7.000	25388.000	.000
BinVis_BinAud	Pillai's Trace	.004	7.959	12.000	50776.000	.000
	Wilks' Lambda	.996	7.964[a]	12.000	50774.000	.000
	Hotelling's Trace	.004	7.968	12.000	50772.000	.000
	Roy's Largest Root	.003	14.414[b]	6.000	25388.000	.000
Con_Incon * AVSOA	Pillai's Trace	.001	2.648	14.000	50776.000	.072
	Wilks' Lambda	.999	2.648[a]	14.000	50774.000	.072
	Hotelling's Trace	.001	2.648	14.000	50772.000	.072
	Roy's Largest Root	.001	2.995[b]	7.000	25388.000	.038
Con_Incon * BinVis_BinAud	Pillai's Trace	.001	2.153	12.000	50776.000	.011
	Wilks' Lambda	.999	2.153[a]	12.000	50774.000	.011
	Hotelling's Trace	.001	2.154	12.000	50772.000	.011
	Roy's Largest Root	.001	3.878[b]	6.000	25388.000	.001
AVSOA * BinVis_BinAud	Pillai's Trace	.005	1.549	82.000	50776.000	.020
	Wilks' Lambda	.995	1.549[a]	82.000	50774.000	.021
	Hotelling's Trace	.005	1.549	82.000	50772.000	.021
	Roy's Largest Root	.003	2.140[b]	41.000	25388.000	.003
Con_Incon * AVSOA * BinVis_BinAud	Pillai's Trace	.003	1.042	82.000	50776.000	.376
	Wilks' Lambda	.997	1.042[a]	82.000	50774.000	.376
	Hotelling's Trace	.003	1.042	82.000	50772.000	.376
	Roy's Largest Root	.002	1.192[b]	41.000	25388.000	.187

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept + Con_Incon + AVSOA + BinVis_BinAud + Con_Incon * AVSOA + Con_Incon * BinVis_BinAud + AVSOA * BinVis_BinAud + Con_Incon * AVSOA * BinVis_BinAud

Figure B1. Study 1A MANOVA output.



Tests of Between-Subjects Effects

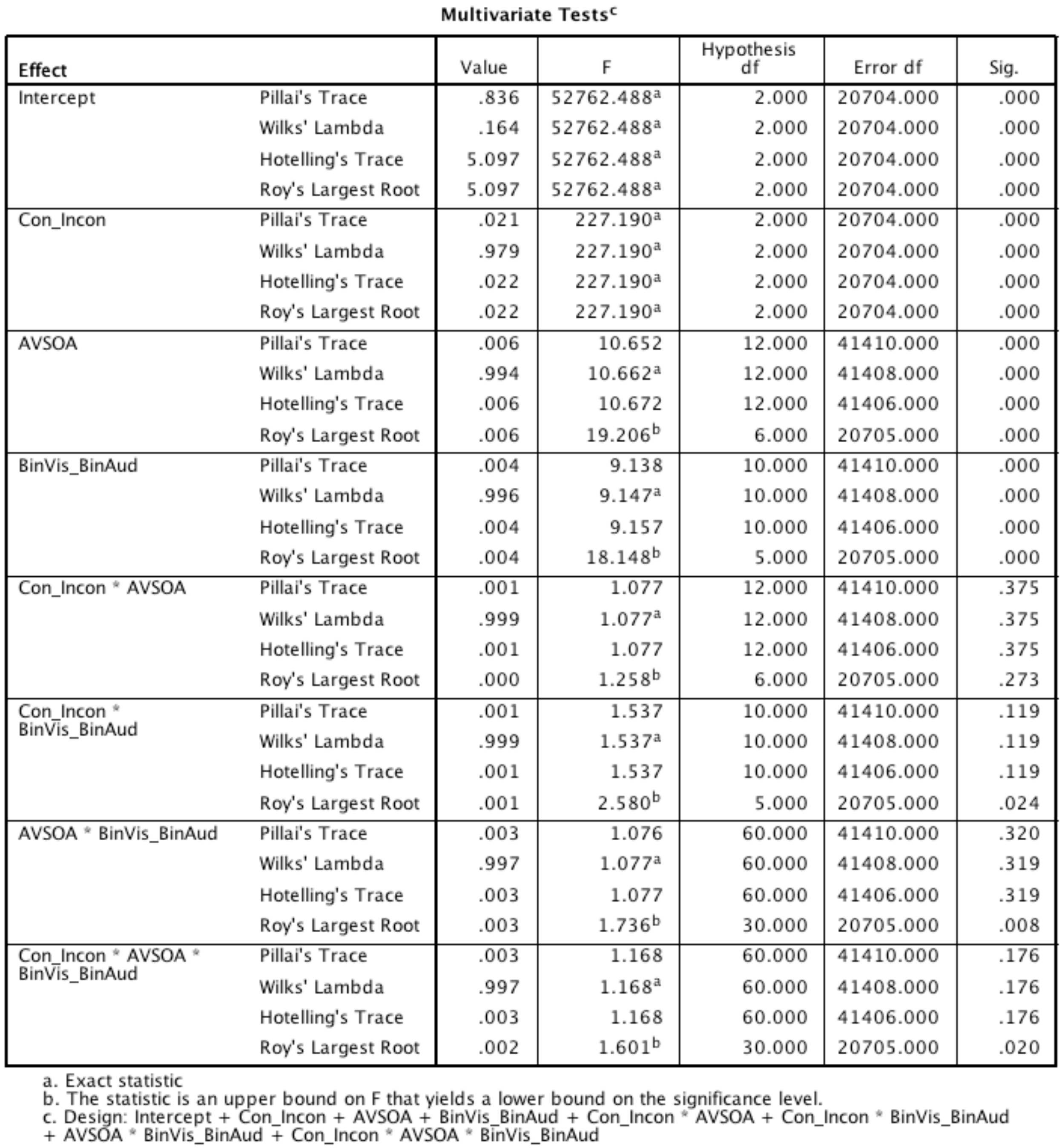
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	AccTimes100	693119.664	109	6358.896	2.692	.000
	LogRT	33.544^b	109	.308	4.999	.000
Intercept	AccTimes100	33821643.7	1	33821643.7	14320.723	.000
	LogRT	71273.379	1	71273.379	1157820.40	.000
Con_Incon	AccTimes100	128234.090	1	128234.090	54.297	.000
	LogRT	.154	1	.154	2.503	.114
AVSOA	AccTimes100	40473.720	7	5781.960	2.448	.017
	LogRT	10.382	7	1.483	24.094	.000
BinVis_BinAud	AccTimes100	24279.830	6	4046.638	3.979	.000
	LogRT	5.310	6	.885	14.377	.000
Con_Incon * AVSOA	AccTimes100	41225.343	7	5889.335	2.494	.147
	LogRT	1.189	7	.170	2.758	.729
Con_Incon * BinVis_BinAud	AccTimes100	52894.876	6	8815.813	3.733	.001
	LogRT	.253	6	.042	.684	.663
AVSOA * BinVis_BinAud	AccTimes100	92933.832	41	2266.679	.960	.544
	LogRT	5.392	41	.132	2.136	.000
Con_Incon * AVSOA * BinVis_BinAud	AccTimes100	86753.319	41	2115.935	.896	.661
	LogRT	3.006	41	.073	1.191	.187
Error	AccTimes100	59959534.2	25388	2361.727		
	LogRT	1562.840	25388	.062		
Total	AccTimes100	1.556E8	25498			
	LogRT	201334.263	25498			
Corrected Total	AccTimes100	60652653.9	25497			
	LogRT	1596.384	25497			

a. R Squared = .011 (Adjusted R Squared = .007)
b. R Squared = .021 (Adjusted R Squared = .017)

Figure B2. Study 1A tests of between-subjects effects output.

APPENDIX C

Study 1B Analysis Tables



Multivariate Tests[c]

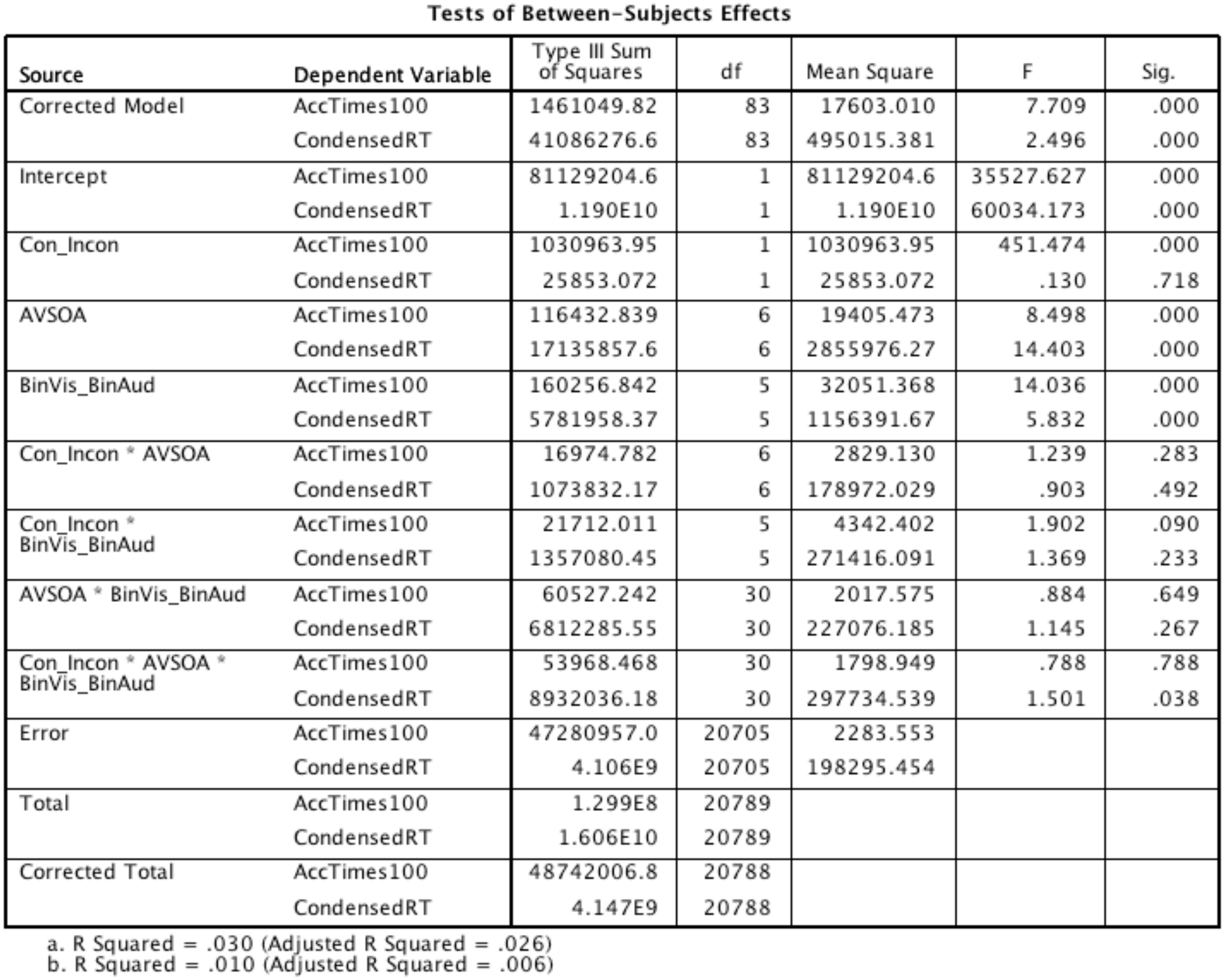
Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.836	52762.488[a]	2.000	20704.000	.000
	Wilks' Lambda	.164	52762.488[a]	2.000	20704.000	.000
	Hotelling's Trace	5.097	52762.488[a]	2.000	20704.000	.000
	Roy's Largest Root	5.097	52762.488[a]	2.000	20704.000	.000
Con_Incon	Pillai's Trace	.021	227.190[a]	2.000	20704.000	.000
	Wilks' Lambda	.979	227.190[a]	2.000	20704.000	.000
	Hotelling's Trace	.022	227.190[a]	2.000	20704.000	.000
	Roy's Largest Root	.022	227.190[a]	2.000	20704.000	.000
AVSOA	Pillai's Trace	.006	10.652	12.000	41410.000	.000
	Wilks' Lambda	.994	10.662[a]	12.000	41408.000	.000
	Hotelling's Trace	.006	10.672	12.000	41406.000	.000
	Roy's Largest Root	.006	19.206[b]	6.000	20705.000	.000
BinVis_BinAud	Pillai's Trace	.004	9.138	10.000	41410.000	.000
	Wilks' Lambda	.996	9.147[a]	10.000	41408.000	.000
	Hotelling's Trace	.004	9.157	10.000	41406.000	.000
	Roy's Largest Root	.004	18.148[b]	5.000	20705.000	.000
Con_Incon * AVSOA	Pillai's Trace	.001	1.077	12.000	41410.000	.375
	Wilks' Lambda	.999	1.077[a]	12.000	41408.000	.375
	Hotelling's Trace	.001	1.077	12.000	41406.000	.375
	Roy's Largest Root	.000	1.258[b]	6.000	20705.000	.273
Con_Incon * BinVis_BinAud	Pillai's Trace	.001	1.537	10.000	41410.000	.119
	Wilks' Lambda	.999	1.537[a]	10.000	41408.000	.119
	Hotelling's Trace	.001	1.537	10.000	41406.000	.119
	Roy's Largest Root	.001	2.580[b]	5.000	20705.000	.024
AVSOA * BinVis_BinAud	Pillai's Trace	.003	1.076	60.000	41410.000	.320
	Wilks' Lambda	.997	1.077[a]	60.000	41408.000	.319
	Hotelling's Trace	.003	1.077	60.000	41406.000	.319
	Roy's Largest Root	.003	1.736[b]	30.000	20705.000	.008
Con_Incon * AVSOA * BinVis_BinAud	Pillai's Trace	.003	1.168	60.000	41410.000	.176
	Wilks' Lambda	.997	1.168[a]	60.000	41408.000	.176
	Hotelling's Trace	.003	1.168	60.000	41406.000	.176
	Roy's Largest Root	.002	1.601[b]	30.000	20705.000	.020

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept + Con_Incon + AVSOA + BinVis_BinAud + Con_Incon * AVSOA + Con_Incon * BinVis_BinAud + AVSOA * BinVis_BinAud + Con_Incon * AVSOA * BinVis_BinAud

Figure C1. Study 1B MANOVA output.



