From Green to Great Dane:
How Multisensory Integration Enhances Working Memory

By

Radhika S. Gosavi

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Educational Psychology)

at the
UNIVERSITY OF WISCONSIN-MADISON
2019

Date of final oral examination: 05/13/2019

The dissertation is approved by the following members of the Final Oral Committee:

Edward M. Hubbard, Assistant Professor, Educational Psychology
Charles Kalish, Professor, Educational Psychology
Haley Vlach, Associate Professor, Educational Psychology
Ferrinne Spector, Associate Professor, Psychology

To my grandparents,

Vasant, Sadhana, Narayan, and Kusum,

who persevered against all odds in the pursuit of education

Table of Contents

Abstract

A large portion of our experience combines information from multiple sensory modalities. The ability to process an abundance of information, and to identify the relevant pieces is highly dependent on mechanisms of multisensory integration (MSI) (Stein & Meredith, 1993). This dissertation focused on audiovisual (AV) MSI, in which sensory information is combined from the auditory and visual modalities to create one, coherent, multisensory percept. While simple stimuli such as flashes and beeps have formed the basis of many past experiments, more recently, studies have started advocating for the use of more naturalistic stimuli, which mimic the complexities present in the environment (Stevenson & Wallace, 2013). Studies have shown that MSI improves performance on a range of tasks from basic perceptual tasks to cognitive tasks involving learning (Newell, Mamassian, & Alais, 2010; Shams & Seitz, 2008). Although past studies have linked MSI and mnemonic processes, studies of MSI and working memory remain sparse. In a series of studies, this dissertation aimed to answer the following outstanding questions: Do multisensory stimuli provide a benefit for working memory processing? Are the auditory and visual modalities linked differently based on stimulus properties? What are the effects of crossmodal congruency on MSI and working memory? To address these questions, we tested the impact of simple and complex multisensory stimuli on working memory in adults using a change detection paradigm. We found enhancements in visual and auditory working memory when presented with multisensory stimuli. However, the attentional demands of the task, complexity of stimuli presented, crossmodal congruency, type of working memory (visual or auditory) tested, and memory load presented mediate the nature of this enhancement. The findings in this thesis make a contribution to the growing scientific literature in the domains of multisensory integration and memory, and also have implications for educational practices.

Acknowledgments

First of all, I would like to thank my advisor and mentor, Dr. Edward Hubbard, who has been the most instrumental person to my success in graduate school. Ed, thank you for sharing your love for science, and for guiding me in every endeavor, from research, to teaching and writing. Your path to academia is inspiring, and your enthusiasm is infectious. Without your encouragement and support, the long nights spent in the lab, and months of data collection, would not have been possible. Thank you for always having my back, and my best interests at heart. I also want to thank my dissertation committee members: your expertise and excellent suggestions have strengthened the contribution and rigor of the studies included in this dissertation.

V.S. Ramachandran, or Rama, was one of the first people I informed when I was admitted to graduate school at UW-Madison. I received an email back saying, "Congrats on becoming a *grandstudent*". I worked with Rama as an undergraduate research assistant at UCSD, and Ed received his PhD at UCSD with Rama as his advisor. Who would've thought that my decision to join Rama's lab as a freshman in college would introduce me to my PhD advisor/mentor, the topic of my dissertation, and lead me to move to Wisconsin?! I want to thank Rama for giving me my first chance to be a part of a research lab and for introducing me to the beauty of scientific inquiry with deep questions and simple, yet, elegant methodologies.

I want to thank the members of the Educational Neuroscience Lab, especially Liz, John, Bella, Katherine, Priya, Zac, Mia, and Irene. Thank you for reading numerous drafts of my writing, and always being around to share ideas, frustrations, and good times. I have been lucky to have an excellent team of undergraduate research assistants throughout grad school. I'd like to thank them for spending dozens of hours in windowless rooms collecting data (sometimes on a very ambitious timeline) and helping me develop my mentorship abilities. I also want to thank the faculty members of the department of Educational Psychology for giving me the opportunity to teach large undergraduate classes. This experience has been crucial in my development as a science communicator, researcher, and instructor.

Thank you to my friends, who have been unwavering pillars of support throughout this journey. To my long-distance friends: Thank you for being available to chat on the phone at all hours of the day and sharing a good laugh when I needed it the most. Additionally, thank you for coming to visit me in Madison, and doing wholesome midwestern activities with me. To my friends in Madison: Thank you for always being willing to go on late night ice-cream runs, to burn the latest frustrations by working out, to spend fabulous evenings at the terrace, and to motivate each other in our writing sessions (shout out to all the amazing coffeeshops in Madison).

Last but not least, I want to thank my family for everything that they have done for me. To my parents: thank you for raising me with a love for science, encouraging me at my lowest, celebrating my successes, and supporting me in all my endeavors. To my grandparents: thank you for paving the path for the generations to come by sharing your passion for education, and social service.

This journey would've been incomplete without every single one of you by my side. I want to thank you from the bottom of my heart, I couldn't have done this without you!

Chapter 1: Introduction & General Background

A large portion of our experience of the world combines information from multiple sensory modalities. We are rarely presented with information about the world that is unisensory, or only involving information from one sensory modality. Perhaps this is why humans and other animals have developed specialized behavioral and neural systems to use multisensory experiences to help make sense of the environment. For example, as a deer stands in a field, processing various sights and sounds, it has to be able to successfully combine the sight and sound of a predator to detect a possible threat. Similarly, humans have to be able to match the sound of a siren with the sight of an ambulance to take the appropriate action in a crowded city square. The ability to process a multitude of sensory stimuli to identify a relevant and salient stimulus is highly dependent on mechanisms of multisensory integration (MSI). In this thesis I focus on audiovisual (AV) MSI, in which sensory information is combined from the auditory and visual sensory modalities.

Although a great deal of research has examined the impact of AV MSI on perception, the implications for cognition have been relatively understudied. In this thesis, I aim to elucidate the importance of multisensory integration by discussing key components that determine its success, and, critically, how AV MSI impacts memory (especially working memory). Beyond the immediate empirical contributions, this thesis also serves as part of an emerging understanding of multisensory memory as influenced by stimulus properties, and as integrated with perceptual and neural systems, rather than being an encapsulated cognitive process (e.g., D'Esposito & Postle, 2015; Wheeler, Petersen, & Buckner, 2000). In the remainder of Chapter 1, I will review multisensory integration processes through a cognitive neuroscience perspective and highlight debates regarding the relationship of multisensory processing and memory in the auditory and visual modalities.

Principles of Multisensory Integration in Perception

In the last two decades, the rate of studies investigating how visual and auditory information are integrated has accelerated. This research boom has dramatically advanced our knowledge of MSI in the audiovisual (AV) domain through behavioral and neural studies (Newell, Mamassian, & Alais, 2010; Stein & Meredith, 1993). These studies have demonstrated a number of "rules" or principles that guide multisensory integration in perception. Here, I review four critical principles of MSI, with a focus on AV MSI: 1) modality appropriateness, 2) temporal and spatial proximity, 3) the impact of stimulus complexity on MSI, and 4) the role of cross-modal correspondences on MSI.

Modality Dominance vs. Modality Appropriateness

Traditionally, visual information has been thought be more "dominant" and salient than information from other sensory modalities (Rock & Victor, 1964; Welch & Warren, 1986). Rock and Victor (1964) asked participants to observe objects using touch and vision. Participants wore distorting prisms, which made the objects appear longer or taller than they actually were, in conflict with how the object felt to the touch. Based on the participants' ability to draw the object or match it to another object depending on touch, Rock and Victor concluded that the distorted visual information dominated over the veridical tactile information. However, more recent evidence demonstrates that audition is superior to vision in temporally mediated tasks (Repp & Penel, 2002, 2004). In a perceptual judgement task, participants were asked to tap their finger with auditory and visual sequences containing event onset shifts (Repp & Penel, 2002). Performance was much poorer in the visual domain compared to the auditory domain, leading researchers to conclude that auditory sensory modality is dominant for temporally-dependent tasks.

These diverging examples of dominance for the visual and auditory modality led researchers to propose the Modality Appropriateness Hypothesis: the influence of each sensory modality depends how appropriate that modality is for the given task (Welch & Warren, 1980). More specifically, the visual modality has higher spatial precision, making it more appropriate for spatial tasks, whereas the auditory modality has higher temporal precision, making it more appropriate for tasks involving judgments of time. This is further demonstrated by the Ventriloquist Illusion, a well-known effect that is modulated by visual and auditory cues. Sound is perceived as originating from a different source due to the perception of a visual stimulus despite a spatial discrepancy in the sound source and visual stimulus (Howard & Templeton, 1966). This effect is regularly experienced in movie theaters where speech is perceived as coming from the actors on the screen due to their lip movements rather than the original source: the speakers. The neural correlates of this effect are found in the planum temporale of the auditory cortex, with larger neural responses in the hemisphere contralateral to the visual stimulus or perceived auditory source (Bonath et al., 2007).

While one sense can be dominant in some instances, studies of MSI have also demonstrated the bidirectional influence of visual and auditory information. One widely cited effect of MSI is the McGurk Effect (McGurk & MacDonald, 1976). The McGurk Effect is a perceptual phenomenon in which the visual component of speech dramatically affects the auditory perception of speech. More specifically, this effect shows the averaging or blending of the senses rather than a dominance of the visual over the auditory sensory modality. The visual stimulus presented is either congruent or incongruent with the auditory stimulus, while the auditory stimulus remains constant across conditions. In the original study, participants were first presented with a video of another individual mouthing the syllable “ba-ba” accompanied by the congruent auditory sound “ba-ba”. Next, the individual in the video mouthed “ga-ga” with the auditory sound of “ba-ba”, which produced the fused perception of “da-da” or “tha-tha” (McGurk & MacDonald, 1976).

More recently, studies have investigated the neural underpinnings of the McGurk effect by presenting participants with matching or mismatching visual and auditory cues during fMRI scanning (Beauchamp, Nath, & Pasalar, 2010; Jones & Callan, 2003; Sekiyama, Kanno, Miura, & Sugita, 2003). The key area of activation was the superior temporal sulcus, which is known to be critically involved in AV MSI. Additional activity was found in a network of the visual and motor areas, including the extrastriate, premotor, and posterior parietal cortices, along with the inferior frontal gyrus (Beauchamp, Nath, & Pasalar, 2010; Jones & Callan, 2003). The McGurk and Ventriloquist effects are two widely-documented examples of the basic principle of the integration of sensory information from the visual and auditory modalities to create one, coherent, multisensory percept.

Temporal and Spatial Proximity

The spatial and temporal proximity of the stimuli presented are also instrumental determining whether two stimuli will be integrated (Stevenson & Wallace, 2013; Wallace et al., 2004). Spatial interactions in MSI are primarily grounded in the ability to integrate visual input with high fidelity auditory information as spatial tasks employ the visual modality by their nature (Rock & Victor, 1964; Welch & Warren, 1980). Overall, research has demonstrated that multisensory stimuli presented in close spatial proximity, or stimuli that fall within the same neuronal receptive fields, are more likely to be integrated than those that are presented distally

(Gepshtein, Burge, Ernst, & Banks, 2005). As previously stated, the Ventriloquist Effect is a widely known example of visual information influencing the perception of an auditory stimulus based on spatial placement. In this effect, the sound is perceived as originating from a different source due to the perception of a visual stimulus despite a spatial discrepancy in the sound source and visual stimulus (Howard & Templeton, 1966). Additionally, the Ventriloquist Effect has been documented as being automatic (Bertelson & Aschersleben, 1998; Vroomen, Bertelson, & de Gelder, 2001). Another study found that visuo-spatial information biased responses when participants had to localize auditory stimuli in a set of multisensory, audio-visual stimuli (Bertelson & Radeau, 1981).

While the importance of spatial proximity is demonstrated by the afore-reviewed literature, temporal synchrony is arguably recognized as the most important determinant of MSI. Studies have widely observed that sensory stimuli presented as linked in time, or being temporally synchronous, have a higher likelihood of being integrated (Stein & Meredith, 1993; Cecere, Gross, & Thut, 2016). Temporal synchrony may be a critical factor due to the properties of neurons that permit temporal summation, in which a high number of presynaptic action potentials induces summated or elevated postsynaptic potentials (Magee, 2000; Zwislocki, 2005; Zwislocki, 2005). Early studies presented auditory and visual stimuli in a sequential manner to understand the impact of temporal factors on crossmodal processing (Hirsh & Sherrick Jr., 1961; Sternberg & Knoll, 1973). One of these early studies made note of the auditory driving effect in which participants were presented with an auditory flutter and a visual flicker at certain rates (Shipley, 1964). If the flicker rate was initially presented as being slower than the flutter rate, the flicker rate was influenced by the flutter rate and was perceived as being faster (Gebhard & Mowbray, 1959). Furthermore, if both the rates start off in a synchronized manner, the flutter rate had to be adjusted by a large degree before the rates are perceived as being desynchronized (Shipley, 1964). More recently, Shams and colleagues (Shams, Kamitani, & Shimojo, 2000) developed the “double-flash” illusion to illustrate that auditory stimuli are able to bias visual perceptions. In this illusion one visual flash is perceived as two when simultaneously presented with two auditory clicks. These effects demonstrate that temporal synchrony is a powerful method to bind and alter the perception of multimodal stimuli.

Numerous studies have noted MSI enhancements by presenting AV stimuli in a temporally synchronized manner (Stein & Meredith, 1993; Stein, Wallace, Stanford, & Jiang, 2002; Zampini, Guest, Shore, & Spence, 2005). Temporal synchrony has been documented to be important for MSI in children and adults (Bahrick & Pickens, 1994; Lewkowicz, 1996; Morrongiello, Fenwick, & Nutley, 1998; Vroomen & Keetels, 2009). For example, three and six month old infants are able to discriminate between audiovisual information based on temporal synchrony (Bahrick, Flom, & Lickliter, 2002; Scheier, Lewkowicz, & Shimojo, 2003). Additionally, if presented in a temporally synchronized manner, newborns who are only a few hours old are able to learn AV associations, specifically between objects and linguistic stimuli (Morrongiello et al., 1998; Slater, Quinn, Brown, & Hayes, 1999). Spelke (1979) demonstrated that infants gazed at a puppet that was bouncing at the same speed as the sound presented for a longer time than a puppet that was bouncing asynchronously, or at a different speed. This evidence shows infants bind auditory and visual input based on temporal synchrony, further highlighting the importance of temporal factors in MSI from early on in development. While investigation regarding the neural bases of temporal effects on MSI is sparse, current evidence has localized these effects to the insula, superior temporal sulcus, inferior frontal gyrus, and the inferior parietal sulcus (Miller, 2005; Stevenson, VanDerKlok, Pisoni, & James, 2011).

Simple vs. Complex Stimuli

The ability to integrate information from multiple senses is also highly dependent on the properties of the stimulus presented. Simple stimuli, such as flashes and beeps, have formed the basis of many experimental studies and have been instrumental in revealing important findings about the organization of cortical neurons and the retina (Barutchu et al., 2010; Hartline, 1940; Hillock, Powers, & Wallace, 2011; Hubel & Wiesel, 1962). Simple visual stimuli include flashes, gratings, spots or bars of light, while simple auditory stimuli include beeps and tones, which can be easily parameterized and used for determining stimulus dependent responses. Due to the fact that these stimuli are generally produced electronically and represent a "purified version" of sight or sound, they are generally regarded to be artificial. Early work to understand the organization of visual and auditory cortices was conducted using such stimuli (Evans & Whitfield, 1964; Hartline, 1940; Hubel & Wiesel, 1962). The same simple stimuli were used even as investigations shifted from understanding unisensory to multisensory processes.

Moving beyond simple stimuli, more recent MSI research has turned to the use of more complex stimuli, which reflect natural sensory stimuli, to understand crossmodal mechanisms (Beauchamp, 2005; Erickson, Heeg, Rauschecker, & Turkeltaub, 2014; Felsen & Dan, 2005; Hocking & Price, 2008; James, VanDerKlok, Stevenson, & James, 2011; Stevenson & James, 2009; Stevenson & Wallace, 2013). Humans have used face-to-face conversations as the primary method of communication historically. Visual information can be extracted from speech by the movements of the mouth, eyes, and head. Therefore, normal speech lends itself to be a natural stimulus set to analyze audiovisual interactions. Research has found that individuals rely on visual properties of speech when the auditory properties are not salient, such as in a noisy room. Evidence in support of this notion notes that the visual percept of the face enhances the auditory discriminability of speech compared to the auditory stimuli alone (Sumby & Pollack, 1954).

It is important to note that the type of auditory stimulus presented impacts AV speech perception (Grant & Seitz, 1998). For example, studies of the McGurk effect present a range of auditory stimuli from nonsense syllables to full sentences (Grant & Seitz, 1998; Sams, Manninen, Surakka, Helin, & Kättö, 1998; Van Engen, Xie, & Chandrasekaran, 2017). Due to individual differences in the strategies employed by participants in such experiments, scores on measures in studies that use nonsense syllables do not correlate with scores on studies that use full sentences (Grant & Seitz, 1998). Even with the nuances in responses, speech stimuli provide a cleaner reflection of the complexities of AV stimuli compared to simple or artificial stimuli. While we have begun to understand the natural complexities in MSI by using complex speech stimuli, there remains a dearth of studies investigating these mechanisms using naturally-occurring, and semantically meaningful crossmodal correspondences.

Crossmodal Correspondences

Studies of MSI that used simple stimuli were instrumental in establishing the crossmodal correspondence between properties of vision and audition (Spence, 2011). Crossmodal correspondences are nonarbitrary associations between features present in different sensory modalities (Spence, 2011). These correspondences are useful for solving the crossmodal binding problem, for which it is important to know which of the many pieces of sensory information present in the environment should be bound together (Ernst, 2007). Studies have found that properties within the domains of audition and vison are mapped onto each other (Marks, 1987). For example, one study found that children link louder sounds with larger shapes (Smith & Sera, 1992). Pitch is associated with visual elevation such that higher pitches are matched with higher spaces (Walker et al., 2010). There has also been found that louder auditory stimuli are associated with brighter visual stimuli (Lewkowicz & Turkewitz, 1980; Marks, 1987; Maurer &

Mondloch, 2005; Maurer, Pathman, & Mondloch, 2006). Children and adults pair light grey color patches with louder sounds, while dark grey color patches are paired with quieter sounds (Bond & Stevens, 1969; Root & Ross, 1965). Additionally, researchers have shown a correspondence between hue and pitch in participants (Simpson, Quinn, & Ausubel, 1956). Simpson and colleagues found that the higher pitch that a tone is presented in, the more likely it is to be associated with yellow than with blue.

Impact of Multisensory Integration on Cognition

As detailed above, multisensory stimuli have an important influence on perception. However, MSI, and a variety of its factors impact not only perception, but also cognitive tasks involving memory and learning. Studies have also shown that MSI improves performance on a range of tasks, from basic perceptual tasks such as low-level target detection, to cognitive tasks which test memory and learning (Botta et al., 2011; Cichy & Teng, 2017; Deroy & Spence, 2013; Quak, London, & Talsma, 2015; Stein & Meredith, 1993). One such study investigated the effects of multisensory processing on perception using an audiovisual Stroop task (Stroop, 1935). Participants were presented with visual color words paired with auditory color words, non-color words, non-speech (music) stimuli, or silence. The results showed a visual Stroop interference when people were presented with spoken color-words, but not when presented with spoken non-color-words (Cowan & Barron, 1987). Furthermore, the interference was additive when participants were presented with spoken and written color words together compared to when they were only presented with a spoken or written word. While the Stroop task is a classic perceptual task, this study has implications for speech memory. It has been proposed that information from multiple sensory modalities is stored in a pre-speech buffer, from which the response must be selected (Cowan & Barron, 1987). Thus, it is more difficult to select the correct response if multiple color-words from the auditory and visual modalities are stored in the pre-speech buffer (Cowan & Barron, 1987; Salamé & Baddeley, 1982).

While the link between MSI and memory is much less studied than the link between MSI and perception (reviewed above), evidence suggests that MSI enhances memory in the same ways that it enhances perception. In an early demonstration of the involvement of multisensory processes in memory, Thompson and Paivio (1994) instructed participants to remember auditory, visual, and audiovisual stimuli to the best of their ability. Participants were able to recall the multisensory (AV) stimuli with higher fidelity than the unisensory stimuli. It has also been noted that this effect is due to the multisensory nature of the stimuli, as presenting the information twice in a visual-visual or audio-audio condition still yielded lower levels of accuracy (Goolkasian & Foos, 2005; Thompson & Paivio, 1994). A recent study revealed that visual objects presented with an irrelevant auditory sensory stimulus were recalled with greater accuracy than objects presented alone in an object discrimination memory task (Matusz, Wallace, & Murray, 2017).

Studies have also measured the multisensory enhancement in learning. In one study, participants were trained on a perceptual learning task using unimodal visual stimuli and multimodal audio-visual stimuli (Seitz, Kim, & Shams, 2006). Across all ten sessions, participants showed significantly enhanced learning when presented with the multisensory information. Similarly, another study found that voice recognition was improved when participants learned to associate auditory information with visual stimuli (e.g. faces and voices) as compared to when only the voice was presented (Von Kriegstein & Giraud, 2006). This study demonstrated that exposure to (via training) natural and redundant stimulus pairs can lead to forming multisensory associations. Overall, studies that show heightened learning in multisensory conditions have led researchers to conclude that conducting training studies using multisensory stimuli is more conducive to cognitive tasks (Shams & Seitz, 2008). Moving from previous research which has investigated the relationship of MSI, perception, and cognition, we now shift our focus to working memory, the memory store that I focused on in this thesis.

Working Memory

The short-term retention of information underlies the cognitive systems of short-term and working memory. Tests of short-term memory (STM) engage the cognitive faculties required to store information for a brief duration (seconds or minutes), and then utilize that information to guide subsequent behavior, by which time that information is no longer present in the external environment.

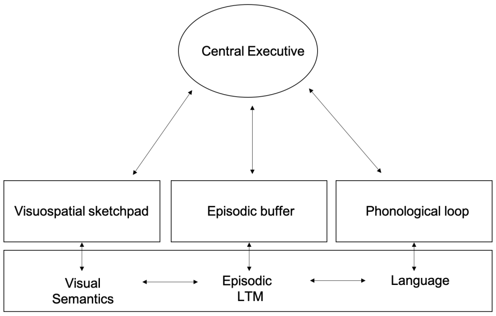


Figure 1.1. A model of working memory as proposed by Baddeley (2000).

While tests of working memory (WM) engage these same cognitive processes, they also require the manipulation or transformation of the stored information into a state most useful for a subsequent behavioral response. The concept of working memory was first coined by Baddeley and Hitch (1974). Even though the stored information appeared to be twice as large, they found that performance was equivalent when participants maintained visuospatial and phonological information in short-term memory, and when they only held information from one of these domains in short-term memory. This result led the researchers to conclude that there must be two separate memory buffers for each domain: a “visuospatial sketchpad” and a “phonological loop”. Additionally, they proposed that there existed a “central executive”, which would manage and gather information from these independent buffers. Baddeley & Hitch (1974, 2000) suggested a model of working memory (Figure 1.1) comprised of three principle information storage buffers: a visuospatial sketchpad, a phonological loop, and an episodic buffer. Additionally, the central executive would coordinate and recruit information simultaneously from these buffers to maintain cognitive control and manipulate information as demanded.

Based on the aforementioned information, the definition of working memory can be formulated as requiring the short-term storage of information and the ability of the central executive to manipulate this information to guide subsequent behavior. Tasks that require working memory often require us to coordinate and hold information for a short time in multiple memory buffers, and then manipulate the representation of that information, such as in tasks that require mental arithmetic, mental rotation, or mental reordering of stimuli to guide a response. While the details of the model, such as the exact number and content-specificity of the storage buffers, are still debated, the central proposition is a useful way to frame discussions of working memory. The concept that the short-term retention of information and the central executive form the faculties of working memories can seem simplistic, albeit functional. One drawback of this model is that it assumes that the sensory buffers are entirely independent and encapsulated. Instead, it is possible that information from various sensory modalities is represented in a connected, if not unified manner in working memory. The experiments conducted in this thesis use various principles of perceptual processing in MSI to test working memory. Therefore, if many of the same principles that impact perception also impact working memory, it would suggest that the sensory buffers are not entirely independent, but are governed by integrated sensory processing mechanisms. For a review of current working memory models, including this one, extensions, and alternatives see Postle (2015) and Postle and Pasternak (2010).

Capacity, Objects, and Features

To more completely understand the mechanism of working memory, it is important to understand its storage capacity, which objects and features are maintained in this memory store, and had how they are bound together. Unlike other memory stores (i.e. long-term memory), studies have shown that working memory has a limited capacity to maintain information across a period of time (Miller, 1956). Adults have generally been found to have a storage capacity of about four items in working memory (Cowan, 2001, 2010). The reason behind this limit remains unknown; however, researchers have proposed that this limit can be viewed as a strength or a weakness (for a review see Cowan, 2005). Simulation studies have suggested a strength of the capacity limit by showing that searching through groups and learning new information (e.g. grammar) in increments of 3.5 items on average is the most advantageous, (Cowan, 2010). However, a weakness of this limit comes from a biologically-based theory which states that having a limited capacity may be too expensive for processing large pieces of information neurally, therefore causing memory errors (Cowan, 2010).

Experiments that use paradigms to test working memory present the participant with a stimulus set to be studied. After a given delay period, typically on the order of seconds or minutes, participants are either instructed to freely recall the previously presented stimuli, or are presented with a new stimulus set and are instructed to make a match/non-match judgment compared to the previous stimuli. Luck and Vogel (1997) investigated the number of objects that can be held in visual working memory and manipulated the features of those objects using a change detection paradigm, in which participants had to indicate whether there had been a change in the two arrays of stimuli presented. In this study, participants could hold approximately four items in working memory. Furthermore, performance on tasks of working memory appears to be mediated by the complexity and number of items presented (Wilken &

Ma, 2004) such that when presented with multiple simple objects that exceed their capacity for storage, humans only store high-resolution representations of a few objects and do not retain information about the others (Woodman, Vogel, & Luck, 2012; Zhang & Luck, 2008).

Studies have tested the capacity limit of working memory by varying the memory load presented, the number of objects presented in the stimulus array, or changing the features of the objects. One feature that studies have manipulated is the spatial location at which the objects are presented. These studies of visual working memory have found that while the spatial location does not automatically bind to the object's features, these features do automatically bind to the object's spatial location (Jiang, Olson, & Chun, 2000; Olson & Marshuetz, 2005). However, other studies have shown that verbal and spatial information are remembered in a bidirectionally integrated fashion (Bao, Li, & Zhang, 2007; Campo et al., 2008, 2010; Meier, Nair, Meyerand, Birn, & Prabhakaran, 2014; Prabhakaran, Narayanan, Zhao, & Gabriel, 2000).

While object features such as spatial location seem to differentially affect visual and verbal working memory, research has widely shown that working memory capacity is estimated to be four items regardless of the features presented, specifically in the visual domain (Luck & Vogel, 1997, 2013; Zhang & Luck, 2008). Studies have also found that this limit is smaller than four items when presented with non-verbal auditory stimuli (e.g. tones) (Lehnert & Zimmer, 2006; Li, Cowan, & Saults, 2012). However, recent evidence calls into question the presence of a discrete fixed item-capacity limit. Schneegans & Bays (2016) argue instead that behavior on visual working memory tasks is better explained by a model where the putative limited storage resource(s) becomes progressively more thinly distributed among items as mnemonic load increases, but that all items are nonetheless being represented (albeit weakly).

Theories

Multisensory Integration & Memory: Dual Coding Theory

One possible explanation for the memory advantage seen in multisensory processing relies on the dual coding model of memory (Paivio, 1969, 2008). This model proposes that having both a visual and verbal code leads to enhanced memory and learning. These two codes are created at the encoding stage and then can be called upon at the retrieval stage. Such an experience has been evoked in individuals by asking participants to associate mental images with meaningful words (Ishai & Sagi, 1997). Paivio (1969) also found that participants had higher recall for concrete words which evoked an image compared to abstract words that did not.

Although the dual coding theory was first developed to explain the behavioral effects and interactions of verbal and non-verbal information, it has been adapted to understand the mechanisms of audiovisual, multisensory processing. It has been postulated that the sensory codes for two pieces of information are stored independently, and do not fuse into a single multi-modal representation. For this reason, based on task demands, these sensory codes can be recalled separately, although they can interact to aid behavioral performance (Quak et al., 2015). In this account, the dual coding mechanism produces enhanced memory recall for multisensory stimuli due to the ability to access a greater amount of information about the stimulus via multiple code formats.

Working Memory: Central Storage vs. Sensory Modality-Specific Storage

There remains an open question in the working memory literature: Is auditory and visual information stored in separate or a combined system in working memory? While some studies have supported the notion of distinct storage systems for sensory information from separate modalities (Baddeley, 1986), others have supported the notion of a central storage system (Cowan, 2005). In particular, studies have used dual-task paradigms to further understand the mechanistic underpinnings of the storage of visual and auditory information in the working memory domain (Fougnie & Marois, 2006; Saults & Cowan, 2007). In these paradigms, two tasks use two different sensory modalities (e.g. auditory and visual), rather than using tasks that tap into only one sensory modality (e.g. visual). These studies have found that task costs are significantly lower if the task-relevant information originates from a different sensory modality in the second task, compared with if it originates from the same sensory modality (Baddeley & Hitch, 1974; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Fougnie & Marois, 2006; Luck & Vogel, 1997b). This evidence has been used to support both the central and modality-specific storage systems of working memory by previous studies as described below.

Saults and Cowan (2007) tested if working memory has a central storage unit in a series of five experiments in which participants had to remember only visual or auditory information (unimodal condition) or both visual and auditory information (bimodal condition). They hypothesized that if a central storage applied to the processing of information in one sensory modality, it should also apply to bimodal or multisensory information. Additionally, if working memory is governed by a modality-specific store, then capacity for multisensory stimuli should be much larger than the four-item threshold, as would be expected for unisensory stimuli. To test this, Saults and Cowan (2007) asked participants to remember unimodal (visual or auditory) or bimodal (visual and auditory) information in a dual task and found dual task tradeoffs for only visual working memory capacity. They concluded that their findings provide support for a central “limited-capacity WM storage that cuts across modalities and codes” (Saults & Cowan, 2007). While evidence from previous research supports the central storage as well as modality-specific storage accounts of memory, fundamental questions about the storage of sensory information in working memory remain open. For this reason, additional tests of multisensory working memory are required to further understand the mechanisms at play.

The Neural Bases of Multisensory Integration and Working Memory

Multisensory Integration

Much of the neurophysiological evidence of MSI originated from studies of the superior colliculus (Wallace & Stein, 1996; Wallace & Stein, 1997). The superior colliculus is a structure in the mammalian mid-brain which has multiple layers (King, 2004). The superficial layers of the superior colliculus get input directly from the retina, while deeper layers process information from various sensory modalities. This structure is involved in behaviors of orientation by shifting gaze and coordination of movement to orient to the salient stimulus (Krauzlis, Lovejoy, & Zénon, 2013; Sparks & Jay, 1986). From the superior colliculus, the information is relayed to the spinal cord and brainstem, along with other sensory and motor regions to initiate the motor actions (e.g. eye movements) (Grantyn & Grantyn, 1982; May, 2005). The superior colliculus has been primarily investigated with animal studies, particularly in cats (for a review see Stein, Jiang, & Stanford, 2004).

Work in the adult cat has shown that sensory information from various sensory modalities, namely auditory and visual, converge onto multisensory neurons located in the superior colliculus (Beauchamp, Lee, Argall, & Martin, 2004; Stein & Wallace, 1996; Wallace & Stein, 1997). Multisensory neurons are defined as neurons which respond to input from multiple sensory modalities, or whose response is altered by input from a secondary sensory modality (Stein & Meredith, 1993). These neurons synchronize multi-modal information to relay a unified sensory message, rather than a series of summated messages. Research has been conducted with young kittens to understand the impact that these neurons and various stimuli have on the development of MSI (Wallace & Stein, 1997). Some of this research has supported the early integration perspective which proposed that information that is redundantly presented across multiple (two or more) modalities is instrumental in capturing attention in young infants (Bahrick & Lickliter, 2000, 2003; Bahrick, Lickliter, Castellanos, & Vaillant-Molina, 2010). Other research has supported the late integration perspective, which proposed that MSI abilities are not present at birth and develop through cognitive and neural reorganization which are driven by life experiences by showing that multisensory neurons are not found till late in postnatal development (Gori, Sandini, & Burr, 2012; Robinson & Sloutsky, 2010; Wallace & Stein, 1997)

Moving beyond single neuron recordings in the superior colliculus, single unit physiology studies in non-human primates have noted activation for MSI processes in the inferior parietal sulcus and superior parietal lobule (Cohen, Cohen, & Gifford, 2004; Molholm, 2006). While homologous neural regions have not been fully established between humans and other animals, multisensory responses have been found in these regions in humans as well (Bremmer et al., 2001; Sereno & Tootell, 2005). More recently, research has also considered the involvement of the primary visual and auditory cortices in MSI (Ghazanfar & Schroeder, 2006; Henschke et al., 2017). Traditionally, the primary sensory cortices have been thought to be unimodal, processing input only from the specified modality (e.g. visual cortex processes visual input). However, studies have found direct anatomical and functional connections between the primary visual and auditory cortices, which demonstrated that AV crossmodal interactions are present from early states of sensory processing (Calvert, Spence, & Stein, 2004; Cappe & Barone, 2005; Christoph Kayser & Logothetis, 2007; Kumar et al., 2017). To gain higher acuity in localizing cross modal interactions, researchers functionally parcellated the auditory cortex of the macaque monkey into functional fields and found that the visual stimuli enhanced activation specifically in the caudal parabelt and medial belt of the auditory cortex (Kayser, Petkov, Augath, & Logothetis, 2007).

Other studies have demonstrated that sensory input from one modality triggers activation in cortical areas that correspond to another sensory modality. In one of the first demonstrations of this effect, Calvert and colleagues showed activation in the primarily auditory cortex while silently lip reading in individuals with normal vision and hearing (Calvert, 1997). Other studies have found activation in the visual cortex of blind individuals when presented with auditory stimuli (Kujala et al., 1995; Rothen, Bartl, Franklin, & Ward, 2017), and in the auditory cortex of deaf individuals when presented with visual stimuli (Finney, Fine, & Dobkins, 2001). Tying together these themes, previous research has hypothesized that this cross-sensory reorganization is made possible by audio-visual connectivity in the primary sensory cortices (Ghazanfar & Schroeder, 2006).

The neural representation of MSI with speech stimuli has been broadly recognized to be processed the superior temporal cortex (Stevenson et al., 2011). A range of studies have localized the neural correlates of multisensory speech to various brain regions such as the dorsal and ventral streams (Hickok & Poeppel, 2007; Rauschecker, 2012), the superior temporal sulcus (Beauchamp et al., 2010; Hein & Knight, 2008; Hocking & Price, 2008; Raij, Uutela, & Hari, 2000), and the inferior frontal gyrus (Ojanen et al., 2005; Sekiyama et al., 2003). A meta-analysis of conflicting (different signals) versus validating (complementary signals) AV speech signals identified a network in the dorsal stream regions for conflicting stimuli and in the ventral stream visual regions of the occipital and temporal lobes for validating stimuli (Erickson et al., 2014). In summary, studies that have investigated the neural underpinnings of MSI have found evidence for this process in the superior colliculus, inferior parietal sulcus, superior parietal lobule, and sensory cortices across studies conducted with cats, non-human primates, and humans using a variety of neuroimaging tools.