Tests of Between-Subjects Effects

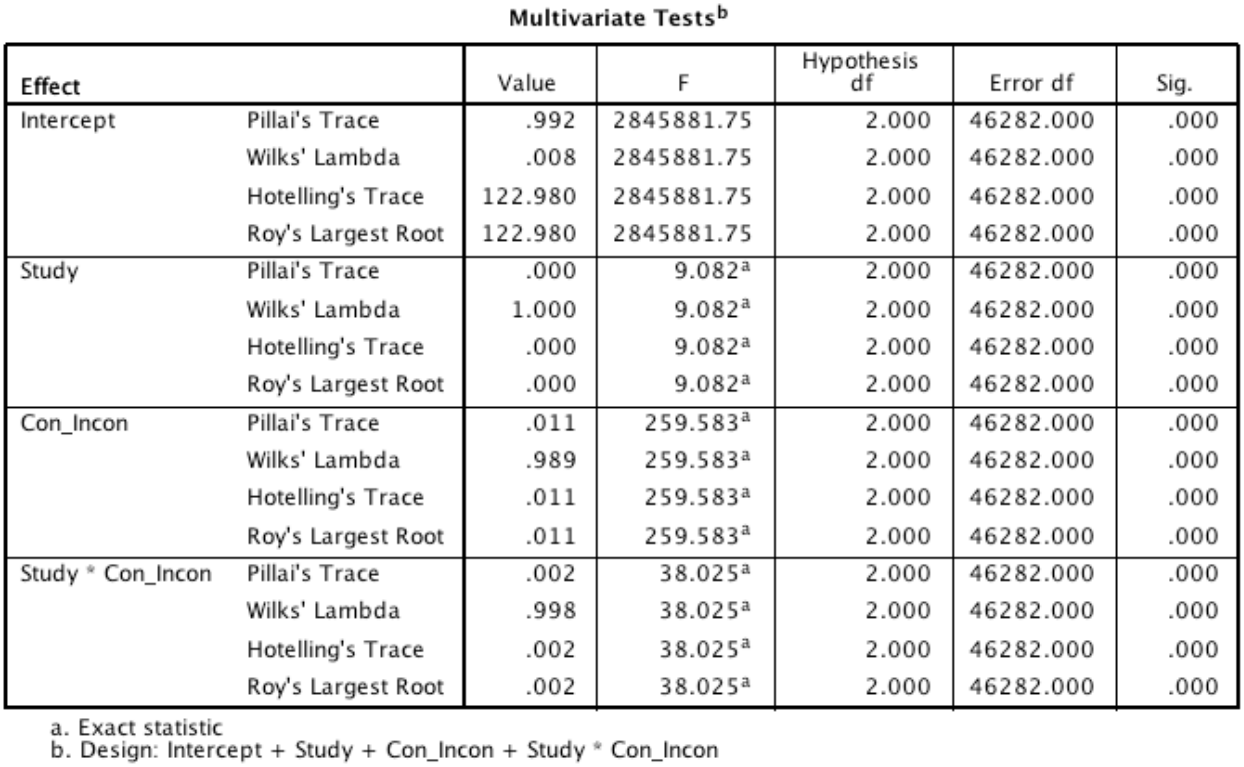
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	AccTimes100	1461049.82	83	17603.010	7.709	.000
	CondensedRT	41086276.6	83	495015.381	2.496	.000
Intercept	AccTimes100	81129204.6	1	81129204.6	35527.627	.000
	CondensedRT	1.190E10	1	1.190E10	60034.173	.000
Con_Incon	AccTimes100	1030963.95	1	1030963.95	451.474	.000
	CondensedRT	25853.072	1	25853.072	.130	.718
AVSOA	AccTimes100	116432.839	6	19405.473	8.498	.000
	CondensedRT	17135857.6	6	2855976.27	14.403	.000
BinVis_BinAud	AccTimes100	160256.842	5	32051.368	14.036	.000
	CondensedRT	5781958.37	5	1156391.67	5.832	.000
Con_Incon * AVSOA	AccTimes100	16974.782	6	2829.130	1.239	.283
	CondensedRT	1073832.17	6	178972.029	.903	.492
Con_Incon * BinVis_BinAud	AccTimes100	21712.011	5	4342.402	1.902	.090
	CondensedRT	1357080.45	5	271416.091	1.369	.233
AVSOA * BinVis_BinAud	AccTimes100	60527.242	30	2017.575	.884	.649
	CondensedRT	6812285.55	30	227076.185	1.145	.267
Con_Incon * AVSOA * BinVis_BinAud	AccTimes100	53968.468	30	1798.949	.788	.788
	CondensedRT	8932036.18	30	297734.539	1.501	.038
Error	AccTimes100	47280957.0	20705	2283.553		
	CondensedRT	4.106E9	20705	198295.454		
Total	AccTimes100	1.299E8	20789			
	CondensedRT	1.606E10	20789			
Corrected Total	AccTimes100	48742006.8	20788			
	CondensedRT	4.147E9	20788			

a. R Squared = .030 (Adjusted R Squared = .026)
b. R Squared = .010 (Adjusted R Squared = .006)

Figure C2. Study 1B tests of between-subjects effects output.

APPENDIX D

Instruction Manipulation Analysis Tables



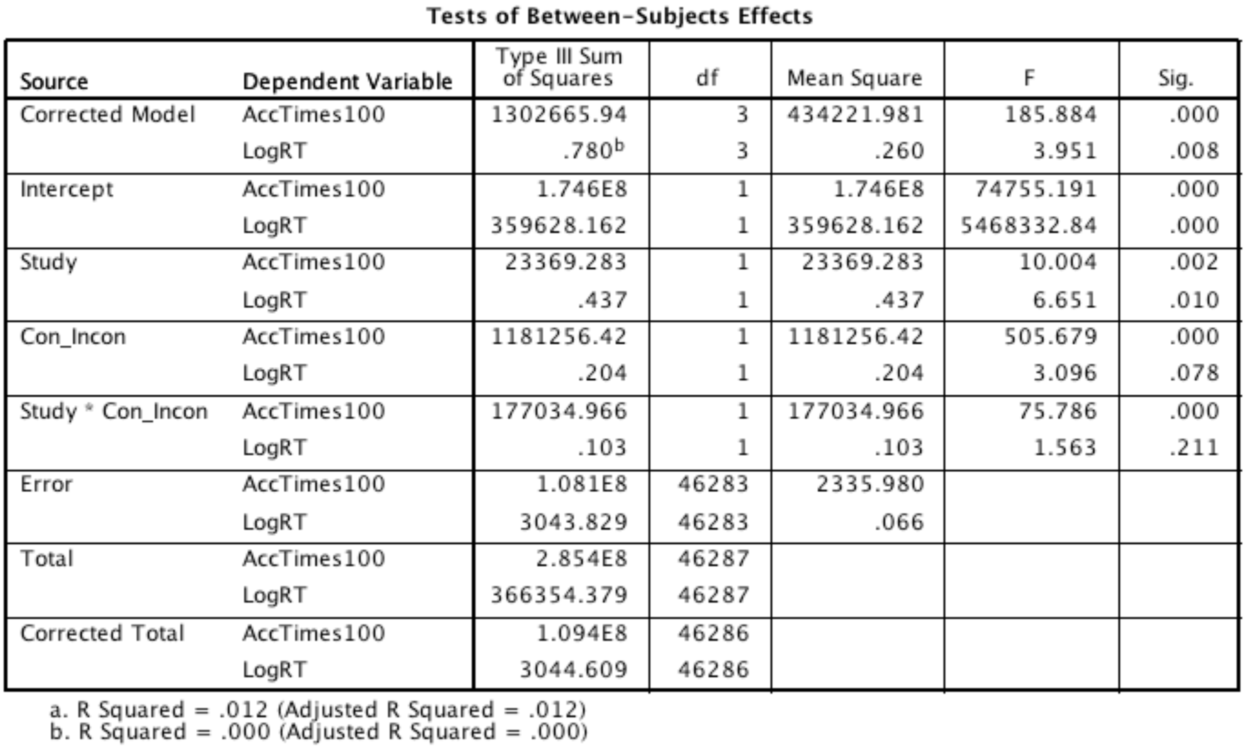
Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.992	2845881.75	2.000	46282.000	.000
	Wilks' Lambda	.008	2845881.75	2.000	46282.000	.000
	Hotelling's Trace	122.980	2845881.75	2.000	46282.000	.000
	Roy's Largest Root	122.980	2845881.75	2.000	46282.000	.000
Study	Pillai's Trace	.000	9.082[a]	2.000	46282.000	.000
	Wilks' Lambda	1.000	9.082[a]	2.000	46282.000	.000
	Hotelling's Trace	.000	9.082[a]	2.000	46282.000	.000
	Roy's Largest Root	.000	9.082[a]	2.000	46282.000	.000
Con_Incon	Pillai's Trace	.011	259.583[a]	2.000	46282.000	.000
	Wilks' Lambda	.989	259.583[a]	2.000	46282.000	.000
	Hotelling's Trace	.011	259.583[a]	2.000	46282.000	.000
	Roy's Largest Root	.011	259.583[a]	2.000	46282.000	.000
Study * Con_Incon	Pillai's Trace	.002	38.025[a]	2.000	46282.000	.000
	Wilks' Lambda	.998	38.025[a]	2.000	46282.000	.000
	Hotelling's Trace	.002	38.025[a]	2.000	46282.000	.000
	Roy's Largest Root	.002	38.025[a]	2.000	46282.000	.000

a. Exact statistic

b. Design: Intercept + Study + Con_Incon + Study * Con_Incon

Figure D1. Study 1A compared with Study 1B (instruction manipulation) MANOVA output.



Tests of Between-Subjects Effects

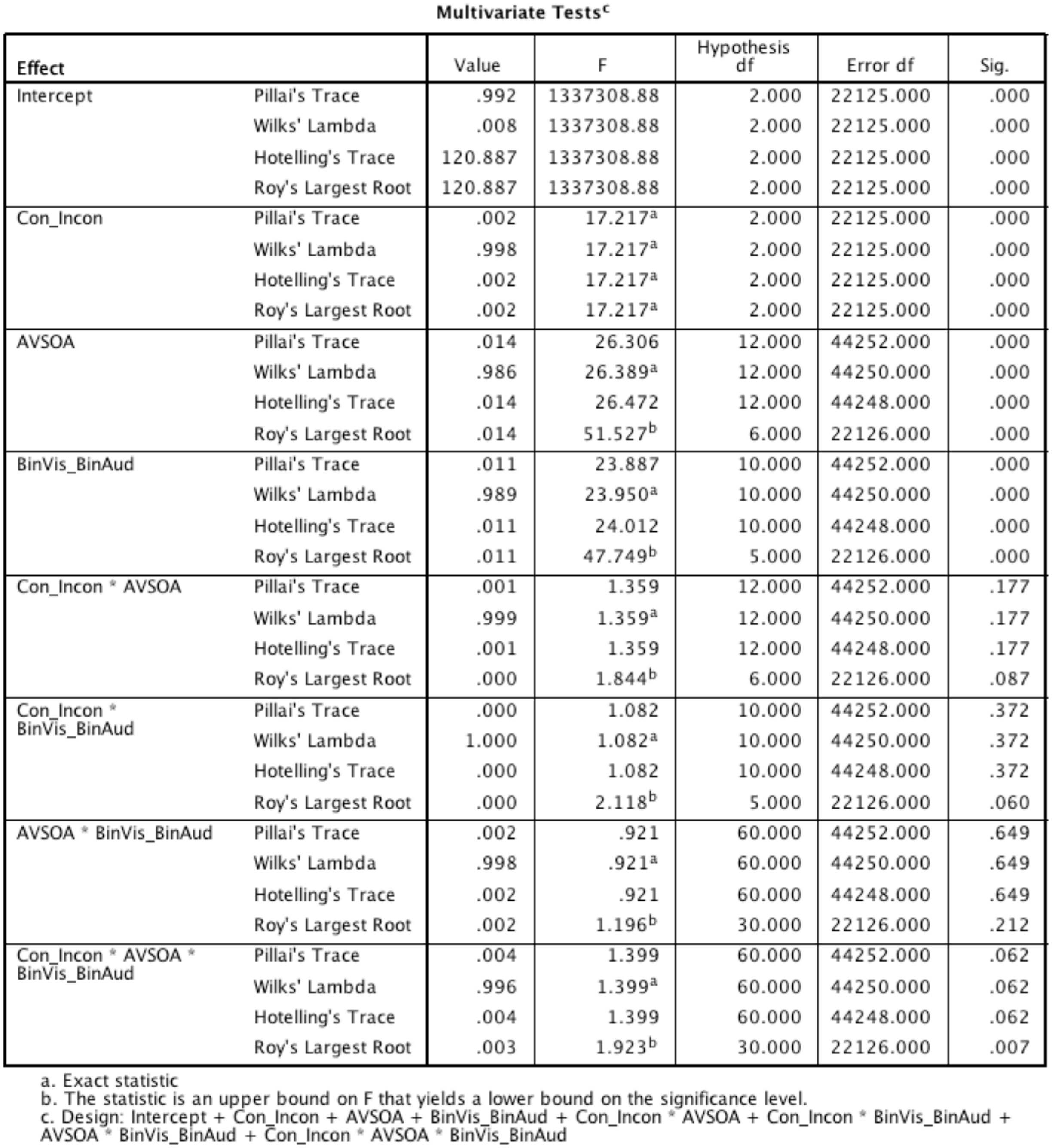
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	AccTimes100	1302665.94	3	434221.981	185.884	.000
	LogRT	.780[b]	3	.260	3.951	.008
Intercept	AccTimes100	1.746E8	1	1.746E8	74755.191	.000
	LogRT	359628.162	1	359628.162	5468332.84	.000
Study	AccTimes100	23369.283	1	23369.283	10.004	.002
	LogRT	.437	1	.437	6.651	.010
Con_Incon	AccTimes100	1181256.42	1	1181256.42	505.679	.000
	LogRT	.204	1	.204	3.096	.078
Study * Con_Incon	AccTimes100	177034.966	1	177034.966	75.786	.000
	LogRT	.103	1	.103	1.563	.211
Error	AccTimes100	1.081E8	46283	2335.980		
	LogRT	3043.829	46283	.066		
Total	AccTimes100	2.854E8	46287			
	LogRT	366354.379	46287			
Corrected Total	AccTimes100	1.094E8	46286			
	LogRT	3044.609	46286			

a. R Squared = .012 (Adjusted R Squared = .012)
b. R Squared = .000 (Adjusted R Squared = .000)