Working memory

Research centered around understanding the neural mechanisms of working memory has aimed to answer a variety of questions such as, "Where is information in working memory stored?", "Is there one area of the brain that holds this information, or is it distributed?", and "How can neural mechanisms account for the ability to manipulate information in working memory?". In the first of these investigations, Lorente de Nó proposed the persistent neural activity model (de Nó, 1933). De Nó observed that information was stored in neural activity after the presented stimulus had ended. In 1964, Pribram and colleagues found that the prefrontal cortex (PFC) was critical for working memory tasks through lesion studies (Pribram, Ahumada, Hartog, & Roos, 1964). More specifically, it was found that lesions in the PFC impaired working memory performance through proactive interference, in which information from previous trials interferes with the to-be-remembered information in the current trial (Milner, 1964). It is important to note that these lesion studies do not necessarily suggest that information from working memory was stored in the PFC. These early studies concluded the contents of WM were stored in PFC, but this conclusion has more recently been called into question.

Since these early studies, various forms of neuroimaging such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) have led to a much deeper understanding of the neural underpinnings of working memory. Studies using these techniques have noted the presence of persistent neural activity in the PFC during tasks of working memory (Braver et al., 1997; Fuster & Alexander, 1971; Goldman-Rakic, 1987; Kane & Engle, 2002; Kubota & Niki, 2017; Kubota, 2003). Researchers have used delayed recognition paradigms and match-to-sample paradigms to locate neural activity in the PFC, and have also found evidence of activity in other neural areas such as the posterior parietal cortex and inferior temporal cortex during working memory tasks (Gnadt & Andersen, 1988; Miller, Li, & Desimone, 2018). More specifically, studies which find persistent neural activity in the PFC have found that the identity of the stimulus was encoded during the delay period, and not when the stimulus was presented (Funahashi, Inoue, & Kubota, 1993; Goldman-Rakic, 1987). While the view that the PFC is the site of storage of information during working memory processes remains prevalent in the cognitive neuroscience field, more recent work has challenged this view by proposing the involvement of sensory regions using sophisticated analytical methods.

Many of the studies conducted using fMRI techniques to investigate neural mechanisms of working memory are limited because they use univariate analysis methods, which smooth over voxels and treat activity from regions as a single metric of "activation" (Lewis-Peacock & Norman, 2014). Recently, many studies have used multivariate pattern analysis, which extracts multidimensional neural activity patterns across voxels that reveal distributed (rather than localized) patterns of processing (Haxby, 2012; Kriegeskorte et al., 2008; Norman, Polyn, Detre, & Haxby, 2006; Pereira, Mitchell, & Botvinick, 2009). This analytical method provides higher sensitivity by extracting activity from multiple neural networks, and also provides greater specificity by overcoming several assumptions made by univariate methods about multiple comparisons (e.g., neurons in close proximity spatially have similar functionality and represent information in similar ways) (Leibenluft, Saad, Cox, Adleman, & Chen, 2014).

Studies investigating working memory using multivariate pattern analysis have been able to decode information from area V1, primary visual cortex, during the delay period, even if these regions do not show elevated delay period activity using a variety of stimulus types (e.g. objects, colors, visuospatial patters, faces, houses, scenes) (Christophel, Hebart, & Haynes, 2012; Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009). Moreover, these studies have not been able to decode information from the PFC, despite finding elevated activity during the delay period (Riggall & Postle, 2012). Together, these findings that sensory regions are recruited during working memory tasks have formulated the sensory recruitment hypothesis (Ester, Anderson, Serences, & Awh, 2013). Based on this hypothesis, the stimulus information stored in the sensory regions during working memory tasks would directly correlate with behavioral performance on these tasks, whereas PFC activity may act as a "central executive", such as that from Baddeley and Hitch's model of working memory (Emrich, Riggall, LaRocque, & Postle, 2013; Sprague, Ester, & Serences, 2014).

Multisensory Integration & Working Memory: Shared Behavioral and Neural Correlates

As noted earlier, behavioral studies have suggested that working memory might have a multisensory or multimodal component (Botta et al., 2011; Quak et al., 2015; Shams & Seitz, 2008). Although this link has been established, there are surprisingly few studies explicitly investigating the impact of multisensory integration on working memory processes. Botta et al. (2011) demonstrated that multimodal (audiovisual) cues allow for more information to be transferred to visuospatial working memory than unimodal (visual) cues, thus having a larger effect on working memory capacity. Importantly, research has established that these enhancements in working memory are due to the multisensory nature of the stimuli and not simply due to the redundancy of information that is presented (e.g. double visual presentations) (Goolkasian & Foos, 2005). For example, presenting participants with an image of a cat twice would not produce an enhancement, but presenting an image of a cat with a cat sound would provide an enhancement due to the information arising from two separate (i.e. visual and auditory) sensory modalities.

Additionally, the encoding, processing, and retrieval of multisensory information is easier, making this material beneficial for storage and learning (Newell & Mitchell, 2016; Quak et al., 2015). A recent study tested the effect of crossmodal correspondences on MSI and working memory in adults (Brunetti, Indraccolo, Mastroberardino, Spence, & Santangelo, 2017). In a series of two experiments this group investigated numerosity, pitch/elevation, and pitch/shape correspondences using an n-back paradigm. Overall, results revealed effects of congruency, such that working memory performance is enhanced when auditory and visual stimuli are presented in a corresponding manner.

Another study investigated how auditory, visual, and audiovisual information is encoded in working memory when presented in verbal and nonverbal formats (Delogu & Raffone, 2009). This study found enhancements in WM when participants were presented with verbal audiovisual material compared to unimodal visual and auditory information. However, they only found enhancements for non-verbal audiovisual material in comparison to unimodal visual, but not auditory information. Critically, performance in both verbal and non-verbal conditions was reduced when participants were instructed not to articulate the stimuli presented. This led the researchers to conclude that stimuli were recoded if necessary, and stored in a verbal format regardless of the format they were presented in. These studies suggest that using multisensory stimuli calls upon mechanisms in which two-codes link the presented sensory modalities together, and hence enhance working memory recall. In summary, studies to date have found that memory recall is enhanced for multisensory compared to unisensory stimuli (Delogu & Raffone, 2009; Quak et al., 2015; Thompson & Paivio, 1994), and working memory capacity is greater for multisensory objects, but only under certain conditions (Fougnie & Marois, 2011; Saults & Cowan, 2007).

Previous research has suggested that MSI and working memory may share specific neural correlates (D'Esposito & Postle, 2015; Wheeler et al., 2000) (See Figure 1.3). The longstanding sensory reactivation hypothesis proposes that sensory areas, which are active during the encoding of information, are reactivated during the retrieval of that same information (Wheeler et al., 2000). More specifically, research has considered the involvement of the visual and auditory cortices in MSI and in working memory (Emrich et al., 2013; Ghazanfar & Schroeder, 2006; Henschke et al., 2017). Overall, these studies suggest that brain regions involved in low-level perceptual and sensory processing are the same as those involved in the memorization of sensory information (D'Esposito & Postle, 2015; Postle, 2006). This evidence shows that working memory and MSI processes both activate early sensory cortices in a top-down manner, and thus the same pathways that mediate memory reactivation may also mediate MSI.

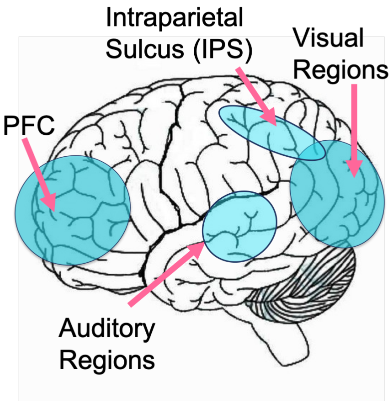


Figure 1.2. A schematic of the neural regions involved in multisensory integration and working memory processes.

Based on these studies, one perspective suggests that multisensory representations should be stored in working memory, along with unisensory representations. Indeed, a study found effects of auditory and tactile integration in the auditory cortex, which had previously been found to process unimodal auditory stimuli (Foxe et al., 2002). Another study found increased activity in the primary somatosensory cortex for bimodal (visual and somatosensory) stimuli compared to unimodal stimuli in a delayed sensory-to-motor task (Dionne, Meehan, Legon, & Staines, 2010).

Conversely, a different perspective suggests that unisensory information is processed in sensory cortices, and higher order brain regions are involved in combining this information to form multisensory neural representations. Support for this view comes from studies that found evidence of multisensory processing in regions of the occipital-temporal cortex, such as the superior temporal sulcus (Beauchamp, 2005). While evidence exists to support both perspectives, additional studies explicitly investigating multisensory working memory processes in the audiovisual domain are required to further understand the overlapping mechanisms of MSI and working memory.

Open Questions & Dissertation Overview

The evidence, studies, and theories reviewed above raise a myriad of unanswered questions regarding the interaction of multisensory integration and working memory. First, do multisensory stimuli provide a benefit for working memory processing? Second, are the auditory and visual modalities linked differently based on stimulus properties? Third, does an investigation in the same group of participants reveal that simple and complex stimuli affect working memory differently? Lastly, what are the effects of crossmodal congruency on MSI and working memory?

This thesis aims to elucidate the importance of multisensory integration by discussing key components that determine its success, and the impact it has on memory mechanisms. To begin answering the aforementioned questions, in a series of four experiments in Chapter 2, I investigated the behavioral basis of multisensory integration as it impacts working memory using simple stimuli. I extended this experimental design in a series of four experiments in Chapter 3 by using a different stimulus set that consists of complex, more semantically meaningful, and naturalistic stimuli. Chapter 4 demonstrates the impact of stimulus complexity on working memory in the same participants. In the concluding chapter (Chapter 5), I discuss the key findings as situated within the theoretical framework of cognitive neuroscience, highlight educational implications, and conclude by identifying common themes and future directions for the field of mind, brain, and education.

Chapter 2: The Impact of Simple Stimuli on Multisensory Integration and Working Memory

Introduction

Much of the information that we gather from everyday situations, such as meeting a new person, or taking a trip to the grocery store is derived from multiple senses. Previous studies have investigated the behavioral, neural and perceptual underpinnings of multisensory integration (Lewkowicz & Turkewitz, 1980; Stein & Meredith, 1993). More recently, research has suggested that working memory and multisensory processes may be intertwined (Botta et al., 2011; Quak et al., 2015; Shams & Seitz, 2008) as recall is enhanced for objects with crossmodal rather than unimodal components (Delogu & Raffone, 2009; Thompson & Paivio, 1994). In an early investigation in the memory domain, Thompson and Paivio (1994) presented undergraduate students with pictures, sounds, or picture-sound pairs and found that students were best able to recall in the third condition, particularly when also presented with a distractor task.

Botta and colleagues (2011) also demonstrated that multimodal (audiovisual) cues allow for more information to be transferred to visuospatial working memory than unimodal (visual) cues, thus having a larger effect on working memory performance. Brunetti and colleagues (2017) found that working memory performance is enhanced when auditory and visual stimuli are presented in a corresponding manner across the sensory modalities. Together the aforementioned investigations suggest that using multisensory stimuli calls upon dual coding mechanisms that link the presented sensory modalities together, and thus enhance working memory recall.

Many of the studies conducted to understand the nature of MSI have used simple stimuli (Barutchu et al., 2010; Hillock et al., 2011). Simple visual stimuli include flashes, gratings, spots or bars of light, while simple auditory stimuli include beeps and tones, which can be easily parameterized and used for determining stimulus dependent responses. Due to the fact that these stimuli are often produced electronically and represent a "purified version" of sight or sound, they are generally regarded to be artificial. Indeed, early work to understand the organization of visual and auditory cortices was conducted using such stimuli (Evans & Whitfield, 1964; Hubel & Wiesel, 1962). Simple stimuli elicit well-understood patterns of neural activity, are advantageous for isolating response properties, and are considered to be well-suited for experiments due to their negligible semantic content. The same simple stimuli have continued to be used even as investigations have shifted from understanding unisensory to multisensory processes. These early studies of MSI were instrumental in establishing the correspondence and mapping between stimulus properties within the domains of audition and vison (Marks, 1987; Spence, 2011). Crossmodal correspondences are nonarbitrary associations between features present in different sensory modalities (Spence, 2011). These correspondences are useful for solving the crossmodal binding problem, for which it is important to know which of the many pieces of sensory information present in the environment should be bound together (Ernst, 2007).

Auditory pitch has been found to correspond to a number of visual factors such as shape angularity, weight, spatial location, color brightness, and speed (Boltz, 2011; Collier & Hubbard, 2001; Eitan & Timmers, 2010; Marks, 1987). Specifically, studies have found that lighter colors correspond to higher pitches while darker colors correspond to lower pitches (Martino & Marks, 1999; Melara, 1989; Walker, Scallon, & Francis, 2017). Pitch is also associated with visual elevation such that higher pitches are matched with higher spaces (Walker et al., 2010). Additionally, researchers have shown a correspondence between hue and pitch in participants (Simpson et al., 1956). Simpson and colleagues (1956) found that the higher pitch that a tone is presented in, the more likely it is to be associated with yellow than with blue. In addition to correspondences between pitch and visual factors, it has also been found that louder auditory stimuli are associated with brighter visual stimuli (Lewkowicz & Turkewitz, 1980; Marks, 1987; Maurer & Mondloch, 2005; Maurer, Pathman, & Mondloch, 2006). Children and adults pair light grey color patches with louder sounds, while dark grey color patches are paired with quieter sounds (Bond & Stevens, 1969; Root & Ross, 1965). In this study, we focus on correspondences between color and pitch, which have previously been identified in adults and children (Mondloch & Maurer, 2004; Simpson et al., 1956).

One way in which studies have investigated the effects of crossmodal correspondences is by presenting them in a congruent and incongruent manner. Crossmodal correspondences are maintained in a congruent presentation, whereas they are not maintained in an incongruent presentation. Overall, studies investigating effects of congruency in multisensory memory with simple stimuli have noted that participants have higher memory performance when presented with congruent compared to incongruent multisensory stimuli (Calvert et al., 2004; Lacey, Martinez, McCormick, & Sathian, 2016; Spence, 2007). A recent study tested the effect of crossmodal correspondences on MSI and working memory in adults (Brunetti et al., 2017). In a series of two experiments this group investigated numerosity, pitch/elevation, and pitch/shape correspondences using a n-back paradigm. Overall, the results of this study reveal effects of congruency, such that working memory performance is enhanced when auditory and visual stimuli are presented in a corresponding manner. Furthermore, a study showed that congruent stimuli (i.e. visual shape/sound) enhanced working memory, but only when the duration between the stimulus array and memory probe was short (150 ms), and diminished at long intervals (1150 ms) (Makovac, Kwok, & Gerbino, 2014).

In many of the studies to date, multisensory information has been the central focus of the task and participants have been instructed to use the multisensory information to complete the task at hand (Koelewijn, Bronkhorst, & Theeuwes, 2010; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). One study found that multisensory distractors have significant effects on attentional mechanisms in adults and children (Matusz et al., 2015). Participants were presented with audiovisual distractors when instructed to complete a visual search task with color-defined objects or graphemes. The distractors and target objects always shared features, leading the researchers to conclude that the participants' goals in completing the task may affect the allocation of attention. Thus, while a few studies have investigated this topic, whether multisensory information being central or peripheral to the focus of attention has an effect on memory performance has largely been an open question in the literature.

Furthermore, researchers have suggested that the availability of attentional resources may impact multisensory integration. At a neural level, this idea is supported by evidence that semantic memory-dependent multimodal processes engage the brain at a relatively late latency that is over 100 ms after the stimulus onset (Alsius, Möttönen, Sams, Soto-Faraco, & Tiippana, 2014; Sophie Molholm, Ritter, Javitt, & Foxe, 2004; Raij et al., 2000; Yuval-Greenberg & Deouell, 2007), and are dependent on the task (Stevenson & Wallace, 2013; van Atteveldt, Formisano, Goebel, & Blomert, 2007). Alsius and colleagues (2014) showed that the available attentional resources influence multisensory processing using complex and naturalistic stimuli. They found that the ability to detect audiovisual congruence between voices and faces was dependent on the number of faces available (Alsius et al., 2014). Additionally, the McGurk effect was significantly reduced when the participant's attention was diverted to a concurrent task (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Alsius, Navarra, & Soto-Faraco, 2007), and typical neural signatures of the McGurk effect were also absent in this context (Alsius et al., 2014).

The Current Study

In a series of four studies, we aimed to answer the following questions: How does audio-visual multisensory integration affect working memory performance when presented with simple stimuli? How do color-pitch crossmodal correspondences influence MSI and working memory? How does attentional cueing affect multisensory processing in the working memory store? Based on principles of crossmodal correspondences, we used congruent and incongruent presentations of instrumental tones, pitch, and color patches answer these questions. We used a standard change detection paradigm with memory loads of two, four, six, or eight items to test working memory in unisensory (auditory alone and visual alone) as well as multisensory (audiovisual) conditions.

Using a change detection paradigm is beneficial for a few key reasons. First, a change detection task is a classic paradigm used to study working memory. Using this type of task allowed us to test if the memory system is flexible enough to detect changes in the identity of the object once the presentation of the object has ended. Change detection paradigms were first conducted using only one stimulus and participants had to detect a change in the given stimulus. Thereafter, researchers began using multiple display stimuli with various changing properties to test the capacity and flexibility of the working memory system using change detection tasks. Adopting this type of paradigm will allow us to test the capacity and flexibility of multisensory working memory in typically functioning adults. Additionally, this paradigm will shed light on the ability to hold multiple objects in memory and how effectively individuals are able to compare incoming sensory information with pre-existing representations in their working memory.

We expected that participants would perform with lower accuracy and would have longer reaction times on trials with a higher compared to a lower memory load. Conversely, based on previous research and analyses methods, we hypothesized the capacity estimates would increase as the load increased (Rouder, Morey, Morey, & Cowan, 2011). Furthermore, we expected that guiding attention towards a to-be-remembered sensory modality would have an influence on working memory performance. More specifically, we predicted that we would find a deficit in working memory performance, and multisensory integration as attentional resources were saturated (Alsius et al., 2014). We expected to find differences at the modality-specific level. In particular, we expected participant to have an advantage in working memory when presented with multisensory compared to unisensory stimuli in the auditory (AV-Test A compared to A) and visual (AV-Test V compared to V) domains. Lastly, in line with findings from previous research and the dual-coding theory (Brunetti et al., 2017; Paivio, 1969), we predicted that we would find effects of congruency, such that working memory performance would be enhanced when participants were presented with congruent rather than incongruent audiovisual stimuli.

General Methods

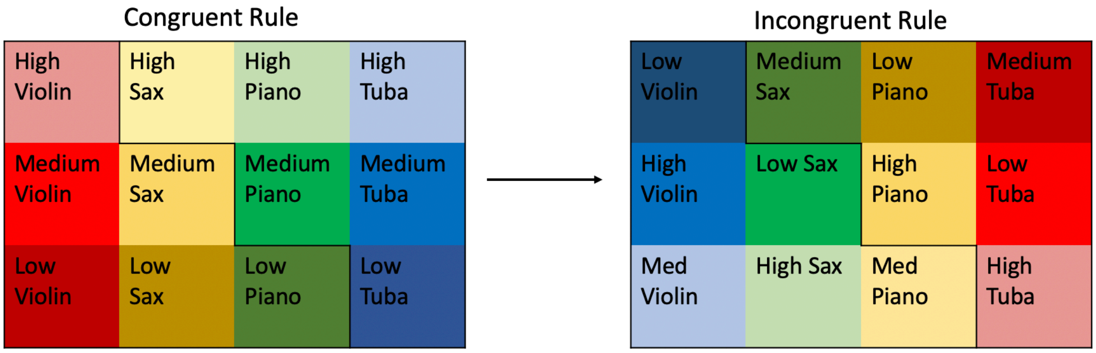
Participants & Procedure

We recruited undergraduate students for the experiments outlined in this chapter. Participants were compensated with partial course credit. All components of the study were approved by the Institutional Review Board (IRB #2018-0395). The procedure was administered in a quiet room in the Educational Neuroscience Lab space. Experiments were programmed with E-prime 2.0.8.90a (Psychology Software Tools, Sharpsburg, PA) on a Dell Optiplex 390

Desktop PC (3.1 GHz, 4 GB RAM) running Windows 7.0 64-bit operating system. Visual stimuli were presented on a Dell UltraSharp U2212H 21.5” flat-screen monitor at a resolution of 1024 x 768 and a refresh rate of 60 Hz. Auditory stimuli were presented using headphones (Sony MDR-ZX110 Black).

Stimuli

To understand the influence of simple auditory-visual correspondence on MSI of WM, we presented participants with one of four different instrumental timbres: violin, piano, tuba, and saxophone, and four colors corresponding to these sounds: red, green, blue, yellow, respectively. We have found evidence for these timbre-color associations a study of synesthetes and non-synesthetes previously conducted in the Educational Neuroscience Lab, and similar associations have been also documented in previous research by other groups (Gosavi, Bade, & Hubbard, in prep; Ward, Huckstep, & Tsakanikos, 2006). Following principles of color-pitch correspondences (Martino & Marks, 1999; Melara, 1989), each of these sounds was presented in either a high, medium, or low pitch, and each color was presented in a light, medium or dark value (shade), resulting in a total of 12 stimuli (Figure 2.1). The stimuli were presented



Congruent Rule

High Violin	High Sax	High Piano	High Tuba
Medium Violin	Medium Sax	Medium Piano	Medium Tuba
Low Violin	Low Sax	Low Piano	Low Tuba

Incongruent Rule

Low Violin	Medium Sax	Low Piano	Medium Tuba
High Violin	Low Sax	High Piano	Low Tuba
Med Violin	High Sax	Med Piano	High Tuba

Figure 2.1. Stimulus set used to test the impact of simple stimuli on working memory. The stimuli were presented as per the cross-modal correspondences in the “congruent” rule. The colors were flipped across the diagonal line and sounds were rotated within the columns to create the schematic that formed the “incongruent” rule.

consistently with the crossmodal correspondences in congruent trials (e.g. high-pitched violin tone paired with the light shade of red). Crossmodal correspondence were systematically disrupted in incongruent trials. To form the incongruent grid, we systematically varied the distance of each stimulus from its position in the congruent grid by exchanging the positions of the colors across the diagonal line and rotating the pitch of the tones for a given instrument.

Experimental Design

On each trial, participants were presented with a different number of to-be-remembered items (i.e. memory load, two, four, six, or eight items). In a given trial, each stimulus was presented for 750 ms with a 250 ms interstimulus interval (ISI). After a delay period (4000 ms), participants were presented with either the same pattern or a different pattern of stimuli, and participants were asked to detect a change in the pattern presented in the response duration (2000

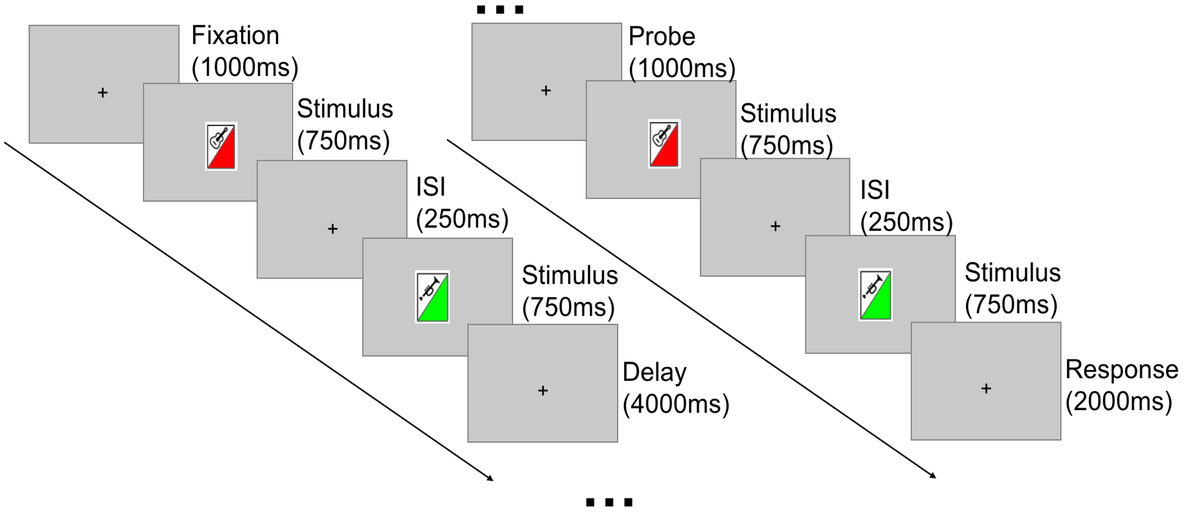


Figure 2.2. The multisensory working memory paradigm used in the experiments in this chapter with a set of simple stimuli. Participants were presented with a randomly selected array of two, four, or eight sounds or color patches. After a delay, they were presented with either the same pattern or a different pattern, and were asked to detect a change in the pattern presented during the response duration.

ms) (Figure 2.2). Only one item out of the string of stimuli changed in the change trials, and none of the stimuli changed in the no-change trials.

In the unisensory auditory (A) and visual (V) baseline conditions, participants were presented with and tested only on stimuli from the given modality. In the multisensory (audio-visual: AV) condition, participants were presented with auditory and visual stimuli in a temporally and spatially synchronized manner, but were only tested on stimuli from one modality (Conditions: AV-Test A or AV-Test V). Participants were cued to the condition (A, V, or AV) in a blocked fashion, which was counterbalanced across participants.

In the multisensory conditions, participants were presented with congruent trials in which cross-sensory correspondences were maintained on half the trials, and incongruent trials in which auditory and visual stimuli were presented as per the varied grid on half the trials (Figure 2.1). All the stimuli were presented with congruent pairings in congruent trials, and incongruent pairings in incongruent trial.

Analysis

Accuracy, capacity estimates, and reaction times were calculated, as together they provide a deeper understanding of the working memory mechanisms at play. While performance accuracy tends to decrease as the mnemonic load increases, individual differences in working memory capacity have also been noted (Linderholm, Cong, & Zhao, 2008; Luck & Vogel, 2013). Accuracy was calculated based on the proportion of correct responses on each trial. All participants with accuracies below 50% (chance performance) at lowest load in baseline A or V conditions were removed from the final analyzed data set. Reaction times were calculated as the time it took participants to respond (in milliseconds) from the offset of the last stimulus in the memory array.

Working memory capacity was calculated using Cowan's k (Cowan, 2001, 2010; Rouder et al., 2011). Under the assumption that participants have a working memory capacity that is fully employed on each trial, this formula to calculate working memory capacity (k) is detailed in the following manner: The probability that the participant makes a hit is $H= k/N + g((N-k)/N)$, where k = working memory capacity, N=total number of stimuli presented, and g is the probability of guessing correctly (based on the number of choices). That is, the probability of a hit is the sum of the probability that the participant detects the change (k/N) plus the probability that the participant guesses correctly, even when they do not detect the change ($g((N-k)/N)$). Based on this formula, the probability that the participant makes a correct rejection is $CR= k/N + (1-g)((N-k)/N)$. Combining these two formulas and solving for k, we can obtain the formula to calculate working memory capacity: $k=N(H+CR-1)$. Working memory capacity was calculated in the same manner here. Statistical results for all the pairwise comparisons of working memory performance (accuracy, reaction time, capacity) across conditions conducted for experiments in this chapter can be found in Appendix B.

Experiment 1

Methods

Thirty-five adult participants (mean=20.46 years, M=3, F=32) were recruited for this experiment. All recruited participants had normal or corrected-to-normal vision and hearing and were naïve to the purposes of the experiment. All participants were included in the final data set after the data cleaning analyses were conducted. Participants were presented with a randomly selected array of two, four, or eight sounds or color patches. The procedure for this study followed the experimental design as described above. In Experiment 1, participants were not cued to the modality that they were going to be tested on (AV-Test A or AV-Test V) in the multisensory condition. Participants were presented with a set of congruent stimuli on half the trials, and incongruent stimuli on the other half of the trials in a randomized manner.

Results

Effects of Modality

Accuracy

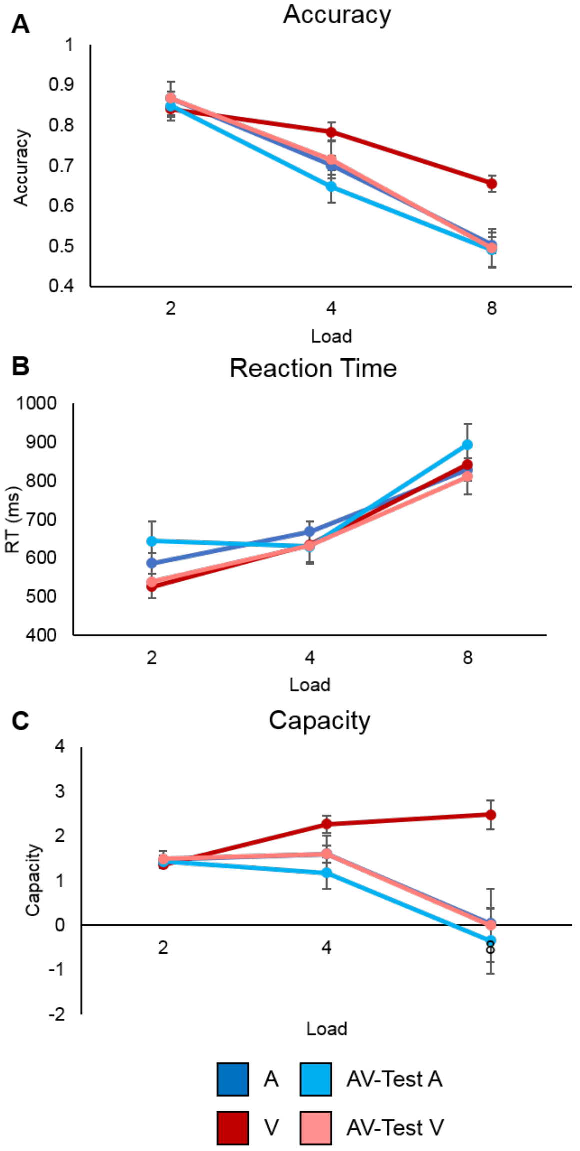


Figure 2.3. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants in the unisensory (V, A) or multisensory (AV-test A, AV-test V) conditions. Error bars represent the standard error of the mean.

As can be seen in Figure 2.3A, accuracy decreased as the memory load increased. The accuracy in the unisensory visual condition was significantly higher than the other four conditions. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,34)= 153.95, p<0.001, η_p^2=0.82), and condition (F(1,34)=8.07, p<0.001, η_p^2=0.192), and a significant load x condition interaction (F(1,34)=3.31, p<0.01, η_p^2=0.089). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. Overall, we found significant differences at the middle load of four items or highest load of eight items when accuracies were compared across the unisensory and multisensory conditions. When comparing accuracy for unisensory auditory and visual conditions, we found a significant main effect of condition (F(1,34)=18.32, p<0.001, η_p^2=0.35), load (F(1,38)=83.08, p<0.001, η_p^2=0.71), and a significant load x condition interaction (F(1,34)=14.81, p<0.001, η_p^2=0.30). Specifically, the accuracy for unisensory visual condition was significantly greater than unisensory auditory condition when at a load of 4 (t(34)=2.69, p=0.01), and 8 (t(34)=7.29, p<0.001) items. We did not find evidence of a visual multisensory enhancement. However, unisensory visual and AV-Test V were significantly different (condition F(1,34)=8.73, p<0.01, η_p^2=0.20, load F(1,34)=57.24, p<0.001, η_p^2=0.63, load x condition interaction F(1,34)=6.25, p<0.01, η_p^2=0.16). Visual alone (V) accuracy was significantly higher than AV- test V at a load of 8 (t(34)=3.78, p=0.001) items. Lastly, we did not find significant differences when comparing the unisensory auditory and AV-Test A conditions. A 2 (condition) x 3 (load) ANOVA revealed a significant main effect of load (F(1,34)=130.02, p<0.001, η_p^2=0.79), but not of condition (F(1,34)=1.78, p=0.19, η_p^2=0.05), or a load x condition interaction (F(1,34)=0.30, p=0.74, η_p^2<0.01).

Reaction Time

As shown in Figure 2.3B, reaction times increased at larger loads, but did not vary by stimulus condition. A 3 (load) x 4 (condition) ANOVA revealed a main effect of load (F(1,34)=95.93, p<0.001, η_p^2=0.74), but no main effect of condition (F(1,34)=2.41, p=0.71, η_p^2=0.07), nor a significant load x condition interaction (F(1,34)=1.91, p=0.08, η_p^2=0.05).

Capacity

As predicted, estimated capacity increased as the memory load increased for some conditions but not others (Figure 2.3C). The largest capacity estimate on a given trial is dictated by the number of stimuli presented on that trial. Thus, the possible capacity on a trial with a load of eight items is larger than on a trial with a load of two items. For this reason, the capacity estimate, *k*, is modulated by the multiplier, *N*, which accounts for the number of stimuli presented. Furthermore, visual inspection of Figure 2.3C revealed that the trend for capacity estimates was still increasing for the unisensory visual (V) condition, whereas it began to decline for the unisensory auditory (A), and multisensory (AV-Test A, AV-Test V) conditions at a load of four items. A 3 (load) x 4 (condition) ANOVA revealed significant main effects of load (F(1,34)=11.85, p<0.001, η_p^2=0.26), condition (F(1,34)=12.99, p<0.001, η_p^2=0.28), and a significant load x condition interaction (F(1,34)=6.29, p<0.001, η_p^2=0.16). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. When comparing accuracy for unisensory auditory and visual conditions, we found a significant main effect of condition (F(1,34)=46.30, p<0.001, η_p^2=0.58), load (F(1,34)=3.98, p=0.02, η_p^2=0.105), and a significant load x condition interaction (F(1,34)=29.20, p<0.001, η_p^2=0.46). This difference in capacity estimates was significant when compared at a load of 4 (t(34)=2.69, p=0.01), and 8 (t(34)=7.29, p<0.001) items. We did not find evidence of an visual multisensory enhancement. However, unisensory visual and AV-Test V were significantly different (condition F(1,34)=15.56, p<0.001, η_p^2=0.31, load F(1,34)=2.11, p=0.13, η_p^2=0.06, load x condition interaction F(1,34)=9.63, p<0.001, η_p^2=0.22). Visual alone (V) accuracy was significantly higher than AV test-V at loads of 4 (t(34)=2.67, p=0.01) and 8 (t(34)=3.69, p=0.001) items. Lastly, we did not find significant differences when comparing the unisensory auditory and AV-Test A. A 2 (condition) x 3 (load) ANOVA revealed a significant main effect of load (F(1,34)=21.72, p<0.001, η_p^2=0.39), but not of condition (F(1,34)=1.62, p=0.21, η_p^2=0.05), or a load x condition interaction (F(1,34)=0.27, p=0.76, η_p^2<0.01).