Figure D2. *Study 1A compared with Study 1B (instruction manipulation) tests of between-subjects effects output.*

APPENDIX E

Study 2 Analysis Tables



Multivariate Tests[c]

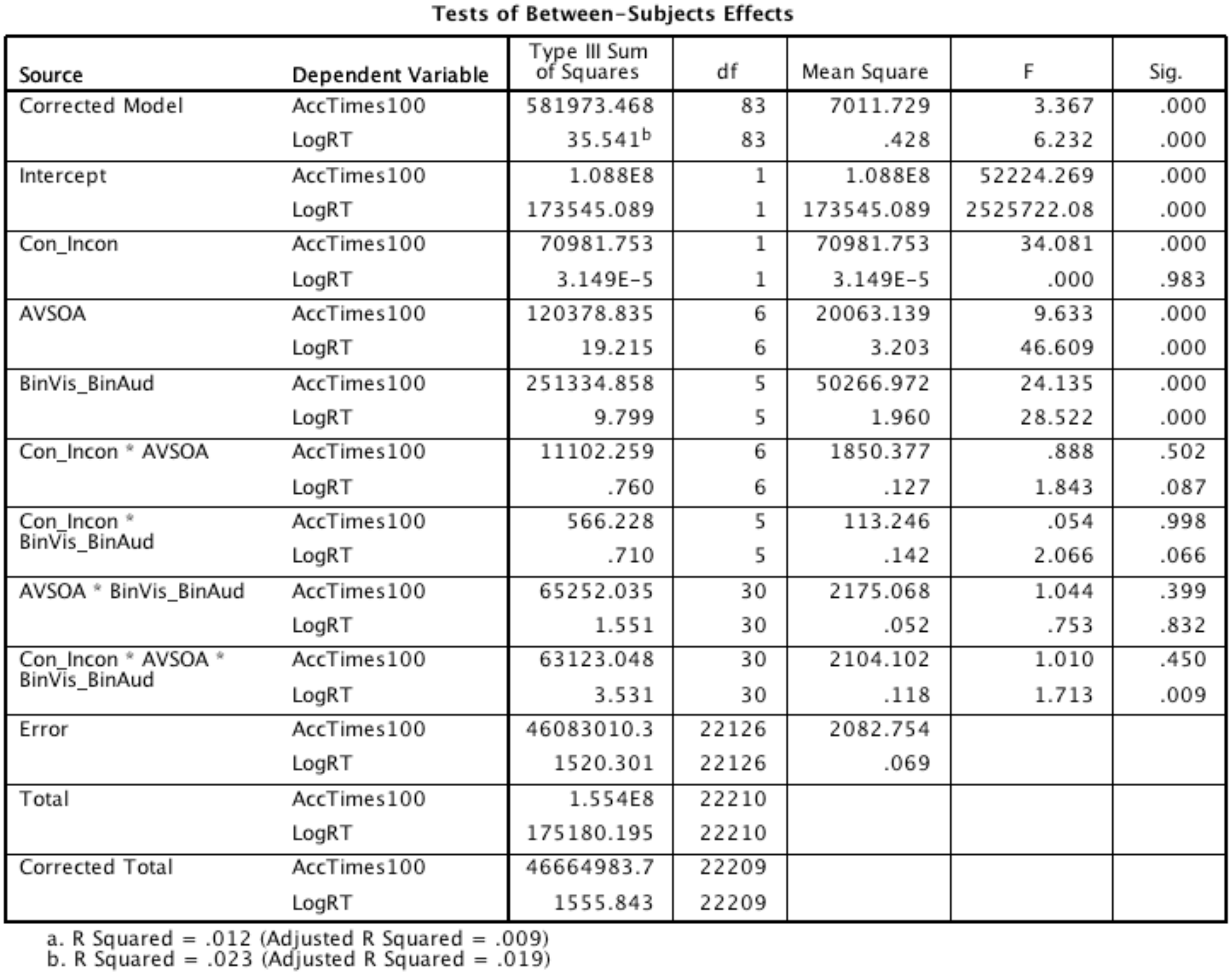
Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.992	1337308.88	2.000	22125.000	.000
	Wilks' Lambda	.008	1337308.88	2.000	22125.000	.000
	Hotelling's Trace	120.887	1337308.88	2.000	22125.000	.000
	Roy's Largest Root	120.887	1337308.88	2.000	22125.000	.000
Con_Incon	Pillai's Trace	.002	17.217[a]	2.000	22125.000	.000
	Wilks' Lambda	.998	17.217[a]	2.000	22125.000	.000
	Hotelling's Trace	.002	17.217[a]	2.000	22125.000	.000
	Roy's Largest Root	.002	17.217[a]	2.000	22125.000	.000
AVSOA	Pillai's Trace	.014	26.306	12.000	44252.000	.000
	Wilks' Lambda	.986	26.389[a]	12.000	44250.000	.000
	Hotelling's Trace	.014	26.472	12.000	44248.000	.000
	Roy's Largest Root	.014	51.527[b]	6.000	22126.000	.000
BinVis_BinAud	Pillai's Trace	.011	23.887	10.000	44252.000	.000
	Wilks' Lambda	.989	23.950[a]	10.000	44250.000	.000
	Hotelling's Trace	.011	24.012	10.000	44248.000	.000
	Roy's Largest Root	.011	47.749[b]	5.000	22126.000	.000
Con_Incon * AVSOA	Pillai's Trace	.001	1.359	12.000	44252.000	.177
	Wilks' Lambda	.999	1.359[a]	12.000	44250.000	.177
	Hotelling's Trace	.001	1.359	12.000	44248.000	.177
	Roy's Largest Root	.000	1.844[b]	6.000	22126.000	.087
Con_Incon * BinVis_BinAud	Pillai's Trace	.000	1.082	10.000	44252.000	.372
	Wilks' Lambda	1.000	1.082[a]	10.000	44250.000	.372
	Hotelling's Trace	.000	1.082	10.000	44248.000	.372
	Roy's Largest Root	.000	2.118[b]	5.000	22126.000	.060
AVSOA * BinVis_BinAud	Pillai's Trace	.002	.921	60.000	44252.000	.649
	Wilks' Lambda	.998	.921[a]	60.000	44250.000	.649
	Hotelling's Trace	.002	.921	60.000	44248.000	.649
	Roy's Largest Root	.002	1.196[b]	30.000	22126.000	.212
Con_Incon * AVSOA * BinVis_BinAud	Pillai's Trace	.004	1.399	60.000	44252.000	.062
	Wilks' Lambda	.996	1.399[a]	60.000	44250.000	.062
	Hotelling's Trace	.004	1.399	60.000	44248.000	.062
	Roy's Largest Root	.003	1.923[b]	30.000	22126.000	.007

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept + Con_Incon + AVSOA + BinVis_BinAud + Con_Incon * AVSOA + Con_Incon * BinVis_BinAud + AVSOA * BinVis_BinAud + Con_Incon * AVSOA * BinVis_BinAud

Figure E1. *Study 2 MANOVA output.*



Tests of Between-Subjects Effects

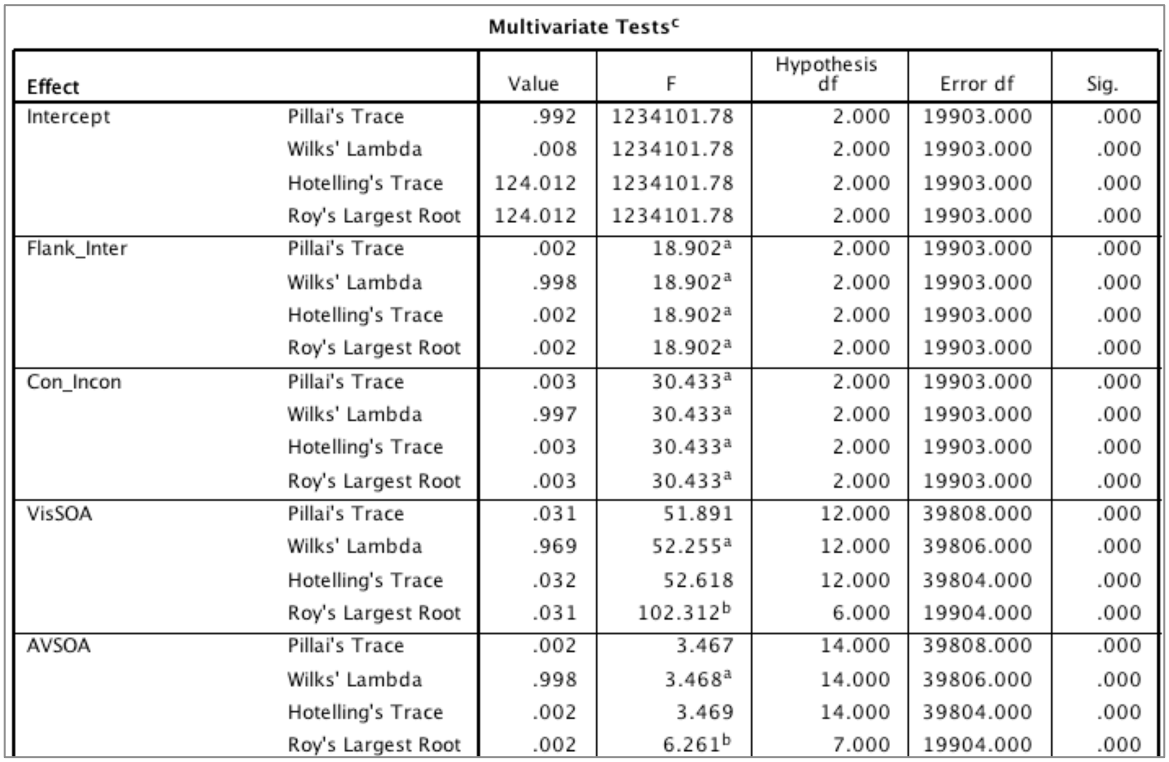
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	AccTimes100	581973.468	83	7011.729	3.367	.000
	LogRT	35.541[b]	83	.428	6.232	.000
Intercept	AccTimes100	1.088E8	1	1.088E8	52224.269	.000
	LogRT	173545.089	1	173545.089	2525722.08	.000
Con_Incon	AccTimes100	70981.753	1	70981.753	34.081	.000
	LogRT	3.149E-5	1	3.149E-5	.000	.983
AVSOA	AccTimes100	120378.835	6	20063.139	9.633	.000
	LogRT	19.215	6	3.203	46.609	.000
BinVis_BinAud	AccTimes100	251334.858	5	50266.972	24.135	.000
	LogRT	9.799	5	1.960	28.522	.000
Con_Incon * AVSOA	AccTimes100	11102.259	6	1850.377	.888	.502
	LogRT	.760	6	.127	1.843	.087
Con_Incon * BinVis_BinAud	AccTimes100	566.228	5	113.246	.054	.998
	LogRT	.710	5	.142	2.066	.066
AVSOA * BinVis_BinAud	AccTimes100	65252.035	30	2175.068	1.044	.399
	LogRT	1.551	30	.052	.753	.832
Con_Incon * AVSOA * BinVis_BinAud	AccTimes100	63123.048	30	2104.102	1.010	.450
	LogRT	3.531	30	.118	1.713	.009
Error	AccTimes100	46083010.3	22126	2082.754		
	LogRT	1520.301	22126	.069		
Total	AccTimes100	1.554E8	22210			
	LogRT	175180.195	22210			
Corrected Total	AccTimes100	46664983.7	22209			
	LogRT	1555.843	22209			

a. R Squared = .012 (Adjusted R Squared = .009)
b. R Squared = .023 (Adjusted R Squared = .019)

Figure E2. Study 2 tests of between-subjects effects output.