<u>Effects of Congruency</u>

Accuracy

Figure 2.4A shows accuracy when participants were presented with congruent versus incongruent stimuli in the multisensory condition. This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a main effect of load

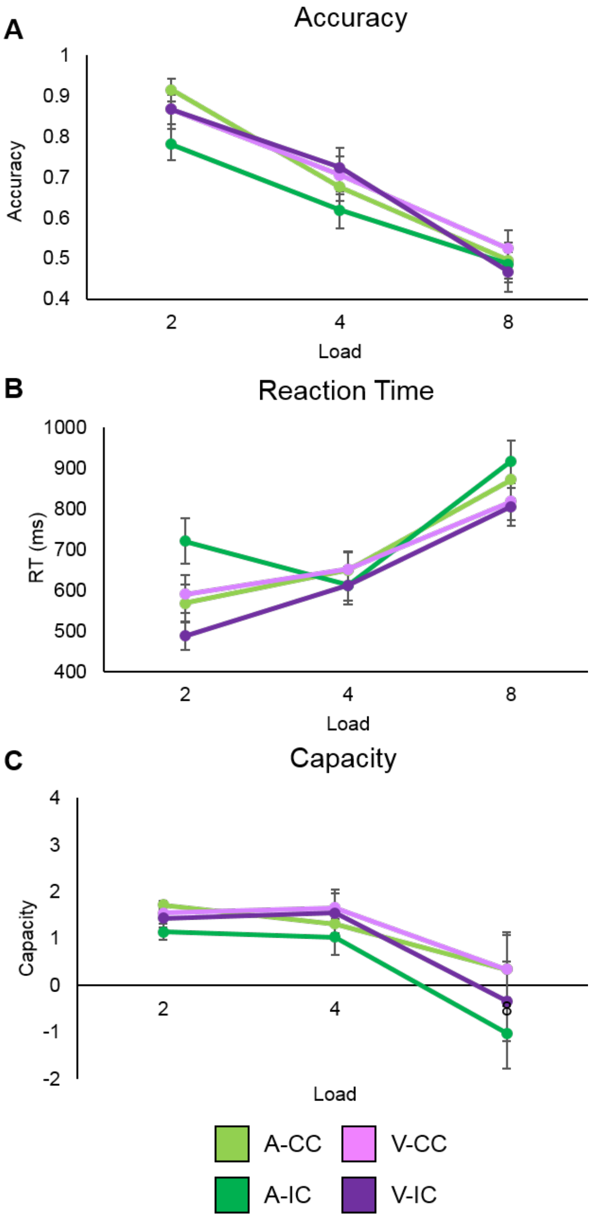


Figure 2.4. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants when presented with congruent and incongruent stimuli and tested on the auditory (A-CC & A-IC) and visual (V-CC & V-IC) working memory. Error bars represent the standard error of the mean.

(F(1,34)=66.85, $p<0.001$, η_p^2=0.663), a marginal effect of condition (F(1,34)=2.43, p=0.07, η_p^2=0.07), but not a significant interaction of load x condition (F(1,34)=0.715, p=0.638, η_p^2=0.02). The pattern of results shows that congruent stimuli did provide an advantage when tested on auditory working memory (condition F(1,34)=4.85, p=0.03, η_p^2=0.13, load F(1,34)=38.17, $p<0.001$, η_p^2=0.53, load x condition interaction F(1,34)=1.58, p=0.21, η_p^2=0.04), but not visual working memory (condition F(1,34)=0.19, p=0.67, $\eta_p^2<0.01$, load F(1,34)=34.94, $p<0.001$, η_p^2=0.51, load x condition interaction F(1,34)=0.33, p=0.72, η_p^2=0.01). We found that participants had significantly higher auditory working memory when presented with congruent stimuli compared to incongruent stimuli at a load of 2 (t(34)=3.22, $p<0.01$) items.

Reaction Time

Visual inspection of Figure 2.4B reveals that reaction times did not differ when presented with congruent versus incongruent stimuli and tested on auditory or visual working memory at loads of four or eight items. This analysis showed results from trials in which participants were presented with congruent and incongruent stimuli but were only tested on auditory (A-CC, A-IC) or visual stimuli (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,34)=47.49, $p<0.001$, η_p^2=0.58), but no main effect of condition (F(1,34)=3.97, p=0.1, η_p^2=0.11), nor a significant load x condition interaction (F(1,34)=1.90, p=0.08, η_p^2=0.05).

Capacity

Figure 2.4C shows that capacity estimates plateaued after a load of two items and declined after a load of four items for all congruent and incongruent conditions. This analysis showed results from trials in which participants were presented with congruent and incongruent stimuli but were only tested on auditory (A-CC, A-IC) or visual stimuli (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,34)=11.78,

$p<0.001$, $\eta_p^2=0.26$), but no main effect of condition ($F(1,34)=1.92$, $p=0.13$, $\eta_p^2=0.05$), nor a significant load x condition interaction ($F(1,34)=0.33$, $p=0.92$, $\eta_p^2=0.01$).

Brief Discussion

In Experiment 1, we had a memory load of two, four, or eight items on a given trial, and participants were not cued on which modality (auditory or visual) they were going to be tested on in the multisensory condition. We did not find evidence of a multisensory advantage in this experiment. Furthermore, contrary to our predictions, we did not find enhanced working memory when presented with congruent stimuli, which are in line with crossmodal correspondences (Spence, 2011). Existing literature demonstrates that the number of available attentional resources influence multisensory processing (Alsius et al., 2014). Thus, it is possible that the participant's attentional and memory resources were saturated as they were not oriented the sensory modality to attend to. Rather, they had to hold visual and information in their memory. For this reason, we hypothesized that we would find a benefit for multisensory stimuli if participants were cued on the modality that they were going to be tested on in the multisensory condition. We have tested this hypothesis in Experiment 2.

Experiment 2

Methods

Thirty-five adult participants were recruited for this experiment. All recruited participants had normal or corrected-to-normal vision and hearing and were naïve to the purposes of the experiment. Thirty-four participants (mean=20.08 years, M=2, F=32) were included in the final data set after the data cleaning analyses were conducted. Participants were presented with a randomly selected array of two, four, or eight sounds or color patches. The procedure for this study followed the experimental design as described in the General Methods section. Unlike in Experiment 1, participants were cued to the modality (auditory or visual) that was going to be tested in the multisensory condition (AV-Test A or AV-Test V). Participants were presented with a set of congruent stimuli on half of the trials, and incongruent stimuli on the other half of the trials in a randomized manner.

Results

Effects of Modality

Accuracy

As can be seen in Figure 2.5A, accuracy decreased as the memory load increased for all conditions. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,33)=215.08, p<0.001, η_p^2=0.87), and condition (F(1,33)=12.28, p<0.001, η_p^2=0.27), and a significant interaction of load x condition (F(1,34)=3.32, p<0.01, η_p^2=0.09). Visual inspection of the Figure 2.5A shows that participants had enhanced performance when presented with multisensory stimuli in the visual and auditory working memory domains. More specifically, accuracy was highest in the AV-test V condition, followed by the visual alone, and then AV- test A and auditory alone condition. To directly contrast

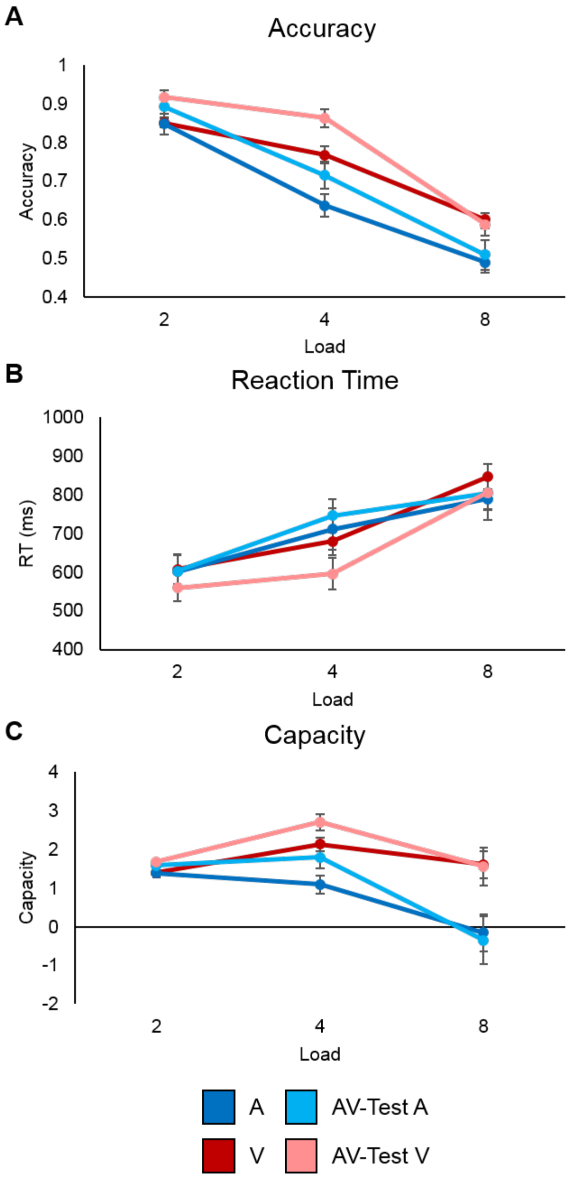


Figure 2.5. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants in the unisensory (V, A) or multisensory (AV-test A, AV-test V) conditions. Error bars represent the standard error of the mean.

performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. When comparing accuracy for unisensory auditory and visual conditions, we found a significant main effect of main effect of condition (F(1,33)=15.03, p<0.001, η_p^2=0.31), load (F(1,33)=111.56, p<0.001, η_p^2=0.77), and a significant load x condition interaction (F(1,33)=5.14, p<0.01, η_p^2=0.14). Specifically, accuracy was higher in the visual alone condition than in the auditory alone condition at loads of 4 (t(33)=3.73, p<0.01), and 8 (t(33)=3.41, p<0.01) items. When comparing accuracy for unisensory visual and AV-Test V conditions, a 2 (condition) x 3 (load) ANOVA revealed a significant main effect of condition (F(1,33)=8.86, p<0.01, η_p^2=0.21), load (F(1,33)=117.02, p<0.001, η_p^2=0.78), and a significant load x condition interaction (F(1,33)=4.56, p=0.01, η_p^2=0.12). Accuracy was significantly higher for the multisensory, AV-Test V, and compared to the unisensory visual condition at loads of 2 (t(33)=3.14, p<0.01), and 4 (t(33)=3.69, p<0.01) items. We did not find evidence of an auditory multisensory enhancement, as unisensory auditory and AV-Test A were not significantly different (condition F(1,33)=2.30, p=0.10, η_p^2=0.08, load F(1,33)=95.91, p<0.001, η_p^2=0.74, load x condition interaction F(1,33)=0.50, p=0.61, η_p^2=0.02).

Reaction Time

Our results show that the reaction time of participants is significantly different at different loads but does not vary by condition of the stimuli presented (Figure 2.5B). A 3 (load) x 4 (condition) ANOVA revealed a main effect of load (F(1,33)=44.82, p<0.001, η_p^2=0.57), but not a main effect of condition (F(1,33)=1.31, p=0.28, η_p^2=0.04), nor a significant interaction of load x condition (F(1,33)=2.06, p=0.6, η_p^2=0.06).

Capacity

Visual inspection of Figure 2.5C reveals that capacity estimates began to decline after a load of four items for all conditions, although the slope of the decline changed by the sensory modality tested. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,33)=18.69, p<0.001, η_p^2=0.36), and condition (F(1,33)=9.522, p<0.001, η_p^2=0.22), and a significant interaction of load x condition (F(1,33)=3.701, p<0.01, η_p^2=0.10). Differences between capacity estimates between the multisensory and unisensory stimuli is negligible at the lowest load of two items, when gets larger at four items, and diminishes again at a load of eight items. To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. This analysis revealed a significant difference in capacity estimates between the unisensory visual and auditory conditions (condition F(1,33)=18.23, p<0.001, η_p^2=0.36, load F(1,33)=6.99, p<0.01, η_p^2=0.18, load x condition interaction F(1,33)=7.22, p<0.01, η_p^2=0.18) at loads of 4 (t(33)=3.73, p<0.01), and 8 (t(33)=3.41, p<0.01) items. We also found evidence of a visual multisensory enhancement, as unisensory visual and AV-Test V were significantly different (condition F(1,33)=1.49, p=0.23, η_p^2=0.04, load F(1,33)=6.40, p<0.01, η_p^2=0.16, load x condition interaction F(1,33)=0.83, p=0.44, η_p^2=0.02) when compared at a loads of 2 (t(33)=3.24, p<0.01), 4 (t(33)=2.44, p=0.02) items. Lastly, we did not find evidence of an auditory multisensory enhancement, as unisensory auditory and AV-Test A were not significantly different (condition F(1,33)=0.55, p=0.46, $\eta_p^2=0.02$, load F(1,33)=20.79, p<0.001, $\eta_p^2=0.39$, load x condition interaction F(1,33)=0.87, p=0.43, $\eta_p^2=0.03$).

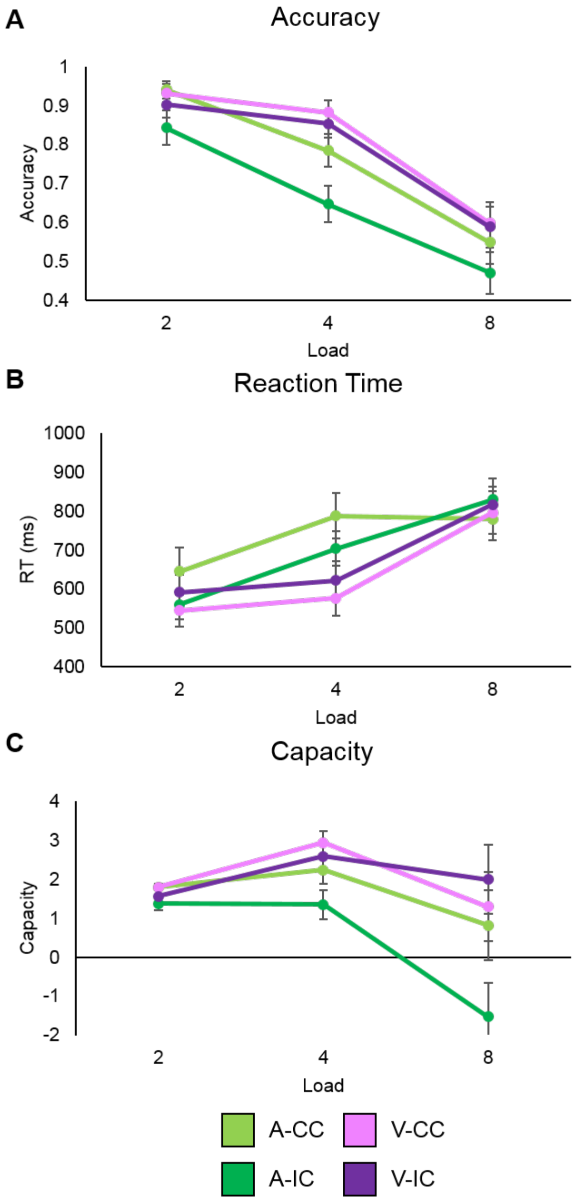


Figure 2.6. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants when presented with congruent and incongruent stimuli and tested on the auditory (A-CC & A-IC) and visual (V-CC & V-IC) working memory. Error bars represent the standard error of the mean.

<u>Effects of Congruency</u>

Accuracy

Figure 2.6A shows participant performance when presented with congruent versus incongruent stimuli in the multisensory condition. A 3 (load) x 4 (condition) ANOVA revealed a main effect of load (F(1,33)=90.68, p<0.001, $\eta_p^2=0.73$), a main effect of condition (F(1,33)=7.618, p<0.001, $\eta_p^2=0.19$), but not a significant interaction of load x condition (F(1,33)=0.87, p=0.52, $\eta_p^2=0.026$). A 2(condition) x 3(load) ANOVA showed that participants had enhanced auditory working memory (condition F(1,33)=8.58, p<0.01, $\eta_p^2=0.21$, load F(1,33)=36.22, p<0.001, $\eta_p^2=0.52$, load x condition interaction F(1,33)=0.26, p=0.78, $\eta_p^2<0.01$), but not visual working memory performance when presented with congruent stimuli (condition F(1,33)=0.37, p=0.55, η_p^2=0.01, load F(1,33)=74,65, p<0.001, η_p^2=0.69, load x condition interaction F(1,33)=0.03, p=0.97, η_p^2=0.001). Specifically, we found a significant difference in accuracies when the congruent and incongruent stimuli tested on auditory working memory (A-CC & A-IC) when compared at a load of 4 (t(33)=2.51, p=0.02) items.

Reaction Time

Visual inspection of Figure 2.6B shows that participants had longer reaction times for congruent stimuli in the auditory, but not visual working memory domain. This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). As shown in Figure 2.6B, reaction times were longest in the A-CC condition, followed by the A-IC, and then V-IC and V-CC condition A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,33)=20.22, p<0.001, η_p^2=0.38), a marginal effect of condition (F(1,33)=2.46, p=0.07, η_p^2=0.07), nor a significant interaction of load x condition (F(1,33)=2.07, p=0.06, η_p^2=0.06). To directly contrast performance between congruent and incongruent conditions, we ran a series of 2 (condition) x 3 (load) ANOVAs. We did not significant statistical differences when reaction times were compared for congruent and incongruent stimuli in the auditory (condition F(1,33)=1.57, p=0.22, η_p^2=0.05, load F(1,33)=9.41, p=0.24, η_p^2=0.04, load x condition interaction F(1,33)=1.62, p=0.21, η_p^2=0.05) or visual (condition F(1,33)=1.42, p=0.24, η_p^2=0.04, load F(1,33)=21.02, p<0.001, η_p^2=0.39, load x condition interaction F(1,33)=0.10, p=0.91, η_p^2<0.01) domain.

Capacity

Figure 2.6C shows participant's capacity estimates when presented with congruent versus incongruent stimuli in the multisensory condition. A 3 (load) x 4 (condition) ANOVA revealed a main effect of load (F(1,33)=10.98, p<0.001, η_p^2=0.25), a significant effect of condition (F(1,33)=5.79, p<0.01, η_p^2=0.15), but not a significant interaction of load x condition (F(1,33)=1.74, p=0.12, η_p^2=0.05). Similar to results in the accuracy domain, we found that while congruent stimuli increase memory capacity in the auditory working memory domain, they do not in the visual working memory domain. A 2 (condition) x 3 (load) ANOVA in the auditory domain revealed a significant main effect of condition (F(1,33)=6.96, p=0.01, η_p^2=0.17), load (F(1,33)=9.91, p<0.001, η_p^2=0.23), but not a significant load x condition interaction (F(1,33)=1.73, p=0.19, η_p^2=0.05). We found significant difference in capacity estimates when the congruent and incongruent stimuli (A-CC & A-IC) tested on the auditory working memory were compared at a load of 2 t(33)=2.17, p=0.04) , and 4 (t(33)=2.04, p=0.05) items. In the visual working memory domain, A 2 (condition) x 3 (load) ANOVA revealed a significant main effect of load (F(1,33)=4.76, p=0.01, η_p^2=0.13), but not of condition (F(1,33)=0.01, p=0.94, η_p^2<0.001), or a load x condition interaction (F(1,33)=0.41, p=0.67, η_p^2=0.01).

Brief Discussion

In Experiment 2, we cued participants on the sensory modality they were going to be tested on the multisensory condition while keeping all other parameters identical to those in Experiment 1. We did find a significant multisensory advantage for auditory and visual working memory stores in this experiment. More specifically, this advantage was greater for visual than auditory working memory. Furthermore, in line with our hypotheses, we did find enhanced working memory when presented with congruent stimuli, which are in line with crossmodal correspondences. The results from this experiment are consistent with the dual coding account of

MSI and memory, which states that having two codes (e.g., verbal and visual codes) for information leads to memory enhancements (Paivio, 1969). By this account, the dual coding mechanism produces enhanced memory recall for multisensory stimuli due to the ability to access a greater amount of information about the stimulus due to the two sensory traces. To better understand whether working memory advantages for multisensory stimuli were affected by the memory load, we presented participants with a larger array of stimuli, and higher memory load in Experiment 3.

Experiment 3

Methods

Thirty-eight adult participants were recruited for this experiment. All recruited participants had normal or corrected-to-normal vision and hearing and were naïve to the purposes of the experiment. Thirty-three participants (mean=20.09 years, M=9, F=24) were included in the final data set after the data cleaning analyses were conducted. Participants were presented with a randomly selected array of four, six, or eight sounds or color patches. The procedure for this study followed the experimental design as described in the General Methods section, and participants were cued on the modality (auditory or visual) that they were going to be tested on in the multisensory condition (AV-Test A or AV-Test V). Due to the increased difficulty of the task with the increase in memory load, participants were presented with practice trials at the beginning of each block of the experiment which were excluded from the final analyses. Participants were presented with a set of congruent stimuli on half the trials, and incongruent stimuli on the other half of the trials in a randomized manner.

Results

Effects of Modality

Accuracy

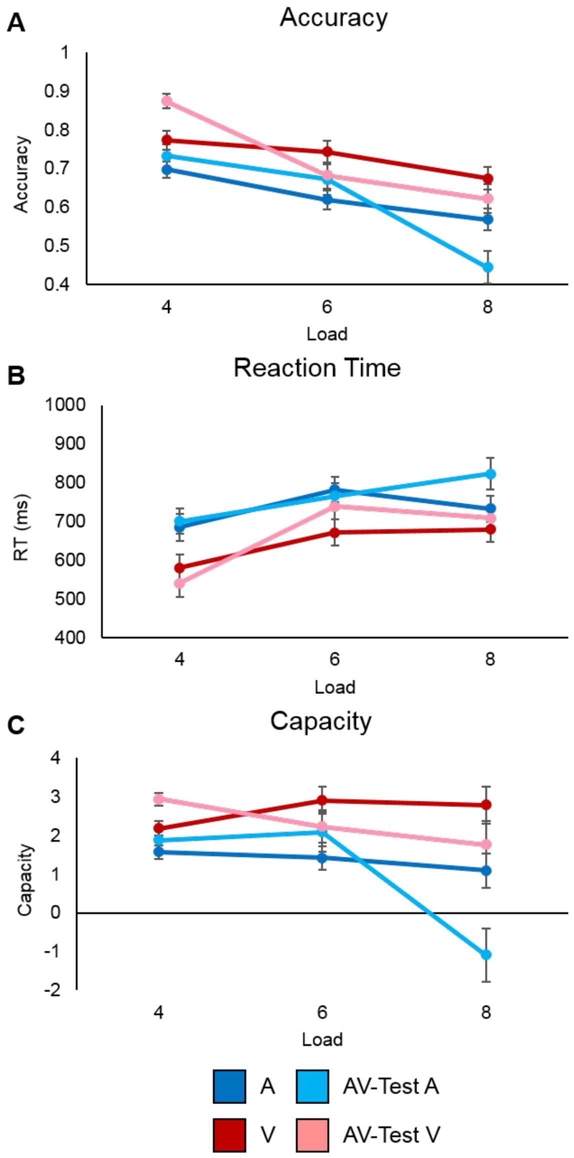


Figure 2.7. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants in the unisensory (V, A) or multisensory (AV-test A, AV-test V) conditions. Error bars represent the standard error of the mean.

Overall, it can be seen in Figure 2.7A that accuracy of participants decreased as the memory load increases, however, this decline is much more distinct for the AV-test A condition than the others. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,32)=, p<0.001, η_p^2=0.59), and condition (F(1,32)=13.14, p<0.001, η_p^2=0.29), and a significant interaction of load x condition (F(1,32)=4.82, p<0.01, η_p^2=0.131. To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. Interestingly, performance was enhanced when presented with multisensory stimuli in the visual working memory domain at low loads, but hindered when presented with multisensory stimuli in the auditory working memory domain at high loads. When comparing accuracy for unisensory visual and AV-Test V conditions, a 2 (condition) x 3 (load) ANOVA revealed a significant main effect of load (F(1,32)=23.68, p<0.001, η_p^2=0.43), but not of condition (F(1,32)=0.04, p=0.84, η_p^2=0.001), and a significant load x condition interaction (F(1,32)=5.59, p<0.01, η_p^2=0.15). Similarly, when comparing accuracy for unisensory auditory and AV-Test A conditions, a 2 (condition) x 3 (load) ANOVA revealed a significant main effect of load (F(1,32)=26.01, p<0.001, η_p^2=0.45), but not of condition (F(1,32)=0.26, p=0.62, η_p^2<0.01), and a significant load x condition interaction (F(1,32)=7.21, p<0.01, η_p^2=0.18). We found a significant difference in accuracy when the unisensory visual and AV-Test V were compared at a load of 4 (t(32)=3.31, p<0.01) items, and when unisensory audio and AV-Test A were compared at a load of 8 (t(32)=3.13, p<0.01) items. Lastly, when comparing accuracy for unisensory auditory and visual conditions, we found a significant main effect of condition (F(1,32)=22.21, p<0.001, η_p^2=0.41), and load (F(1,32)=9.88, p<0.001, η_p^2=0.24), and but not a significant load x condition interaction (F(1,33)=0.73, p=0.49, η_p^2=0.02). Specifically, we found significant differences in accuracy when the unisensory visual and auditory conditions were compared at a load of 4 (t(32)=3.46, p<0.01), 6 (t(32)=3.57, p<0.01) and 8 (t(32)=2.91, p<0.01) items.

Reaction Time

As seen in Figure 2.7B, reaction times were overall longer when tested on auditory working memory than visual workign memory. The differences between these were greater at the lowest load of 4 items, and diminished as the load increased. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,32)=18.22, p<0.001, η_p^2=0.36), condition (F(1,32)=6.92, p<0.001, η_p^2=0.18), and a significant interaction of load x condition (F(1,32)=2.84, p=0.01, η_p^2=0.0.08). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. When comparing accuracies for unisensory auditory and visual conditions, we found a significant main effect of condition (F(1,32)=6.08, p=0.02, η_p^2=0.16), load (F(1,32)=11.90, p<0.001, η_p^2=0.27), but not a significant load x condition interaction (F(1,32)=1.05, p=0.36, η_p^2=0.03). Specifically, this difference was significant when the unisensory auditory and visual conditions conditions were compared at a load of 4 (t(32)=2.22, p=0.03), and 6 (t(32)=2.48, p=0.02) items. When comparing accuracy for unisensory auditory and AV-Test A conditions, a 2 (condition) x 3 (load) ANOVA revealed a significant main effect of load (F(1,32)=5.29, p<0.01, η_p^2=0.14), not of condition (F(1,32)=0.99, p<=0.33 η_p^2=0.03), and did show a significant load x condition interaction (F(1,32)=4.81, p=0.04, η_p^2=0.13). We found significant difference in reaction times when these conditions were compared at a load of 8 (t(32)=2.51, p=0.04) items. We did not find evidence of an visual multisensory enhancement, as unisensory visual and AV-Test V were not significantly different (condition F(1,38)=0.36, p=0.55, η_p^2=0.01, load F(1,38)=2.25, p<0.001, η_p^2=0.41, load x condition interaction F(1,38)=4.81, p=0.01, η_p^2=0.13).

Capacity

Visual inspection of Figure 2.7C revealed that capacity estimates began to stabilize and slightly decline after a load of 4 items and declined more for the multisensory AV-Test A condition than other conditions. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,32)=7.91, p<0.01, η_p^2=0.20), condition (F(1,32)=11.57, p<0.001, η_p^2=0.27), and a significant interaction of load x condition (F(1,32)=6.25, p<0.001, η_p^2=0.16). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. Overall, we found that capacity estimates were significantly higher in the unisensory visual compared to unisenosry auditory modality (condition F(1,32)=19.73, p<0.001, η_p^2=0.38, load F(1,32)=0.42, p=0.66, η_p^2=0.01, load x condition interaction F(1,32)=2.21, p=0.12, η_p^2=0.07) at a loads of 4 (t(32)=3.46, p<0.01), 6 (t(32)=3.57, p<0.01) and 8 (t(32)=2.91, p<0.01) items. When comparing accuracy for unisensory auditory and AV-Test A, we found a significant main effect of load (F(1,32)=12.00, p<0.001, η_p^2=0.27), but not a main effect of condition (F(1,32)=1.59, p=0.22, η_p^2=0.05), and a significant load x condition interaction (F(1,32)=11.47, p<0.001, η_p^2=0.26). Specifically, this difference was significant when compared at a loads of 8 (t(32)=3.45, p<0.01) items. We did not find evidence of an visual multisensory enhancement, as unisensory visual and AV-Test V were not significantly different (condition F(1,32)=1.28, p=0.27, η_p^2=0.04, load F(1,32)=0.40, p=0.67, η_p^2=0.01, load x condition interaction F(1,32)=3.32, p=0.04, η_p^2=0.09) at a load of 2 (t(32)=2.96, p<0.01 items.

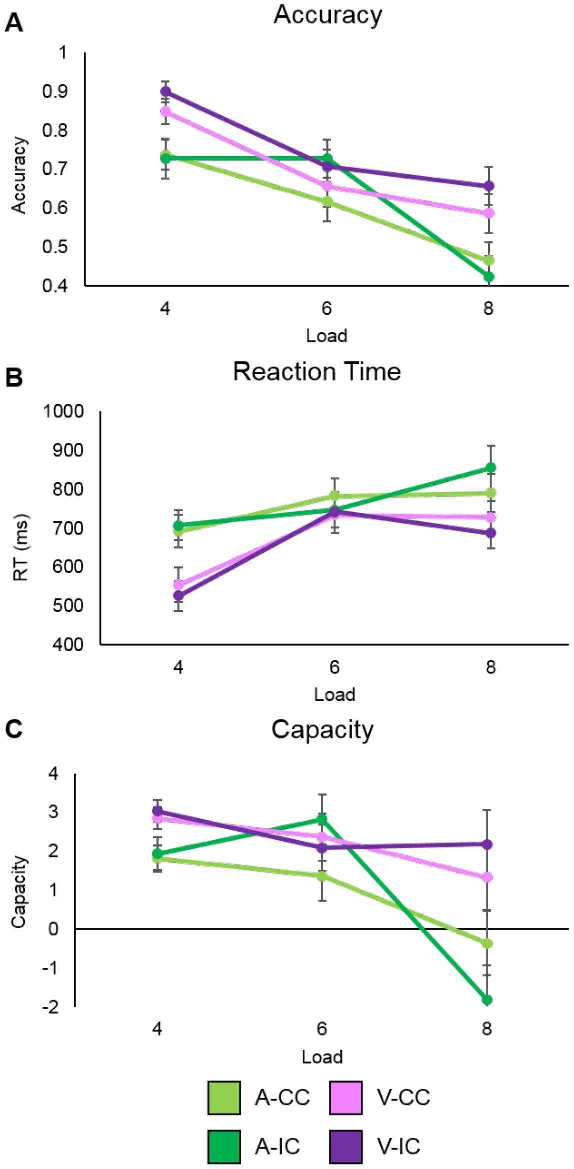


Figure 2.8. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants when presented with congruent and incongruent stimuli and tested on the auditory (A-CC & A-IC) and visual (V-CC & V-IC) working memory. Error bars represent the standard error of the mean.

<u>Effects of Congruency</u>

Accuracy

Overall, Figure 2.8A showed that participants' performance increased when presented with incongruent stimuli. This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (CC) stimuli but were only tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a main effect of load (F(1,32)=37.66, p<0.001, η_p^2=0.541), condition (F(1,32)=5.71, p<0.01, η_p^2=0.), and a marginally significant interaction of load x condition (F(1,32)=2.17, p=0.05, η_p^2=0.06). To directly contrast performance between congruent and incongruent conditions, we ran a series of 2 (condition) x 3 (load) ANOVAs. We did not find significant statistical differences when accuracy was compared for congruent and incongruent stimuli in the auditory (condition F(1,32)=0.28, p=0.60, η_p^2=0.01, load F(1,32)=21.73, p<0.001, η_p^2=0.40, load x condition interaction F(1,32)=1.74, p=0.18, η_p^2=0.05) or visual (condition F(1,32)=2.03, p=0.16, η_p^2=0.06, load F(1,32)=21.30, p<0.001, η_p^2=0.40, load x condition interaction F(1,32)=0.05, p=0.95, η_p^2=0.001) domain.

Reaction Times

Figure 2.8B showed participants' reaction times when presented with congruent versus incongruent stimuli in the multisensory condition. Overall, we can see that that participants had longer reaction times when tested on auditory working memory, although this difference diminished as the load increased. This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (IC) stimuli but were only tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,32)=15.01, p<0.001, η_p^2=0.32), condition (F(1,32)=7.69, p<0.001, η_p^2=0.19), but not a significant interaction of load x condition (F(1,32)=1.39, p=0.29, η_p^2=0.04). A 2 (condition) x 3 (load) ANOVA did not reveal significant differences when presented with congruent versus incongruent stimuli in auditory (condition F(1,32)=0.26, p=0.62, η_p^2<0.01, load F(1,32)=4.39, p=0.02, η_p^2=0.12, load x condition interaction F(1,32)=0.72, p=0.49, η_p^2=0.02) or visual (condition F(1,32)=0.65, p=0.43, η_p^2=0.02, load

F(1,32)=19.63, p<0.001, η_p^2=0.38, load x condition interaction F(1,32)=0.22, p=0.80, η_p^2<0.01) working memory.

Capacity

This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli but were only tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). As can be seen in Figure 2.8C, overall, capacity estimates did not differ by the congruency of the stimuli presented or the type of working memory tested. Interestingly, however, there was a large drop in capacity estimates for the A-IC condition at a load of 8 items. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,32)=12.13, p<0.001, η_p^2=0.28), condition (F(1,32)=4.27, p<0.01, η_p^2=0.12), and a significant interaction of load x condition (F(1,32)=3.40, p<0.01, η_p^2=0.10). To directly contrast performance between congruent and incongruent conditions, we ran a series of 2 (condition) x 3 (load) ANOVAs. We did not find significant statistical differences when capacity estimates were compared for congruent and incongruent stimuli in the auditory (condition F(1,32)=0.01, p=0.93, η_p^2=0.00, load F(1,32)=16.68, p<0.001, η_p^2=0.34, load x condition interaction F(1,32)=3.71, p=0.03, η_p^2=0.10) or visual (condition F(1,32)=,0.211 p=0.65, η_p^2<0.01, load F(1,32)=2.01, p=0.14, η_p^2=0.06, load x condition interaction F(1,32)=0.49, p=0.61, η_p^2=0.02) domains.

Brief Discussion

To better understand whether working memory advantages for multisensory stimuli were affected by the memory load, in Experiment 3 we presented participants with four, six, or eight auditory, visual, or audiovisual stimuli. All other aspects of the experimental paradigm remained identical to those of Experiment 2. In this experiment, we found a multisensory advantage in visual working memory, specifically at the at smallest memory load of four items. However, most of the other multisensory advantages, including those in the auditory working memory domain, diminished. Overall, studies investigating effects of congruency in multisensory processing with simple stimuli have noted that participants have higher memory performance when presented with congruent compared to incongruent multisensory stimuli (Calvert et al., 2004; Lacey, Martinez, McCormick, & Sathian, 2016; Spence, 2007). In line with previous literature, we had hypothesized that we would find evidence for increased working memory abilities when presented with congruent stimuli. Interestingly, contrary to our predictions, we found higher working memory capacity for incongruent stimuli. This advantage for incongruent stimuli was most pronounced at the load of six items in auditory working memory, and at the highest load of eight items for visual working memory. To further understand the influence of congruency for crossmodal stimuli, we manipulated the congruency of the stimuli presented in Experiment 4.