APPENDIX F

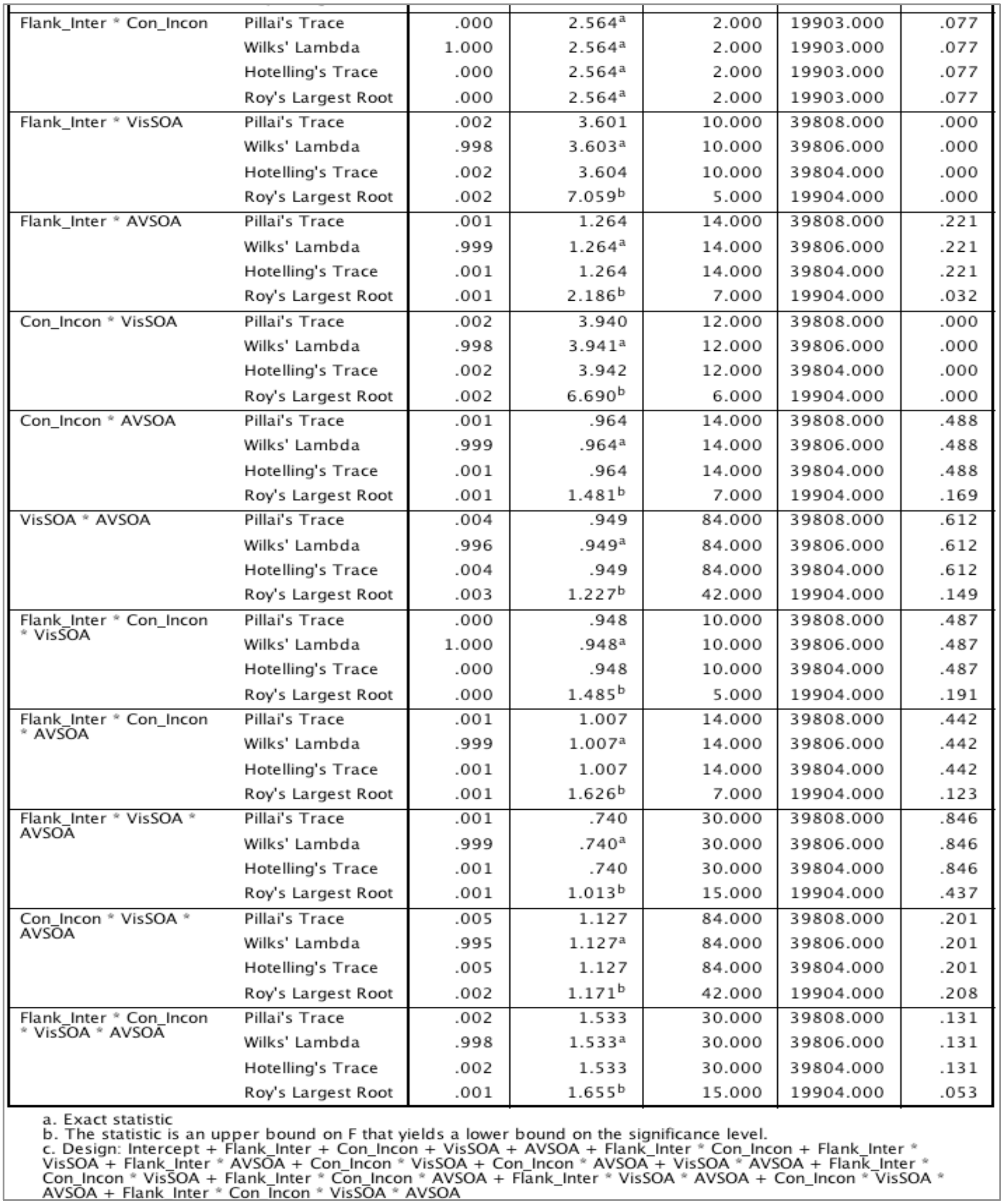
Study 3 Analysis Tables



Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.992	1234101.78	2.000	19903.000	.000
	Wilks' Lambda	.008	1234101.78	2.000	19903.000	.000
	Hotelling's Trace	124.012	1234101.78	2.000	19903.000	.000
	Roy's Largest Root	124.012	1234101.78	2.000	19903.000	.000
Flank_Inter	Pillai's Trace	.002	18.902[a]	2.000	19903.000	.000
	Wilks' Lambda	.998	18.902[a]	2.000	19903.000	.000
	Hotelling's Trace	.002	18.902[a]	2.000	19903.000	.000
	Roy's Largest Root	.002	18.902[a]	2.000	19903.000	.000
Con_Incon	Pillai's Trace	.003	30.433[a]	2.000	19903.000	.000
	Wilks' Lambda	.997	30.433[a]	2.000	19903.000	.000
	Hotelling's Trace	.003	30.433[a]	2.000	19903.000	.000
	Roy's Largest Root	.003	30.433[a]	2.000	19903.000	.000
VisSOA	Pillai's Trace	.031	51.891	12.000	39808.000	.000
	Wilks' Lambda	.969	52.255[a]	12.000	39806.000	.000
	Hotelling's Trace	.032	52.618	12.000	39804.000	.000
	Roy's Largest Root	.031	102.312[b]	6.000	19904.000	.000
AVSOA	Pillai's Trace	.002	3.467	14.000	39808.000	.000
	Wilks' Lambda	.998	3.468[a]	14.000	39806.000	.000
	Hotelling's Trace	.002	3.469	14.000	39804.000	.000
	Roy's Largest Root	.002	6.261[b]	7.000	19904.000	.000

Figure F1. Study 2 main effects MANOVA output.



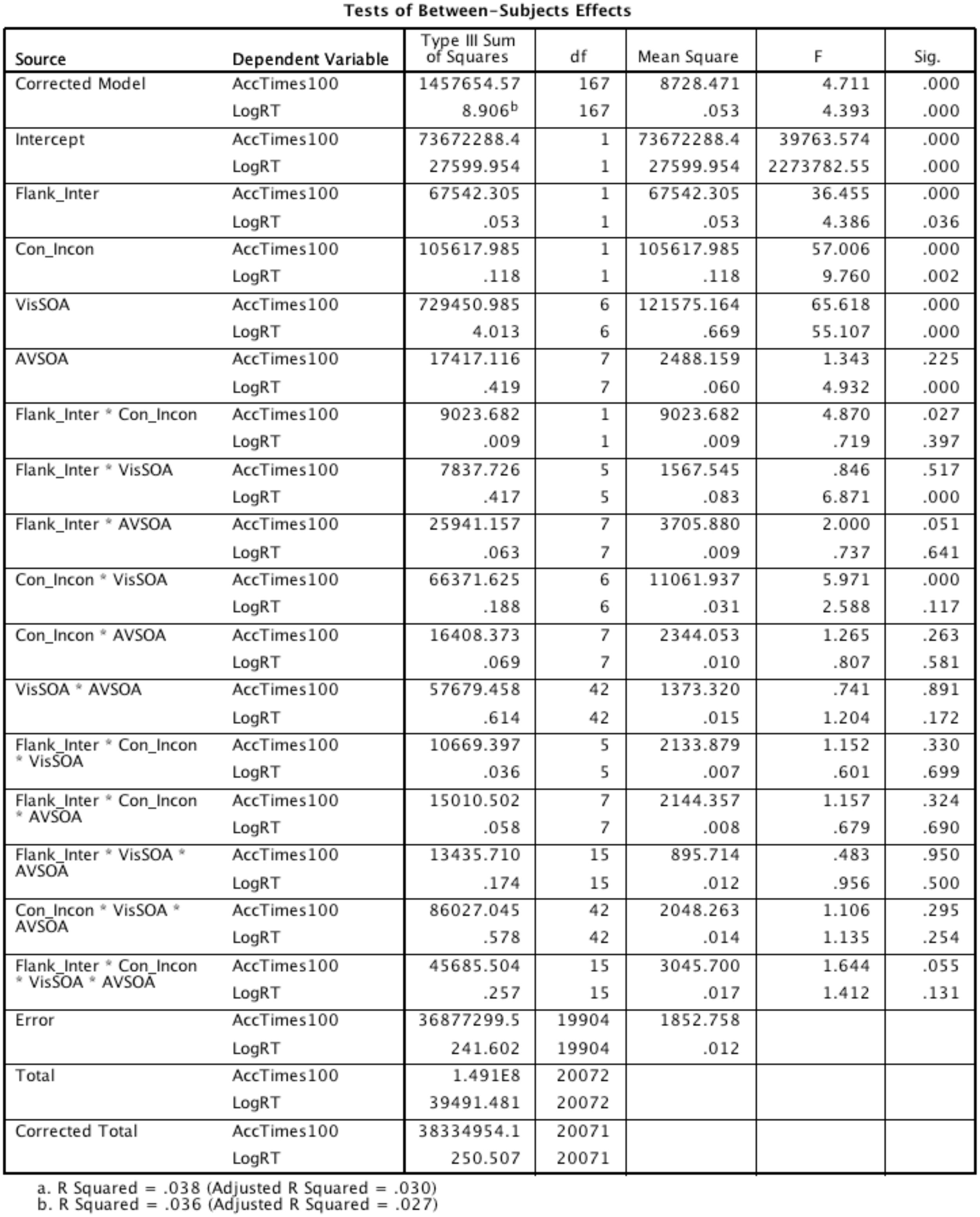
Flank_Inter * Con_Incon	Pillai's Trace	.000	2.564[a]	2.000	19903.000	.077
	Wilks' Lambda	1.000	2.564[a]	2.000	19903.000	.077
	Hotelling's Trace	.000	2.564[a]	2.000	19903.000	.077
	Roy's Largest Root	.000	2.564[a]	2.000	19903.000	.077
Flank_Inter * VisSOA	Pillai's Trace	.002	3.601	10.000	39808.000	.000
	Wilks' Lambda	.998	3.603[a]	10.000	39806.000	.000
	Hotelling's Trace	.002	3.604	10.000	39804.000	.000
	Roy's Largest Root	.002	7.059[b]	5.000	19904.000	.000
Flank_Inter * AVSOA	Pillai's Trace	.001	1.264	14.000	39808.000	.221
	Wilks' Lambda	.999	1.264[a]	14.000	39806.000	.221
	Hotelling's Trace	.001	1.264	14.000	39804.000	.221
	Roy's Largest Root	.001	2.186[b]	7.000	19904.000	.032
Con_Incon * VisSOA	Pillai's Trace	.002	3.940	12.000	39808.000	.000
	Wilks' Lambda	.998	3.941[a]	12.000	39806.000	.000
	Hotelling's Trace	.002	3.942	12.000	39804.000	.000
	Roy's Largest Root	.002	6.690[b]	6.000	19904.000	.000
Con_Incon * AVSOA	Pillai's Trace	.001	.964	14.000	39808.000	.488
	Wilks' Lambda	.999	.964[a]	14.000	39806.000	.488
	Hotelling's Trace	.001	.964	14.000	39804.000	.488
	Roy's Largest Root	.001	1.481[b]	7.000	19904.000	.169
VisSOA * AVSOA	Pillai's Trace	.004	.949	84.000	39808.000	.612
	Wilks' Lambda	.996	.949[a]	84.000	39806.000	.612
	Hotelling's Trace	.004	.949	84.000	39804.000	.612
	Roy's Largest Root	.003	1.227[b]	42.000	19904.000	.149
Flank_Inter * Con_Incon * VisSOA	Pillai's Trace	.000	.948	10.000	39808.000	.487
	Wilks' Lambda	1.000	.948[a]	10.000	39806.000	.487
	Hotelling's Trace	.000	.948	10.000	39804.000	.487
	Roy's Largest Root	.000	1.485[b]	5.000	19904.000	.191
Flank_Inter * Con_Incon * AVSOA	Pillai's Trace	.001	1.007	14.000	39808.000	.442
	Wilks' Lambda	.999	1.007[a]	14.000	39806.000	.442
	Hotelling's Trace	.001	1.007	14.000	39804.000	.442
	Roy's Largest Root	.001	1.626[b]	7.000	19904.000	.123
Flank_Inter * VisSOA * AVSOA	Pillai's Trace	.001	.740	30.000	39808.000	.846
	Wilks' Lambda	.999	.740[a]	30.000	39806.000	.846
	Hotelling's Trace	.001	.740	30.000	39804.000	.846
	Roy's Largest Root	.001	1.013[b]	15.000	19904.000	.437
Con_Incon * VisSOA * AVSOA	Pillai's Trace	.005	1.127	84.000	39808.000	.201
	Wilks' Lambda	.995	1.127[a]	84.000	39806.000	.201
	Hotelling's Trace	.005	1.127	84.000	39804.000	.201
	Roy's Largest Root	.002	1.171[b]	42.000	19904.000	.208
Flank_Inter * Con_Incon * VisSOA * AVSOA	Pillai's Trace	.002	1.533	30.000	39808.000	.131
	Wilks' Lambda	.998	1.533[a]	30.000	39806.000	.131
	Hotelling's Trace	.002	1.533	30.000	39804.000	.131
	Roy's Largest Root	.001	1.655[b]	15.000	19904.000	.053

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept + Flank_Inter + Con_Incon + VisSOA + AVSOA + Flank_Inter * Con_Incon + Flank_Inter * VisSOA + Flank_Inter * AVSOA + Con_Incon * VisSOA + Con_Incon * AVSOA + VisSOA * AVSOA + Flank_Inter * Con_Incon * VisSOA + Flank_Inter * Con_Incon * AVSOA + Flank_Inter * VisSOA * AVSOA + Con_Incon * VisSOA * AVSOA + Flank_Inter * Con_Incon * VisSOA * AVSOA

Figure F2. Study 3 interactions MANOVA output



Tests of Between-Subjects Effects

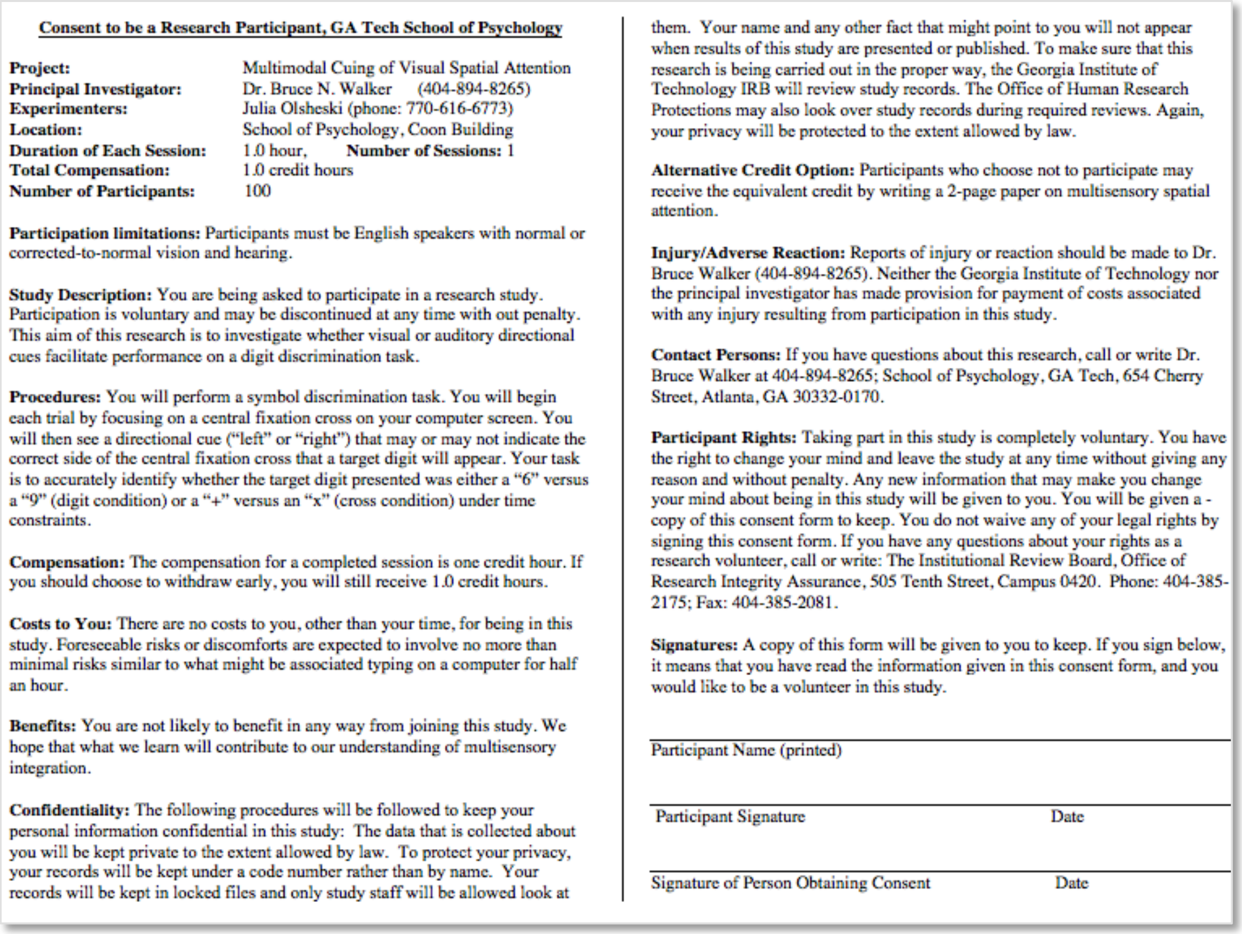
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	AccTimes100	1457654.57	167	8728.471	4.711	.000
	LogRT	8.906[b]	167	.053	4.393	.000
Intercept	AccTimes100	73672288.4	1	73672288.4	39763.574	.000
	LogRT	27599.954	1	27599.954	2273782.55	.000
Flank_Inter	AccTimes100	67542.305	1	67542.305	36.455	.000
	LogRT	.053	1	.053	4.386	.036
Con_Incon	AccTimes100	105617.985	1	105617.985	57.006	.000
	LogRT	.118	1	.118	9.760	.002
VisSOA	AccTimes100	729450.985	6	121575.164	65.618	.000
	LogRT	4.013	6	.669	55.107	.000
AVSOA	AccTimes100	17417.116	7	2488.159	1.343	.225
	LogRT	.419	7	.060	4.932	.000
Flank_Inter * Con_Incon	AccTimes100	9023.682	1	9023.682	4.870	.027
	LogRT	.009	1	.009	.719	.397
Flank_Inter * VisSOA	AccTimes100	7837.726	5	1567.545	.846	.517
	LogRT	.417	5	.083	6.871	.000
Flank_Inter * AVSOA	AccTimes100	25941.157	7	3705.880	2.000	.051
	LogRT	.063	7	.009	.737	.641
Con_Incon * VisSOA	AccTimes100	66371.625	6	11061.937	5.971	.000
	LogRT	.188	6	.031	2.588	.117
Con_Incon * AVSOA	AccTimes100	16408.373	7	2344.053	1.265	.263
	LogRT	.069	7	.010	.807	.581
VisSOA * AVSOA	AccTimes100	57679.458	42	1373.320	.741	.891
	LogRT	.614	42	.015	1.204	.172
Flank_Inter * Con_Incon * VisSOA	AccTimes100	10669.397	5	2133.879	1.152	.330
	LogRT	.036	5	.007	.601	.699
Flank_Inter * Con_Incon * AVSOA	AccTimes100	15010.502	7	2144.357	1.157	.324
	LogRT	.058	7	.008	.679	.690
Flank_Inter * VisSOA * AVSOA	AccTimes100	13435.710	15	895.714	.483	.950
	LogRT	.174	15	.012	.956	.500
Con_Incon * VisSOA * AVSOA	AccTimes100	86027.045	42	2048.263	1.106	.295
	LogRT	.578	42	.014	1.135	.254
Flank_Inter * Con_Incon * VisSOA * AVSOA	AccTimes100	45685.504	15	3045.700	1.644	.055
	LogRT	.257	15	.017	1.412	.131
Error	AccTimes100	36877299.5	19904	1852.758		
	LogRT	241.602	19904	.012		
Total	AccTimes100	1.491E8	20072			
	LogRT	39491.481	20072			
Corrected Total	AccTimes100	38334954.1	20071			
	LogRT	250.507	20071			

a. R Squared = .038 (Adjusted R Squared = .030)
b. R Squared = .036 (Adjusted R Squared = .027)

Figure F3. Study 3 tests of between-subjects effects output

APPENDIX G

Participant Materials for Study 4



Consent to be a Research Participant, GA Tech School of Psychology

Project: Multimodal Cuing of Visual Spatial Attention
Principal Investigator: Dr. Bruce N. Walker (404-894-8265)
Experimenters: Julia Olsheski (phone: 770-616-6773)
Location: School of Psychology, Coon Building
Duration of Each Session: 1.0 hour, **Number of Sessions:** 1
Total Compensation: 1.0 credit hours
Number of Participants: 100

Participation limitations: Participants must be English speakers with normal or corrected-to-normal vision and hearing.

Study Description: You are being asked to participate in a research study. Participation is voluntary and may be discontinued at any time with out penalty. This aim of this research is to investigate whether visual or auditory directional cues facilitate performance on a digit discrimination task.

Procedures: You will perform a symbol discrimination task. You will begin each trial by focusing on a central fixation cross on your computer screen. You will then see a directional cue ("left" or "right") that may or may not indicate the correct side of the central fixation cross that a target digit will appear. Your task is to accurately identify whether the target digit presented was either a "6" versus a "9" (digit condition) or a "+" versus an "x" (cross condition) under time constraints.

Compensation: The compensation for a completed session is one credit hour. If you should choose to withdraw early, you will still receive 1.0 credit hours.

Costs to You: There are no costs to you, other than your time, for being in this study. Foreseeable risks or discomforts are expected to involve no more than minimal risks similar to what might be associated typing on a computer for half an hour.

Benefits: You are not likely to benefit in any way from joining this study. We hope that what we learn will contribute to our understanding of multisensory integration.

Confidentiality: The following procedures will be followed to keep your personal information confidential in this study: The data that is collected about you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code number rather than by name. Your records will be kept in locked files and only study staff will be allowed look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB will review study records. The Office of Human Research Protections may also look over study records during required reviews. Again, your privacy will be protected to the extent allowed by law.

Alternative Credit Option: Participants who choose not to participate may receive the equivalent credit by writing a 2-page paper on multisensory spatial attention.

Injury/Adverse Reaction: Reports of injury or reaction should be made to Dr. Bruce Walker (404-894-8265). Neither the Georgia Institute of Technology nor the principal investigator has made provision for payment of costs associated with any injury resulting from participation in this study.

Contact Persons: If you have questions about this research, call or write Dr. Bruce Walker at 404-894-8265; School of Psychology, GA Tech, 654 Cherry Street, Atlanta, GA 30332-0170.

Participant Rights: Taking part in this study is completely voluntary. You have the right to change your mind and leave the study at any time without giving any reason and without penalty. Any new information that may make you change your mind about being in this study will be given to you. You will be given a - copy of this consent form to keep. You do not waive any of your legal rights by signing this consent form. If you have any questions about your rights as a research volunteer, call or write: The Institutional Review Board, Office of Research Integrity Assurance, 505 Tenth Street, Campus 0420. Phone: 404-385-2175; Fax: 404-385-2081.

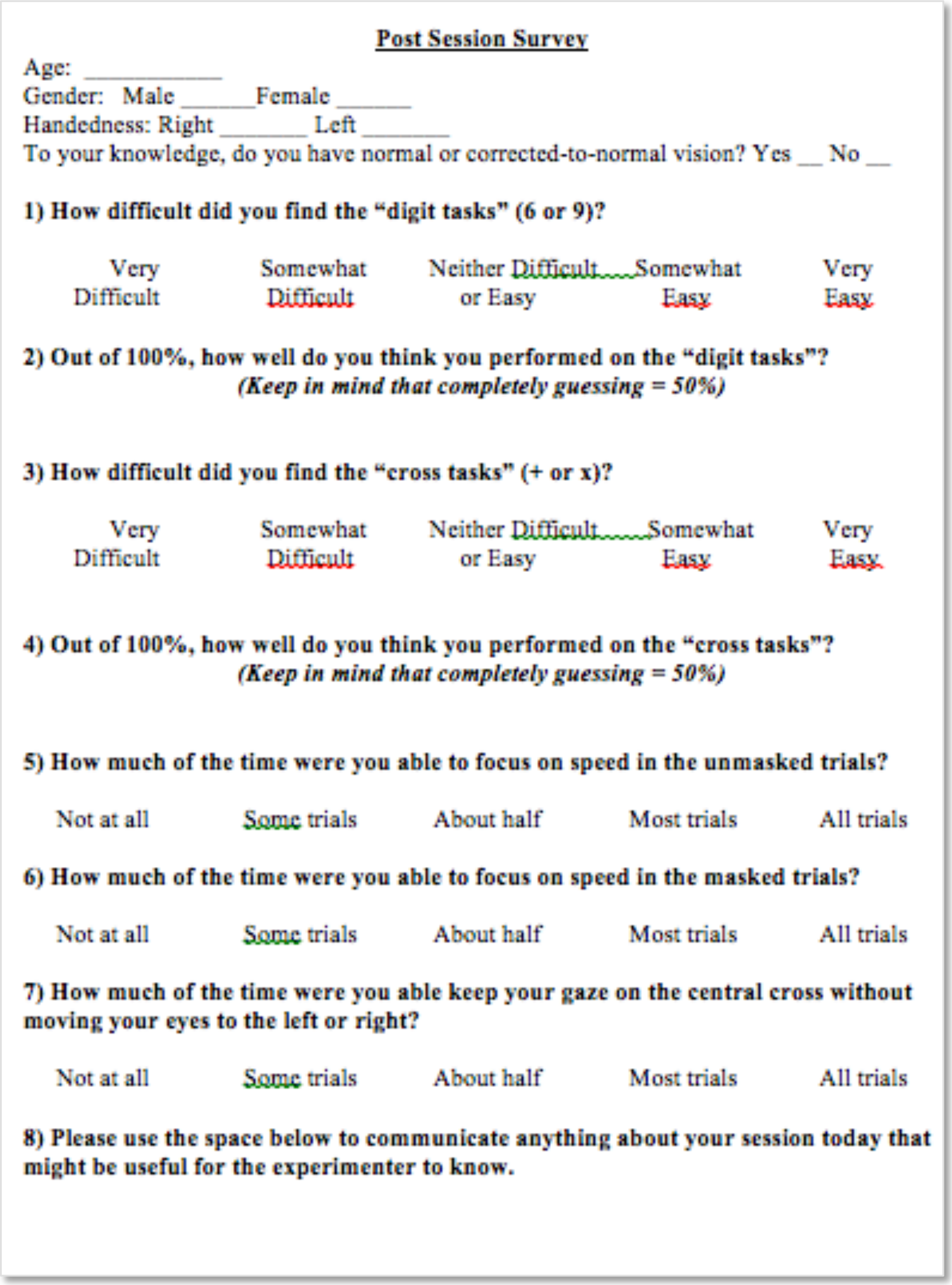
Signatures: A copy of this form will be given to you to keep. If you sign below, it means that you have read the information given in this consent form, and you would like to be a volunteer in this study.

Participant Name (printed)

Participant Signature Date

Signature of Person Obtaining Consent Date

Figure G1. *Consent form for participants in Study 4.*



Post Session Survey

Age: ____________

Gender: Male ______Female ______

Handedness: Right ________ Left ________

To your knowledge, do you have normal or corrected-to-normal vision? Yes __ No __

1) How difficult did you find the "digit tasks" (6 or 9)?

Very Difficult	Somewhat Difficult	Neither Difficult or Easy	Somewhat Easy	Very Easy

2) Out of 100%, how well do you think you performed on the "digit tasks"?
(Keep in mind that completely guessing = 50%)

3) How difficult did you find the "cross tasks" (+ or x)?

Very Difficult	Somewhat Difficult	Neither Difficult or Easy	Somewhat Easy	Very Easy

4) Out of 100%, how well do you think you performed on the "cross tasks"?
(Keep in mind that completely guessing = 50%)

5) How much of the time were you able to focus on speed in the unmasked trials?

Not at all	Some trials	About half	Most trials	All trials

6) How much of the time were you able to focus on speed in the masked trials?

Not at all	Some trials	About half	Most trials	All trials

7) How much of the time were you able keep your gaze on the central cross without moving your eyes to the left or right?