Experiment 4

Methods

Thirty-nine adult participants were recruited for this experiment. All recruited participants had normal or corrected-to-normal vision and hearing and were naïve to the purposes of the experiment. All participants were included in the final data set. Participants were presented with a randomly selected array of four, six, or eight sounds and color patches. The procedure for this study followed the experimental design as described in the General Methods except, participants were only presented with multisensory audiovisual stimuli and were cued on the modality (auditory or visual) that they were going to be tested on (AV-Test A or AV-Test V). Participants were presented with a set of congruent stimuli (distance 0; e.g. light yellow color patch presented with a high pitch saxophone tone), stimuli that had within category incongruency (distance 1; e.g. light yellow color patch presented with a low pitch saxophone tone) and stimuli that had cross-category incongruency (distance 2; e.g. light yellow color patch presented with a medium pitch tuba tone) in a randomized manner. In this way, we were able to systematically test the effects of stimulus congruency on working memory performance.

Results

<u>Accuracy</u>

As can be seen in Figure 2.9A, accuracy of participants decreased as the memory load increased across all conditions. Upon visual inspection, it was seen that overall, accuracies were higher when tested in the visual working memory domain than the auditory working memory domain. To directly contrast performance when presented with congruent, within-category incongruent, and cross-category

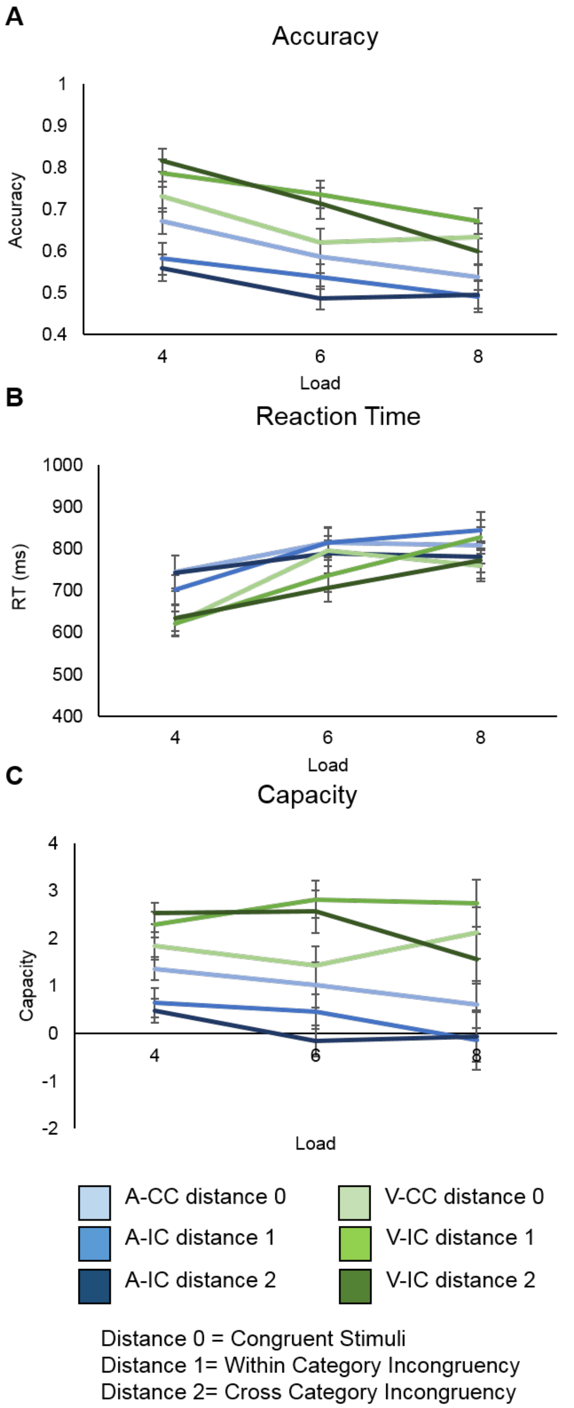


Figure 2.9. Working memory accuracies (A) reaction times (B) and capacity estimates (C) when participants were tested with congruent (distance 0), within-category incongruent (distance 1), and cross-category incongruent (distance 2) stimuli in the multisensory, AV, condition. Error bars represent the standard error of the mean.

incongruent, we ran 3 (congruency/distance) x 3(load) ANOVAs when tested on auditory and visual working memory. In the auditory domain, we found a significant main effect of load (F(1,38)=6.94, p<0.01, η_p^2=0.15), and distance (F(1,38)=5.35, p<0.01, η_p^2=0.12), but not a significant interaction of load x distance (F(1,38)=0.50, p=0.74, η_p^2=0.01). Specifically, we found a significant difference when congruent stimuli (distance 0) were compared with incongruent stimuli (distance 2) at a load of 4 (t(38)=3.05, p<0.01), and 6 (t(38)=2.61, p=0.01) items. A 3 (load) x 3 (distance) ANOVA in the visual domain revealed a significant main effect of load (F(1,38)=14.79, p<0.001, η_p^2=0.28), and distance (F(1,38)=4.04, p=0.02, η_p^2=0.10), but not a significant interaction of load x distance (F(1,38)=1.70, p=0.15, η_p^2=0.04). We found a significant difference when congruent stimuli (distance 0) were compared with incongruent stimuli with a distance of 1 (within-category incongruency) at a load of 6 (t(38)=2.61, p=0.01) items.

Reaction Time

As can be seen in Figure 2.9B, overall, reaction times of participants increased as the memory load increased across all conditions. Reaction times for all conditions are between approximately the 600 ms and 800 ms range. To directly contrast performance when presented with congruent, within-category incongruent, and cross-category incongruent, we ran 3 (congruency/distance) x 3(load) ANOVAs when tested on auditory and visual working memory. We did not find significant differences in the auditory working memory domain (load F(1,38)=4.79, p=0.01, η_p^2=0.11, distance F(1,38)=0.24, p=0.79, η_p^2=0.01, load x distance interaction F(1,38)=0.91, p=0.46, η_p^2=0.02). We found differences in capacity estimates when congruent stimuli (distance 0) were compared with incongruent stimuli (distance 2; cross-category incongruency) at loads of 4 (t(38)=3.05, p<0.01), and 6 (t(38)=2.61, p=0.01) items. A 3 x 3 ANOVA in the visual modality revealed a significant main effect of load (F(1,38)=20.86, p<0.001, η_p^2=0.35), but not a significant effect of distance (F(1,38)=0.55, p=0.58, η_p^2=0.01), nor a significant interaction of load x distance (F(1,38)=2.18, p=0.07, η_p^2=0.0.05). Specifically, we found a significant difference in reaction times when congruent stimuli (distance 0) with incongruent stimuli with a distance of 1 (within-category incongruency) at a load of 6 (t(38)=2.61, p=0.01) items.

Capacity

As can be seen in Figure 2.9C, overall, capacity estimates were lower when tested on auditory working memory than when tested on visual working memory, with the congruency (distance) of the stimuli having a differential effect on the capacity estimates across the three loads. To directly contrast performance when presented with congruent, within-category incongruent, and cross-category incongruent, we ran 3 (congruency/distance) x 3(load) ANOVAs when tested on auditory and visual working memory. This analysis in the auditory domain revealed did not show a significant main effect of load (F(1,38)=2.42, p=0.09, η_p^2=0.06), but showed a significant effect of distance (F(1,38)=3.87, p=0.03, η_p^2=0.09), and a non-significant interaction of load x distance (F(1,38)=0.22, p=0.93, η_p^2=0.01). A 3 (load) x 3 (distance) ANOVA in the visual domain revealed a non-significant main effect of load (F(1,38)=0.08, p=0.93, η_p^2=0.002), a significant effect of distance (F(1,38)=3.23, p=0.05, η_p^2=0.08), and a non-significant interaction of load x distance (F(1,38)=1.73, p=0.15, η_p^2=0.04). Specifically, we found a significant difference in capacity estimates when congruent stimuli (distance 0) were compared with incongruent stimuli with a distance of 1 (within-category incongruency) at a load of 8 items (t(38)=2.17, p=0.04), and with incongruent stimuli with a distance of 2 (cross-category incongruency) at a load of 6 items (t(38)=2.10, p=0.04).

Brief Discussion

We manipulated the congruency of the stimuli presented in Experiment 4 by presenting congruent, within-category incongruent, and cross-category incongruent stimuli in the change detection paradigm with a memory load of four, six, or eight items. Overall, we found that working memory performance was higher for visual compared to auditory stimuli. In line with previous research, we found that performance was highest when presented with congruent stimuli (distance 0), followed by within-category incongruent (distance 1), and cross-category incongruent (distance 2) when tested on auditory working memory. However, we found that performance was higher for incongruent (distance 1 and 2) compared to congruent stimuli (distance 0) when tested on visual working memory, particularly at the lowest load of four items. This result indicates that congruent auditory information is not helping visual working memory, while congruent visual information is indeed helping auditory working memory.

General Discussion

In this study, we tested the impact of simple stimuli on working memory capacity by varying load and congruency in a change detection paradigm. In Experiment 1, we had a memory load of two, four, or eight items on a given trial, and participants were not cued on which modality (auditory or visual) they were going to be tested on in the multisensory (audiovisual) condition. In Experiment 2, we cued participants on the sensory modality they were going to be tested on in the multisensory condition while keeping all other parameters identical to those in Experiment 1. We did not find a significant multisensory advantage for auditory and visual working memory stores in Experiment 1, but did find a multisensory advantage in Experiment 2.

The difference in findings between Experiments 1 and 2, namely for the audiovisual condition, may be explained by the impact of attentional mechanisms on MSI. While a few studies have investigated the involvement of attention in multisensory memory, one relevant study found that the available attentional resources influence multisensory processing (Alsius et al., 2014). It is therefore plausible that the absence of the multisensory memory enhancement in Experiment 1 is due to the lack of available attentional resources. Specifically, because participants were not cued to which sensory modality was going to be tested in AV trials, their attention was divided between the two modalities, and their mnemonic resources were overburdened by twice the amount of information compared with Experiment 2. Therefore, participants were not able to retrieve the appropriate traces from their memory to detect a change in the tested modality. A larger discussion of the involvement of attentional mechanisms in multisensory memory processes is included in Chapter 5.

To better understand whether working memory advantages for multisensory stimuli were affected by the memory load, in Experiment 3 we presented participants with four, six, or eight auditory, visual, or audiovisual stimuli. In this experiment, we found a multisensory advantage in visual working memory, specifically at the at smallest memory load of four items. Additionally, we found higher visual working memory capacity for incongruent stimuli at larger memory loads. To test this finding, we manipulated the congruency of the multisensory stimuli presented in Experiment 4 while keeping other parameters identical to those in Experiment 3. Overall, we found that working memory performance was higher for visual compared to auditory stimuli. As expected, we found that performance was highest when presented with congruent stimuli compared to incongruent stimuli when tested on auditory working memory. However, we found that performance was higher for incongruent compared to congruent stimuli when tested on visual working memory, particularly at the lowest load of four items.

Overall, previous studies investigating effects of congruency in multisensory processing with simple stimuli have noted that participants had higher memory performance when presented with congruent compared to incongruent multisensory stimuli (Calvert et al., 2004; Lacey, Martinez, McCormick, & Sathian, 2016; Spence, 2007). We found variable results across experiment from test of congruency in this chapter, therefore it is important to contextualize them with each other. Along with a lack of a memory advantage for multisensory stimuli overall, we found negligible advantages when presented with congruent stimuli in Experiment 1. On the contrary, we found a sustained memory benefit for multisensory stimuli, and found that participants had significantly higher auditory working memory performance when presented with congruent stimuli in Experiment 2. However, we also found that participants did not have enhanced visual working memory when presented with congruent stimuli in this experiment. Findings related to congruency in Experiment 3 were comparable to Experiment 1, in that we did not find an advantage when presented with congruent stimuli. In fact, with a higher memory load in Experiment 3, we found enhanced memory processing when presented with incongruent stimuli.

Our findings in the explicit congruency manipulation in Experiment 4 match with the existing literature in the domain of auditory working memory as we found that participants had the highest memory performance when presented with congruent stimuli. On the contrary, we found that participants had the lowest memory performance when presented with congruent stimuli and tested on visual working memory. This result indicates that congruent auditory information is not helping visual working memory, while congruent visual information is indeed helping auditory working memory. Previous research has shown that visual working memory has higher retrieval rates and capacity estimates (Fougnie & Marois, 2011; Saults & Cowan, 2007) compared to auditory working memory. For this reason, it could be true that participants don't rely on crossmodal congruence to complete the task in the visual domain. In fact, since participants were cued on the modality they were going to be tested on, they could've found that incongruent stimuli were easier to unbind compared to congruent stimuli, which were linked by implicit crossmodal correspondences. A larger discussion of the enhancements found when presented with incongruent stimuli is included in Chapter 5. While simple stimuli may start to reveal the cognitive mechanisms at work, it is important to further understand the effect of crossmodal congruency using more naturalistic and semantically meaningful stimuli. For this reason, we tested the impact of complex stimuli on working memory in the series of studies described in Chapter 3.

Chapter 3: The Impact of Complex Stimuli on Multisensory Integration and Working Memory

Introduction

The ability to parse through a multitude of sensory information to identify a relevant stimulus is highly dependent on processes of multisensory integration (MSI). In the last two decades, the rate of studies investigating the integration process of visual and auditory information has accelerated. This research boom has dramatically advanced our knowledge of MSI in the audio-visual domain through behavioral and neural studies (Newell, Mamassian, & Alais, 2010; Stein & Meredith, 1993). The ability to integrate information from multiple senses is highly dependent on the properties of the stimulus presented. Simple stimuli (e.g. flashes, gratings, spots of light, beeps and tones) which can be easily parameterized have formed the basis of many experimental studies and have been instrumental in determining stimulus dependent responses, and revealing important findings about the organization of cortical neurons and the retina (Barutchu et al., 2010; Hartline, 1940; Hillock et al., 2011; Hubel & Wiesel, 1962).

Moving beyond simple stimuli, recent MSI research has turned to the use of more complex stimuli, which use semantics, to understand crossmodal mechanisms (Beauchamp, 2005; Erickson, Heeg, Rauschecker, & Turkeltaub, 2014; Hocking & Price, 2008). Several studies have used auditory speech and visual facial expressions to understand more routine interactions in crossmodal processing (McGurk & MacDonald, 1976; Sumby & Pollack, 1954). Other studies have used more dynamic stimuli such as tools, which include complex processes such as biological motion, and the presentation of an asocial body part (e.g. a hand holding/using a hammer) (Stevenson & Wallace, 2013). While these studies have begun to step away from the basic sensory stimuli such as flashes, pure tones, checker boards, and noise bursts, which are limited in scope, there remains a dearth of studies exploring crossmodal stimuli that are naturally occurring in the environment.

One example of how researchers have begun to study the integration of multiple complex AV stimuli is seen in the literature on face-to-face communication. Visual information can be extracted from speech by the movements of the mouth, eyes, and head. Therefore, normal speech lends itself to be a natural stimulus set to analyze naturalistic audio-visual interactions. Research has found that individuals rely on visual properties of speech when the auditory properties are not salient, such as in a noisy room. Evidence in support of this notion notes that the visual percept of the face enhances the auditory discriminability of speech compared to the auditory stimuli alone (Sumby & Pollack, 1954). Studies have noted that non-human primates can detect correspondences between faces and voices, and use this information to guide actions by forming one, unified multisensory percept (Ghazanfar & Logothetis, 2003; Izumi & Kojima, 2004; Plant, 1999). More specifically, rhesus monkeys use their ability to recognize facial and vocal calls to facilitate communication in a natural setting (Ghazanfar & Logothetis, 2003). Researchers have also hypothesized that this ability to match complex and naturalistic crossmodal stimuli may be the precursor to the ability that humans have to match speech and facial expressions.

The McGurk Effect is a demonstration of the faction that visual features can alter the perception of auditory speech (McGurk & MacDonald, 1976). It is important to note that the type of auditory stimulus presented impacts audio-visual speech perception (Grant & Seitz, 1998). For example, studies of the McGurk effect present a range of auditory stimuli from nonsense syllables to full sentences (Grant & Seitz, 1998; Sams et al., 1998; Van Engen et al., 2017). Due to individual differences in the strategy employed by participants in such experiments, scores on studies that use nonsense syllables do not correlate with scores on studies that use full sentences (Grant & Seitz, 1998). A recent study (Van Engen, Dey, Sommers, & Peelle, n.d.) has suggested that tasks of the McGurk Effect should not be used to study complex AV MSI as there are individual differences in susceptibility to this effect and the stimuli used to test this effect do not mimic naturally occurring stimuli.

Even with nuances and variations in responses, complex stimuli provide a cleaner reflection of the intricacies of audio-visual stimuli compared to simple or artificial stimuli. Along with speech stimuli, researchers have used stimuli with cultural relevance such as pictures of tools as visual stimuli (James et al., 2011; Stevenson & James, 2009; Stevenson & Wallace, 2013). These studies use various types of tools such as hammers, paper cutters, paper towel dispensers, and scissors. Another study used auditory and visual clips of water splashing as a naturalistic multisensory stimuli (Senkowski, Saint-Amour, Kelly, & Foxe, 2007). Another element included in making the stimuli more naturalistic is that of motion. For example, James and colleagues (2011) made videos of tool use, and Senkowski and colleagues (2007) recorded a drop of water falling on to a water surface. By using naturalistic stimuli in motion, these studies have found that MSI takes places at various states of processing from perception to semantics, as well as in various neural regions (Senkowski et al., 2007).

The neural representation of MSI with naturalistic stimuli have been broadly recognized to be stored the superior temporal cortex (Stevenson et al., 2011). A range of studies have localized the neural correlates to various brain regions such as the dorsal and ventral steams (Hickok & Poeppel, 2007; Rauschecker, 2012), the superior temporal sulcus (Beauchamp et al., 2010; Hein & Knight, 2008; Hocking & Price, 2008; Raij, Uutela, & Hari, 2000), and the inferior frontal gyrus (Ojanen et al., 2005; Sekiyama et al., 2003). A meta-analysis of conflicting (different signals) versus validating (complementary signals) audio-visual speech signals identified a network in the dorsal stream regions for conflicting stimuli and in the ventral stream visual regions of the occipital and temporal lobes for validating stimuli (Erickson et al., 2014). Overall, the use of various types of complex stimuli have demonstrated that naturally occurring AV interactions and are key determinants of successful multisensory integration.