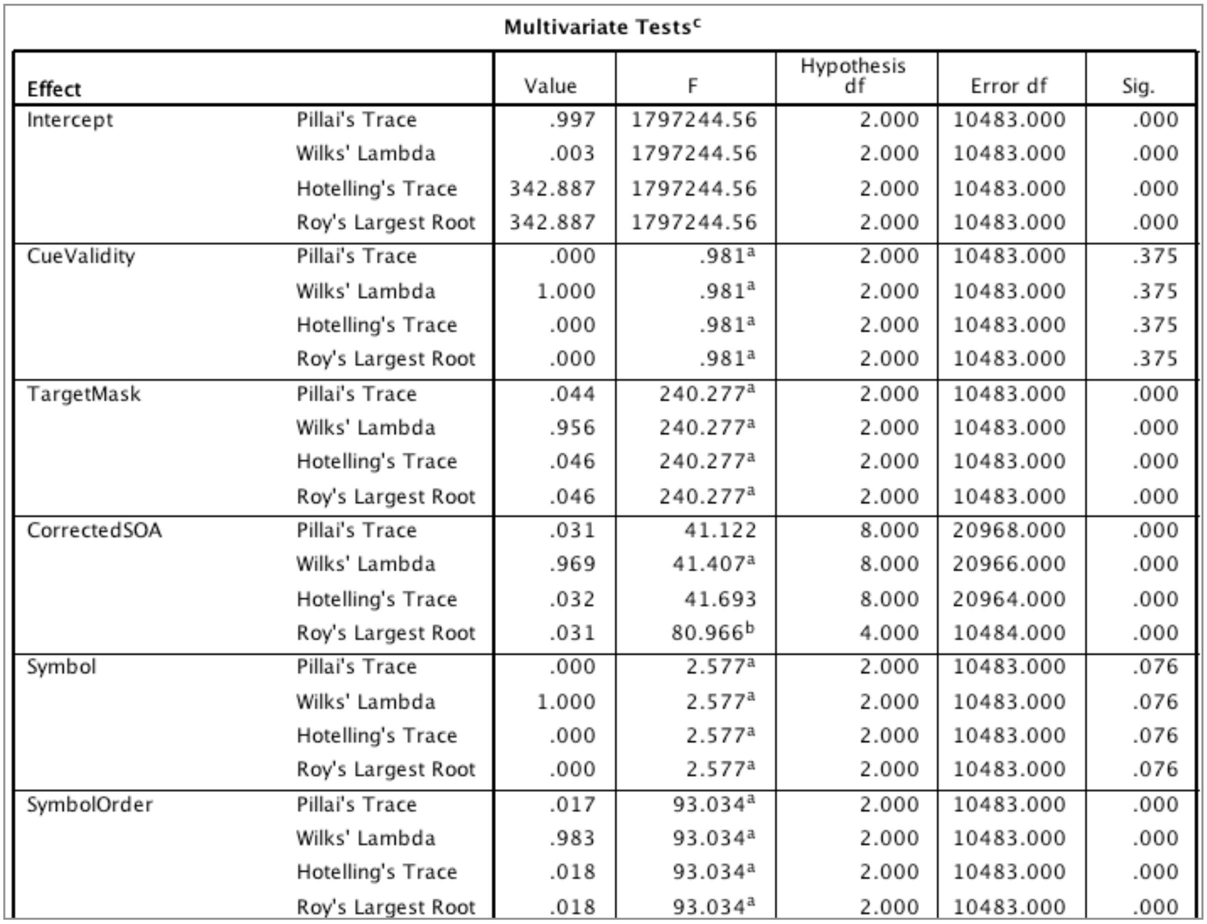
Not at all	Some trials	About half	Most trials	All trials

8) Please use the space below to communicate anything about your session today that might be useful for the experimenter to know.

Figure G2. Post-session survey for participants in Study 4.

APPENDIX H

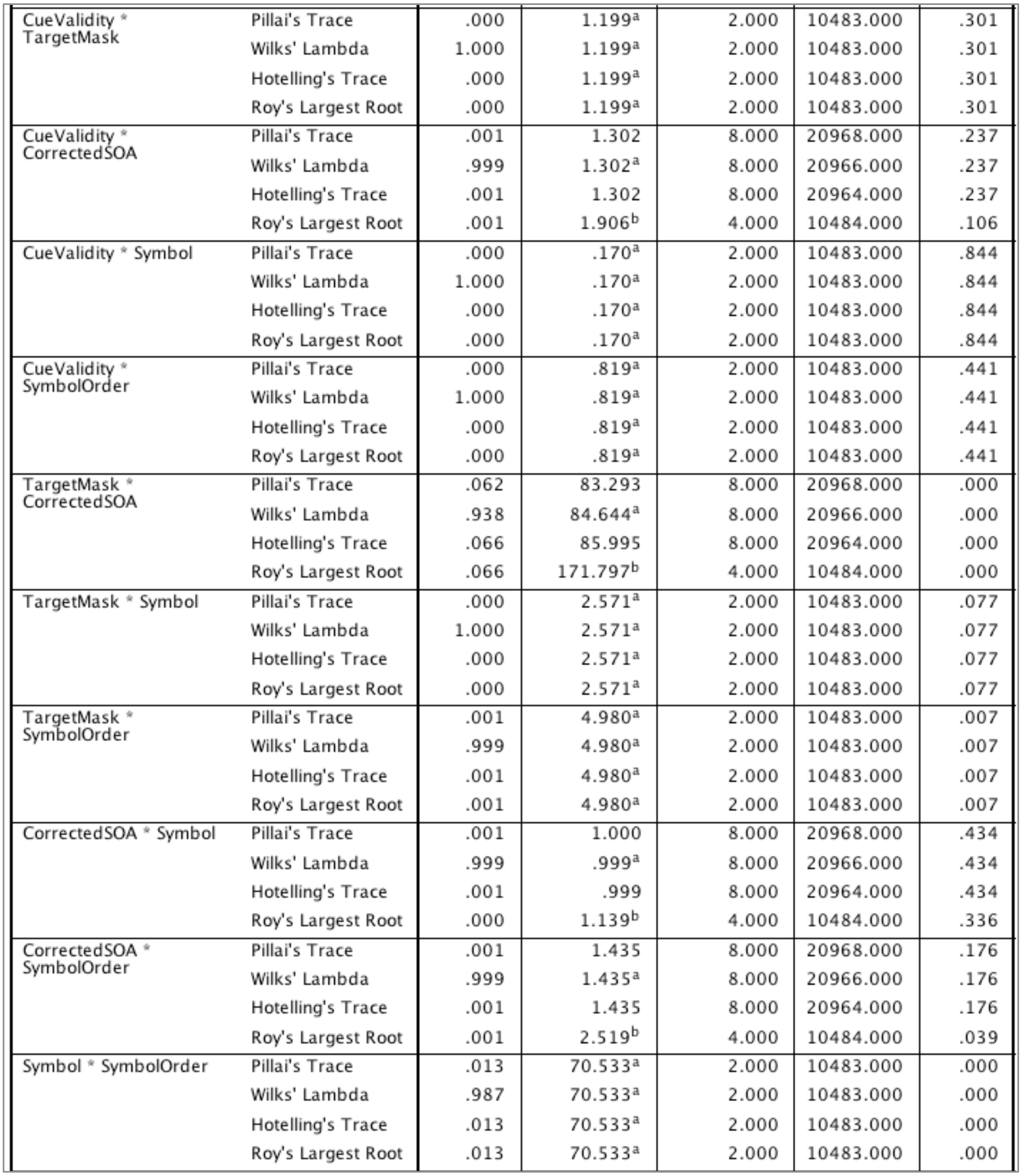
Study 4 Data Analysis Tables



Multivariate Tests[c]

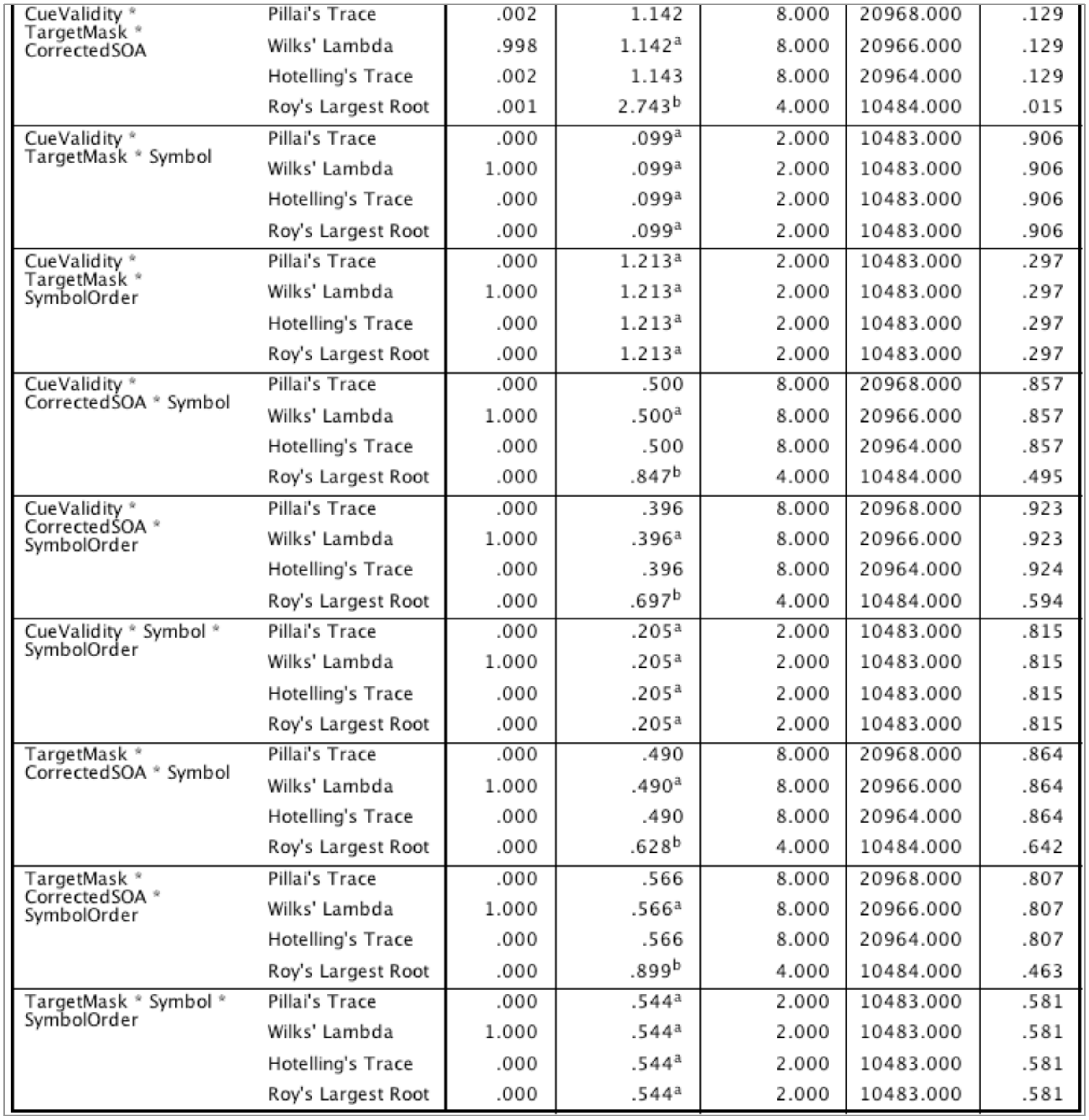
Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.997	1797244.56	2.000	10483.000	.000
	Wilks' Lambda	.003	1797244.56	2.000	10483.000	.000
	Hotelling's Trace	342.887	1797244.56	2.000	10483.000	.000
	Roy's Largest Root	342.887	1797244.56	2.000	10483.000	.000
CueValidity	Pillai's Trace	.000	.981[a]	2.000	10483.000	.375
	Wilks' Lambda	1.000	.981[a]	2.000	10483.000	.375
	Hotelling's Trace	.000	.981[a]	2.000	10483.000	.375
	Roy's Largest Root	.000	.981[a]	2.000	10483.000	.375
TargetMask	Pillai's Trace	.044	240.277[a]	2.000	10483.000	.000
	Wilks' Lambda	.956	240.277[a]	2.000	10483.000	.000
	Hotelling's Trace	.046	240.277[a]	2.000	10483.000	.000
	Roy's Largest Root	.046	240.277[a]	2.000	10483.000	.000
CorrectedSOA	Pillai's Trace	.031	41.122	8.000	20968.000	.000
	Wilks' Lambda	.969	41.407[a]	8.000	20966.000	.000
	Hotelling's Trace	.032	41.693	8.000	20964.000	.000
	Roy's Largest Root	.031	80.966[b]	4.000	10484.000	.000
Symbol	Pillai's Trace	.000	2.577[a]	2.000	10483.000	.076
	Wilks' Lambda	1.000	2.577[a]	2.000	10483.000	.076
	Hotelling's Trace	.000	2.577[a]	2.000	10483.000	.076
	Roy's Largest Root	.000	2.577[a]	2.000	10483.000	.076
SymbolOrder	Pillai's Trace	.017	93.034[a]	2.000	10483.000	.000
	Wilks' Lambda	.983	93.034[a]	2.000	10483.000	.000
	Hotelling's Trace	.018	93.034[a]	2.000	10483.000	.000
	Roy's Largest Root	.018	93.034[a]	2.000	10483.000	.000

Figure H1. Study 4 MANOVA results for main effects.



CueValidity * TargetMask	Pillai's Trace	.000	1.199[a]	2.000	10483.000	.301
	Wilks' Lambda	1.000	1.199[a]	2.000	10483.000	.301
	Hotelling's Trace	.000	1.199[a]	2.000	10483.000	.301
	Roy's Largest Root	.000	1.199[a]	2.000	10483.000	.301
CueValidity * CorrectedSOA	Pillai's Trace	.001	1.302	8.000	20968.000	.237
	Wilks' Lambda	.999	1.302[a]	8.000	20966.000	.237
	Hotelling's Trace	.001	1.302	8.000	20964.000	.237
	Roy's Largest Root	.001	1.906[b]	4.000	10484.000	.106
CueValidity * Symbol	Pillai's Trace	.000	.170[a]	2.000	10483.000	.844
	Wilks' Lambda	1.000	.170[a]	2.000	10483.000	.844
	Hotelling's Trace	.000	.170[a]	2.000	10483.000	.844
	Roy's Largest Root	.000	.170[a]	2.000	10483.000	.844
CueValidity * SymbolOrder	Pillai's Trace	.000	.819[a]	2.000	10483.000	.441
	Wilks' Lambda	1.000	.819[a]	2.000	10483.000	.441
	Hotelling's Trace	.000	.819[a]	2.000	10483.000	.441
	Roy's Largest Root	.000	.819[a]	2.000	10483.000	.441
TargetMask * CorrectedSOA	Pillai's Trace	.062	83.293	8.000	20968.000	.000
	Wilks' Lambda	.938	84.644[a]	8.000	20966.000	.000
	Hotelling's Trace	.066	85.995	8.000	20964.000	.000
	Roy's Largest Root	.066	171.797[b]	4.000	10484.000	.000
TargetMask * Symbol	Pillai's Trace	.000	2.571[a]	2.000	10483.000	.077
	Wilks' Lambda	1.000	2.571[a]	2.000	10483.000	.077
	Hotelling's Trace	.000	2.571[a]	2.000	10483.000	.077
	Roy's Largest Root	.000	2.571[a]	2.000	10483.000	.077
TargetMask * SymbolOrder	Pillai's Trace	.001	4.980[a]	2.000	10483.000	.007
	Wilks' Lambda	.999	4.980[a]	2.000	10483.000	.007
	Hotelling's Trace	.001	4.980[a]	2.000	10483.000	.007
	Roy's Largest Root	.001	4.980[a]	2.000	10483.000	.007
CorrectedSOA * Symbol	Pillai's Trace	.001	1.000	8.000	20968.000	.434
	Wilks' Lambda	.999	.999[a]	8.000	20966.000	.434
	Hotelling's Trace	.001	.999	8.000	20964.000	.434
	Roy's Largest Root	.000	1.139[b]	4.000	10484.000	.336
CorrectedSOA * SymbolOrder	Pillai's Trace	.001	1.435	8.000	20968.000	.176
	Wilks' Lambda	.999	1.435[a]	8.000	20966.000	.176
	Hotelling's Trace	.001	1.435	8.000	20964.000	.176
	Roy's Largest Root	.001	2.519[b]	4.000	10484.000	.039
Symbol * SymbolOrder	Pillai's Trace	.013	70.533[a]	2.000	10483.000	.000
	Wilks' Lambda	.987	70.533[a]	2.000	10483.000	.000
	Hotelling's Trace	.013	70.533[a]	2.000	10483.000	.000
	Roy's Largest Root	.013	70.533[a]	2.000	10483.000	.000

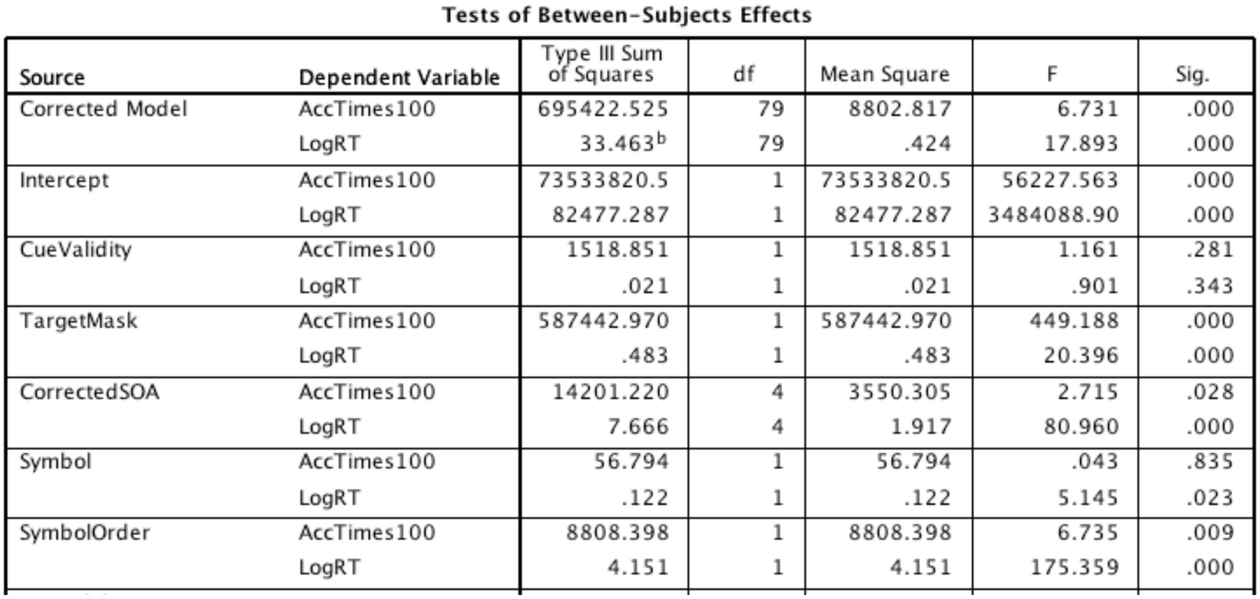
Figure H2. Study 4 MANOVA results for two-way interactions.



CueValidity * TargetMask * CorrectedSOA	Pillai's Trace	.002	1.142	8.000	20968.000	.129
	Wilks' Lambda	.998	1.142[a]	8.000	20966.000	.129
	Hotelling's Trace	.002	1.143	8.000	20964.000	.129
	Roy's Largest Root	.001	2.743[b]	4.000	10484.000	.015
CueValidity * TargetMask * Symbol	Pillai's Trace	.000	.099[a]	2.000	10483.000	.906
	Wilks' Lambda	1.000	.099[a]	2.000	10483.000	.906
	Hotelling's Trace	.000	.099[a]	2.000	10483.000	.906
	Roy's Largest Root	.000	.099[a]	2.000	10483.000	.906
CueValidity * TargetMask * SymbolOrder	Pillai's Trace	.000	1.213[a]	2.000	10483.000	.297
	Wilks' Lambda	1.000	1.213[a]	2.000	10483.000	.297
	Hotelling's Trace	.000	1.213[a]	2.000	10483.000	.297
	Roy's Largest Root	.000	1.213[a]	2.000	10483.000	.297
CueValidity * CorrectedSOA * Symbol	Pillai's Trace	.000	.500	8.000	20968.000	.857
	Wilks' Lambda	1.000	.500[a]	8.000	20966.000	.857
	Hotelling's Trace	.000	.500	8.000	20964.000	.857
	Roy's Largest Root	.000	.847[b]	4.000	10484.000	.495
CueValidity * CorrectedSOA * SymbolOrder	Pillai's Trace	.000	.396	8.000	20968.000	.923
	Wilks' Lambda	1.000	.396[a]	8.000	20966.000	.923
	Hotelling's Trace	.000	.396	8.000	20964.000	.924
	Roy's Largest Root	.000	.697[b]	4.000	10484.000	.594
CueValidity * Symbol * SymbolOrder	Pillai's Trace	.000	.205[a]	2.000	10483.000	.815
	Wilks' Lambda	1.000	.205[a]	2.000	10483.000	.815
	Hotelling's Trace	.000	.205[a]	2.000	10483.000	.815
	Roy's Largest Root	.000	.205[a]	2.000	10483.000	.815
TargetMask * CorrectedSOA * Symbol	Pillai's Trace	.000	.490	8.000	20968.000	.864
	Wilks' Lambda	1.000	.490[a]	8.000	20966.000	.864
	Hotelling's Trace	.000	.490	8.000	20964.000	.864
	Roy's Largest Root	.000	.628[b]	4.000	10484.000	.642
TargetMask * CorrectedSOA * SymbolOrder	Pillai's Trace	.000	.566	8.000	20968.000	.807
	Wilks' Lambda	1.000	.566[a]	8.000	20966.000	.807
	Hotelling's Trace	.000	.566	8.000	20964.000	.807
	Roy's Largest Root	.000	.899[b]	4.000	10484.000	.463
TargetMask * Symbol * SymbolOrder	Pillai's Trace	.000	.544[a]	2.000	10483.000	.581
	Wilks' Lambda	1.000	.544[a]	2.000	10483.000	.581
	Hotelling's Trace	.000	.544[a]	2.000	10483.000	.581
	Roy's Largest Root	.000	.544[a]	2.000	10483.000	.581

Figure H3. Study 4 MANOVA results for three-way interactions

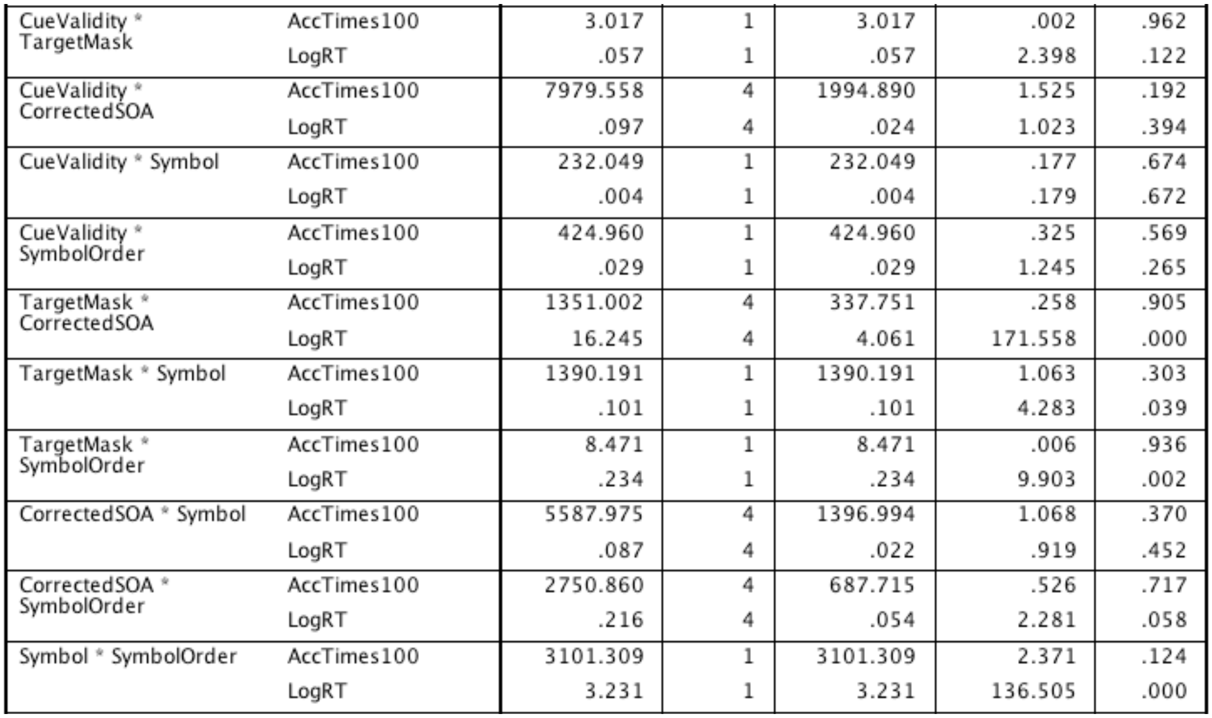
.



Tests of Between-Subjects Effects

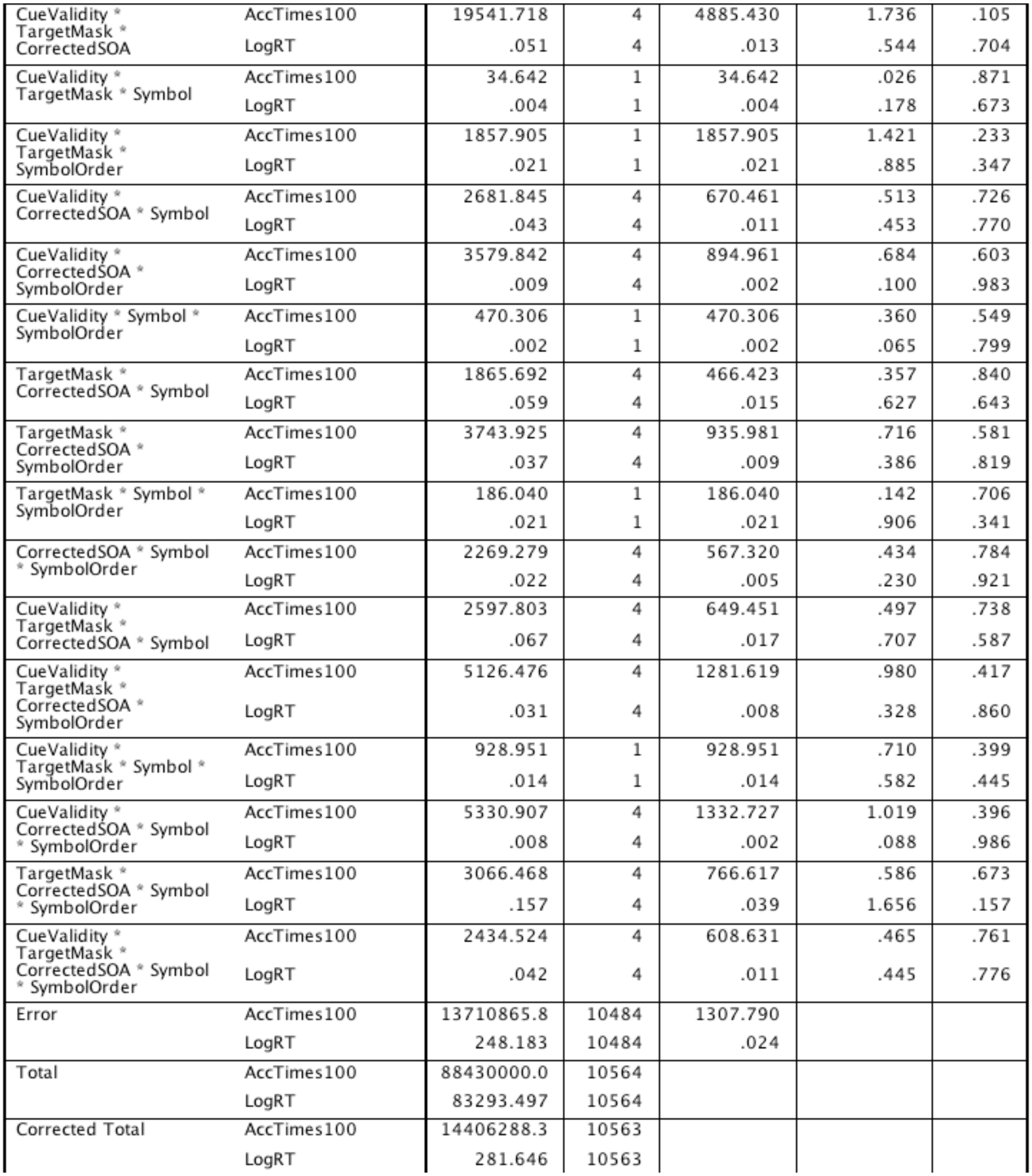
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	AccTimes100	695422.525	79	8802.817	6.731	.000
	LogRT	33.463^{b}	79	.424	17.893	.000
Intercept	AccTimes100	73533820.5	1	73533820.5	56227.563	.000
	LogRT	82477.287	1	82477.287	3484088.90	.000
CueValidity	AccTimes100	1518.851	1	1518.851	1.161	.281
	LogRT	.021	1	.021	.901	.343
TargetMask	AccTimes100	587442.970	1	587442.970	449.188	.000
	LogRT	.483	1	.483	20.396	.000
CorrectedSOA	AccTimes100	14201.220	4	3550.305	2.715	.028
	LogRT	7.666	4	1.917	80.960	.000
Symbol	AccTimes100	56.794	1	56.794	.043	.835
	LogRT	.122	1	.122	5.145	.023
SymbolOrder	AccTimes100	8808.398	1	8808.398	6.735	.009
	LogRT	4.151	1	4.151	175.359	.000

Figure H4. Study 4 tests of between-subjects effects for main effects.



CueValidity * TargetMask	AccTimes100	3.017	1	3.017	.002	.962
	LogRT	.057	1	.057	2.398	.122
CueValidity * CorrectedSOA	AccTimes100	7979.558	4	1994.890	1.525	.192
	LogRT	.097	4	.024	1.023	.394
CueValidity * Symbol	AccTimes100	232.049	1	232.049	.177	.674
	LogRT	.004	1	.004	.179	.672
CueValidity * SymbolOrder	AccTimes100	424.960	1	424.960	.325	.569
	LogRT	.029	1	.029	1.245	.265
TargetMask * CorrectedSOA	AccTimes100	1351.002	4	337.751	.258	.905
	LogRT	16.245	4	4.061	171.558	.000
TargetMask * Symbol	AccTimes100	1390.191	1	1390.191	1.063	.303
	LogRT	.101	1	.101	4.283	.039
TargetMask * SymbolOrder	AccTimes100	8.471	1	8.471	.006	.936
	LogRT	.234	1	.234	9.903	.002
CorrectedSOA * Symbol	AccTimes100	5587.975	4	1396.994	1.068	.370
	LogRT	.087	4	.022	.919	.452
CorrectedSOA * SymbolOrder	AccTimes100	2750.860	4	687.715	.526	.717
	LogRT	.216	4	.054	2.281	.058
Symbol * SymbolOrder	AccTimes100	3101.309	1	3101.309	2.371	.124
	LogRT	3.231	1	3.231	136.505	.000

Figure H5. Study 4 tests of between-subjects effects for two-way interactions.

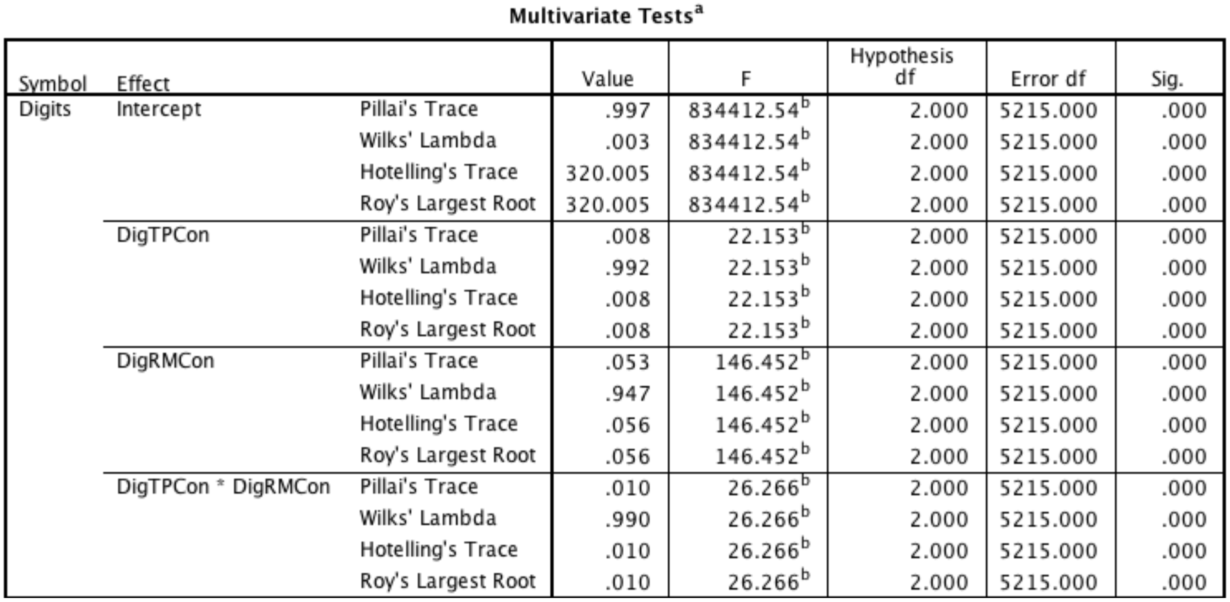


CueValidity * TargetMask * CorrectedSOA	AccTimes100	19541.718	4	4885.430	1.736	.105
	LogRT	.051	4	.013	.544	.704
CueValidity * TargetMask * Symbol	AccTimes100	34.642	1	34.642	.026	.871
	LogRT	.004	1	.004	.178	.673
CueValidity * TargetMask * SymbolOrder	AccTimes100	1857.905	1	1857.905	1.421	.233
	LogRT	.021	1	.021	.885	.347
CueValidity * CorrectedSOA * Symbol	AccTimes100	2681.845	4	670.461	.513	.726
	LogRT	.043	4	.011	.453	.770
CueValidity * CorrectedSOA * SymbolOrder	AccTimes100	3579.842	4	894.961	.684	.603
	LogRT	.009	4	.002	.100	.983
CueValidity * Symbol * SymbolOrder	AccTimes100	470.306	1	470.306	.360	.549
	LogRT	.002	1	.002	.065	.799
TargetMask * CorrectedSOA * Symbol	AccTimes100	1865.692	4	466.423	.357	.840
	LogRT	.059	4	.015	.627	.643
TargetMask * CorrectedSOA * SymbolOrder	AccTimes100	3743.925	4	935.981	.716	.581
	LogRT	.037	4	.009	.386	.819
TargetMask * Symbol * SymbolOrder	AccTimes100	186.040	1	186.040	.142	.706
	LogRT	.021	1	.021	.906	.341
CorrectedSOA * Symbol * SymbolOrder	AccTimes100	2269.279	4	567.320	.434	.784
	LogRT	.022	4	.005	.230	.921
CueValidity * TargetMask * CorrectedSOA * Symbol	AccTimes100	2597.803	4	649.451	.497	.738
	LogRT	.067	4	.017	.707	.587
CueValidity * TargetMask * CorrectedSOA * SymbolOrder	AccTimes100	5126.476	4	1281.619	.980	.417
	LogRT	.031	4	.008	.328	.860
CueValidity * TargetMask * Symbol * SymbolOrder	AccTimes100	928.951	1	928.951	.710	.399
	LogRT	.014	1	.014	.582	.445
CueValidity * CorrectedSOA * Symbol * SymbolOrder	AccTimes100	5330.907	4	1332.727	1.019	.396
	LogRT	.008	4	.002	.088	.986
TargetMask * CorrectedSOA * Symbol * SymbolOrder	AccTimes100	3066.468	4	766.617	.586	.673
	LogRT	.157	4	.039	1.656	.157
CueValidity * TargetMask * CorrectedSOA * Symbol * SymbolOrder	AccTimes100	2434.524	4	608.631	.465	.761
	LogRT	.042	4	.011	.445	.776
Error	AccTimes100	13710865.8	10484	1307.790		
	LogRT	248.183	10484	.024		
Total	AccTimes100	88430000.0	10564			
	LogRT	83293.497	10564			
Corrected Total	AccTimes100	14406288.3	10563			
	LogRT	281.646	10563			

Figure H6. Study 4 Test of between-subjects effects for three-way interactions.

APPENDIX I

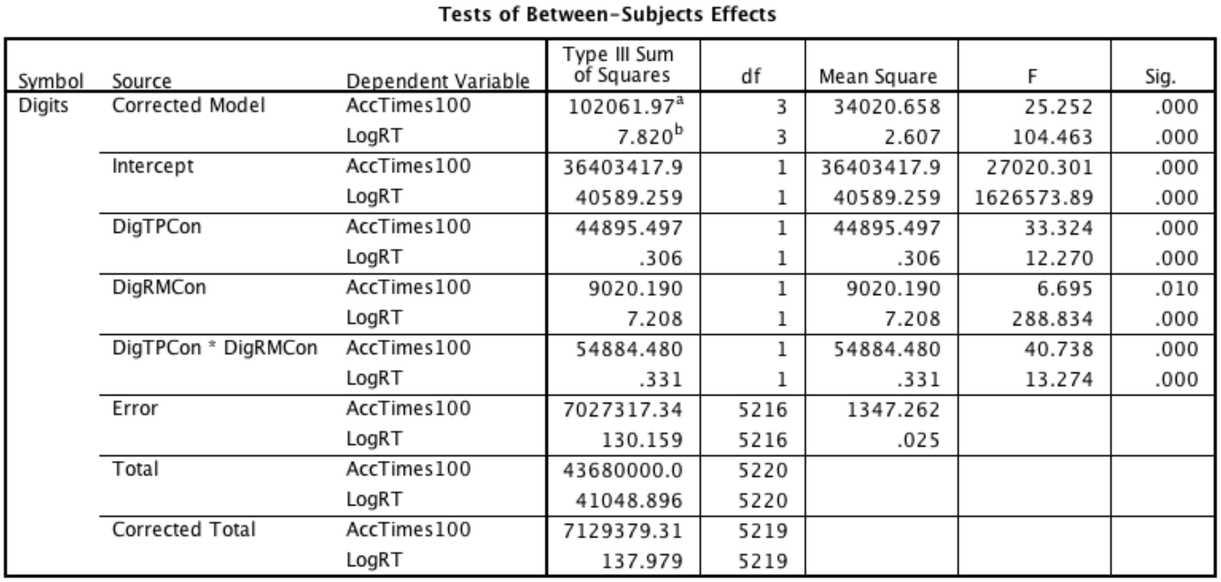
Study 4 Digit Trials Data Analysis Tables.



Multivariate Tests[a]

Symbol	Effect		Value	F	Hypothesis df	Error df	Sig.
Digits	Intercept	Pillai's Trace	.997	834412.54[b]	2.000	5215.000	.000
		Wilks' Lambda	.003	834412.54[b]	2.000	5215.000	.000
		Hotelling's Trace	320.005	834412.54[b]	2.000	5215.000	.000
		Roy's Largest Root	320.005	834412.54[b]	2.000	5215.000	.000
	DigTPCon	Pillai's Trace	.008	22.153[b]	2.000	5215.000	.000
		Wilks' Lambda	.992	22.153[b]	2.000	5215.000	.000
		Hotelling's Trace	.008	22.153[b]	2.000	5215.000	.000
		Roy's Largest Root	.008	22.153[b]	2.000	5215.000	.000
	DigRMCon	Pillai's Trace	.053	146.452[b]	2.000	5215.000	.000
		Wilks' Lambda	.947	146.452[b]	2.000	5215.000	.000
		Hotelling's Trace	.056	146.452[b]	2.000	5215.000	.000
		Roy's Largest Root	.056	146.452[b]	2.000	5215.000	.000
	DigTPCon * DigRMCon	Pillai's Trace	.010	26.266[b]	2.000	5215.000	.000
		Wilks' Lambda	.990	26.266[b]	2.000	5215.000	.000
		Hotelling's Trace	.010	26.266[b]	2.000	5215.000	.000
		Roy's Largest Root	.010	26.266[b]	2.000	5215.000	.000

Figure I1. Study 4 MANOVA output for Digit trials.



Tests of Between-Subjects Effects

Symbol	Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Digits	Corrected Model	AccTimes100	102061.97[a]	3	34020.658	25.252	.000
		LogRT	7.820[b]	3	2.607	104.463	.000
	Intercept	AccTimes100	36403417.9	1	36403417.9	27020.301	.000
		LogRT	40589.259	1	40589.259	1626573.89	.000
	DigTPCon	AccTimes100	44895.497	1	44895.497	33.324	.000
		LogRT	.306	1	.306	12.270	.000
	DigRMCon	AccTimes100	9020.190	1	9020.190	6.695	.010
		LogRT	7.208	1	7.208	288.834	.000
	DigTPCon * DigRMCon	AccTimes100	54884.480	1	54884.480	40.738	.000
		LogRT	.331	1	.331	13.274	.000
	Error	AccTimes100	7027317.34	5216	1347.262		
		LogRT	130.159	5216	.025		
	Total	AccTimes100	43680000.0	5220			
		LogRT	41048.896	5220			
	Corrected Total	AccTimes100	7129379.31	5219			
		LogRT	137.979	5219			

Figure I2. Study 4 tests of between-subjects effects for Digit trials.