The Current Study

The aim of this study is to extend the findings of Chapter 2 by testing the impact of multisensory integration on working memory using natural stimuli that are more complex and semantically meaningful than simple stimuli. In a series of four experiments, we used pictures of animals, and their corresponding animal sounds. Additionally, we built our stimulus set on size-pitch crossmodal correspondences that are naturally occurring in the environment: larger animals tend to produce lower pitched sounds and smaller animals tend to produce higher pitched sounds (Ohala, 1983). After developing an analogous stimulus set to that used in Chapter 2, we replicated the working memory change detection paradigm used to test the impact of simple stimuli on working memory (Chapter 2) with this new stimulus set.

We expected to find differences at the modality-specific level. In particular, we expected participant to have an advantage in working memory when presented with multisensory compared to unisensory stimuli in the auditory (AV-Test A compared to A) and visual (AV-Test V compared to V) domains. In line with findings from previous research and the dual-coding theory (Brunetti et al., 2017; Paivio, 1969), we predicted that we would find effects of congruency, such that working memory performance would be enhanced when participants were presented with congruent rather than incongruent audiovisual stimuli. We also expected that participants would have lower accuracies and longer reaction times when presented with higher, compared to lower, memory loads. Conversely, we hypothesized the capacity estimates would increase as the memory load increased. While we conducted a behavioral study, we would expect a neuroimaging study using the same paradigm to find robust activation in the superior temporal regions, as found in previous studies using complex multisensory stimuli. Specifically, we would expect this activation to be greater in this study compared to the study using simple stimuli (Chapter 2), as previous studies have found that this region is associated with the processing of naturalistic multisensory stimuli.

General Methods

Participants & Procedure

We recruited undergraduate students for the experiments outlined in this chapter. Participants were compensated with partial course credit. All components of the study were approved by the Institutional Review Board (IRB #2018-0395). The procedure was administered in a quiet room in the Educational Neuroscience Lab space. Experiments were programmed with E-prime 2.0.8.90a (Psychology Software Tools, Sharpsburg, PA) on a Dell Optiplex 390 Desktop PC (3.1 GHz, 4 GB RAM) running Windows 7.0 64-bit operating system. Visual stimuli were presented on a Dell UltraSharp U2212H 21.5” flat-screen monitor at a resolution of 1024 x 768 and a refresh rate of 60 Hz. Auditory stimuli were presented using headphones (Sony MDR-ZX110 Black).

Stimuli

In this study, we used visual and auditory stimuli from animals to study working memory with complex and naturalistic stimuli in comparison to the experiments in Chapter 2. For this, we used naturalistic size-pitch crossmodal correspondences found in animals as our stimulus set (Evans & Treisman, 2010; Gallace & Spence, 2006). Visual stimuli comprised of black-and-white line drawings of the animals, and auditory stimuli were clips of the corresponding sounds of that animal (Figure 3.1). We extracted pictures of animals from the Snodgrass image database, a standardized set of 260 black and white line drawings (Snodgrass & Vanderwart, 1980). Auditory clips of animal sounds were downloaded from Cornell Lab of Ornithology's Macaulay Library, which has a freely available database of over 150,000 animal sounds. Congruent pairs were those in which the animals and sounds matched (e.g. a lion roar paired with the visual of a lion), and incongruent pairs were those in which there is no match (e.g. a lion roar paired with the visual of a House cat). In this way, we were able to investigate the effects of cross-modal correspondences on MSI and working memory using more naturalistic stimuli.

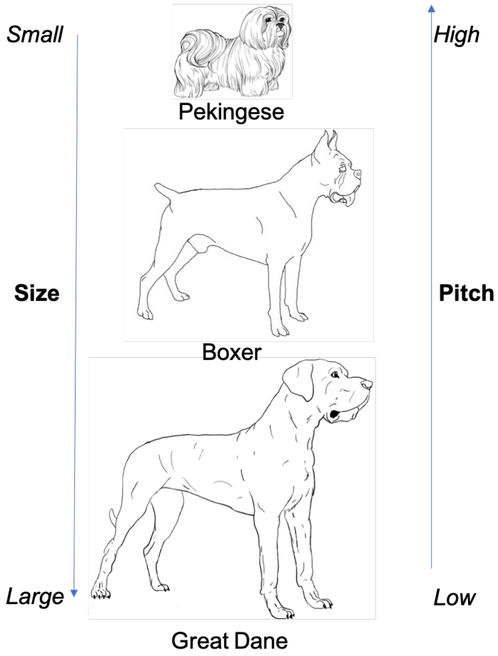


Figure 3.1. An example of the naturalistic size-pitch crossmodal correspondences used for the complex stimulus set for the experiments in this chapter.

To mitigate the possibility of participants using the animal silhouettes to rehearse the information in a verbal format, we used a stimulus set which incorporates images and sounds from different families of animals (Figure 3.2). For example, we extracted images and corresponding sounds of a natal leaf-folding frog, a common river frog, and a giant frog from the family of frogs. Since all three of these animals are frogs with similar features and produce variations of a croak that is not markedly different to the non-expert, the specific differences within these stimuli would be difficult to rehearse verbally.

Experiment 1

Methods

Forty adult participants (mean=19.94 years, M=6, F=34) were recruited for this experiment. All recruited participants had normal or corrected-to-normal vision and hearing and were naïve to the purposes of the experiment. All participants were included in the final data set. In this experiment we aimed to test categories of animals and animal sounds and select the stimuli that we would use to test the impact of complex stimuli on working memory (experiments 2, 3, and 4). We tested eight categories of animals that showed size-pitch cross-modal correspondences: cats, dogs, monkeys, frogs, insects, birds, whales, and pinnipeds. We then selected three animals in a large, medium, and small size to test within these categories (Figure 3.2).

	Cats	Dogs	Monkeys	Frogs	Insects	Birds	Whales	Pinnipeds
Small/ High pitch	House Cat	Pekingese	Pygmy Marmoset	Natal Leaf-folding Frog	Mosquito	Sparrow	Orca	Seal
Med/ Med pitch	Bobcat	Boxer	Barbary Macaques	Common River Frog	House fly	Crow	Humpback Whale	Sea lion
Large/ Low pitch	Lion	Great Dane	Gorilla	Giant Bullfrog	Wasp	Ostrich	Blue Whale	Walrus

Figure 3.2. Animal stimuli used to test categories of animals and animal sounds to select the stimuli that we would use to test the impact of complex stimuli on working memory.

Because the animals that we selected varied on a degree of familiarly, we first presented participants with a list of definitions from the Merriam-Webster, Cambridge, or Oxford Dictionaries (Appendix A). Next, after a fixation cross (500 ms), participants were presented with a sound of each of the animals (2000 ms) in a randomized manner and were instructed to indicate which of the animal categories the stimulus belonged to with a click (Figure 3.3). Next, participants were presented with a within-category or cross-category comparison in a blocked fashion (blocks were counter balanced across participants). Animals from within a category of animals (e.g. House cat vs. lion) were compared in the within-category condition while animals from across two categories of animals were compared in the cross-category condition (e.g. House cat vs. gorilla). In a two-alternative forced choice task, after a fixation cross (500 ms) an auditory sound produced by one of the animals was played (2000 ms). Next, two visual images of animals were presented on the screen and participants were instructed to click on the image of the animal that the sound corresponded to (Figure 3.3). Accuracy was calculated based on the proportion of correct responses on each trial.

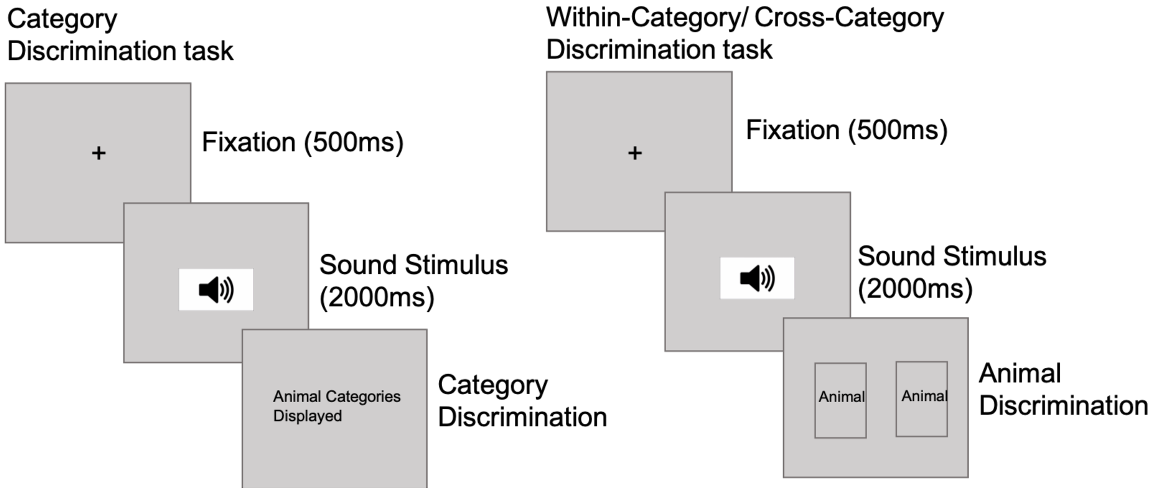


Figure 3.3. Paradigm used to test animal stimuli at a categorical level, within a category, and across a category of animals.

Results

Category Comparison

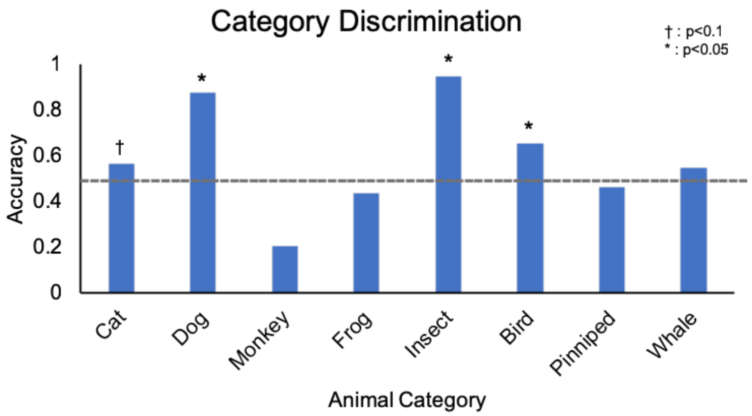


Figure 3.4. Participant accuracy when determining which category of animals a stimulus belonged to. Asterisks indicate significantly higher accuracy compared to chance performance (0.5).

Correct Response	Response									
Categories	Bird	Cat	Dog	Frog	Insect	Monkey	Pinniped	Whale	No Resp	Total	Accuracy
Bird	157		2	8		12	13	47	1	240	0.6542
Cat	7	135	31	11	1	34	19	2		240	0.5625
Dog	3		210		1	24	2			240	0.8750
Frog	37			104	68	1	19	11		240	0.4333
Insect	8				227		3	2		240	0.9458
Monkey	102	17	31	1	13	49	23	3	1	240	0.2042
Pinniped	2	26	59	1		19	111	22		240	0.4625
Whale	31	3		1	5	53	16	131		240	0.5458
No Resp											

Figure 3.5. Analysis of errors to further understand the pattern of response for each of the animal categories. Numbers within the grid indicate the number of trials on which participants chose a particular stimulus as their response.

A visual inspect of figure 3.4 shows that participants were more accurate in determining that the appropriate animal stimulus belonged to some categories compared to others. We found that participants were able to determine that stimuli belonged to the dog (t(39)=15.46, p<0.001) , insect (t(39)=27.47, p<0.001), and bird (t(39)=12.47, p<0.001) categories significantly above chance performance levels (0.5), and the cat (t(39)=1.70, p=0.09) category marginally above chance performance.

We conducted an analysis of errors to further understand the pattern of response for each of the animal categories (Figure 3.5). When looking at performance in the top four animal categories in this analysis, we found that participants had the least confusion when identifying stimuli in the insect category. Out of 240 trials, they chose the insect category for 227 trials (94.5%), the bird category for 8 trials (3.33%), the pinniped category for trials 3 (1.25%), and the whale category for 2 trials (0.83%). Conversely, participants had the most confusion when determining stimuli in the cat category. Out of 240 trials, they chose the cat category for 135 trials (56.25%), the monkey category for 34 trials (14.16%), the dog category for 31 trials (12.91%), the pinniped category for 19 trials (7.91%), the frog category for 11 trials, the bird category for 7 trials (4.58%), and the whale category for 2 trials (0.83%). To choose the final stimulus set, and further understand participant's ability to discriminate between stimuli, it is important to understand discrimination accuracies within animal categories.

Within-Category Discrimination

A visual inspect of figure 3.6 shows that participants had high levels of accuracy when discriminating animals within a category. We found that participants were able to determine that a stimulus was the appropriate animal significantly above chance (0.5) for all animals expect the seal: House cat (t(39)=27.22, p<0.001), bobcat (t(39)=39.00, p<0.001), lion

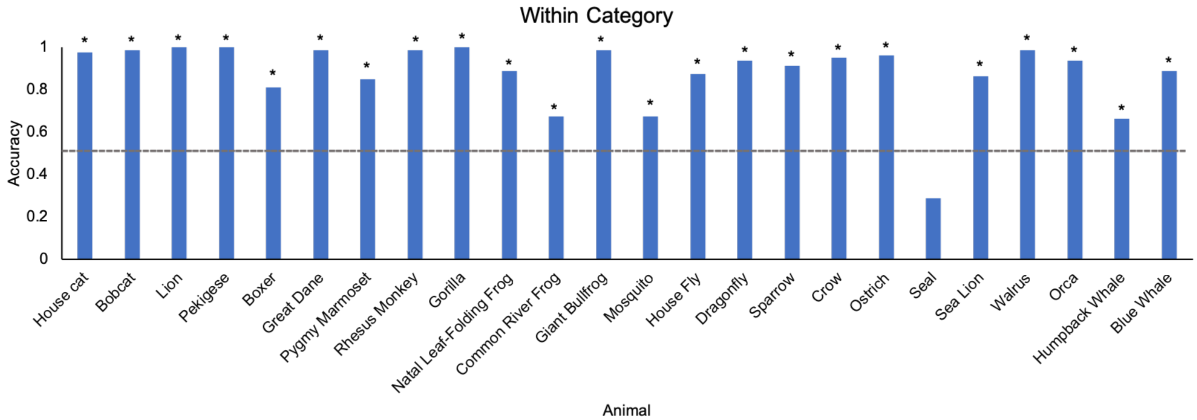


Figure 3.6. Participant accuracy when performing within-category discriminations of animal stimuli. Asterisks indicate significantly higher accuracy compared to chance performance (0.5).

(SD=0, t(39)=na, p<0.001), pekingese (SD=0, t(39)=na, p<0.001), boxer (t(39)=6.75, p<0.001), Great Dane (t(39)=39.00, p<0.001), pygmy marmoset (t(39)=7.85, p<0.001), Rhesus monkey (t(39)=39.00, p<0.001), gorilla (SD=0, t(39)=na, p<0.001), natal leaf-folding frog (t(39)=10.22, p<0.001), common river frog (t(39)=2.88, p<0.01), giant bullfrog (t(39)=39.00, p<0.001), mosquito (t(39)=2.66, p=0.01), House fly (t(39)=8.06, p<0.001), dragonfly (t(39)=13.69, p<0.001), sparrow (t(39)=13.56, p<0.001), crow (t(39)=18.74, p<0.001), ostrich (t(39)=21.93, p<0.001), seal (t(39)=-3.76, p<0.01), sea lion (t(39)=8.28, p<0.001), walrus (t(39)=39.00, p<0.001), orca (t(39)=13.69, p<0.001), humpback whale (t(39)=3.13, p<0.01), blue whale (t(39)=8.49, p<0.001).

Correct Response	Response		
Animal	House Cat	Bobcat	Lion
House Cat	78	2	
Bobcat		79	1
Lion			80

Correct Response	Response		
Animal	Pekingese	Boxer	Great Dane
Pekingese	80		
Boxer		65	15
Great Dane	1		79

Correct Response	Response		
Animal	Pygmy Marmoset	Rhesus Monkey	Gorilla
Pygmy Marmoset	68	12	
Rhesus Monkey		79	1
Gorilla			80

Correct Response	Response		
Animal	Natal Leaf-Folding Frog	Common River Frog	Giant Bullfrog
Natal Leaf-Folding Frog	71	9	
Common River Frog		54	26
Giant Bullfrog	1		79

Correct Response	Response		
Animal	Mosquito	House Fly	Dragonfly
Mosquito	54	26	
House Fly		70	10
Dragonfly	5		75

Correct Response	Response		
Animal	Sparrow	Crow	Ostrich
Sparrow	73	7	
Crow		76	4
Ostrich	3		77

Correct Response	Response		
Animal	Seal	Sea Lion	Walrus
Seal	23	57	
Sea Lion		69	11
Walrus	1		79

Correct Response	Response		
Animal	Orca	Humpback Whale	Blue Whale
Orca	75	5	
Humpback Whale		53	27
Blue Whale	9		71

Figure 3.7. Analysis of errors to further understand the pattern of response for each of the animals within a category. Numbers within the grids indicate the number of trials on which participants chose a particular stimulus as their response.

We conducted an analysis of errors to further understand the pattern of response for each of the animals (Figure 3.7). When looking at performance in the top four animal categories this analysis we found that participants had the most confusion when determining stimuli in the insect category. More specifically, participants chose the correct insect when presented with a mosquito with 67.5% accuracy, and chose the House fly with 32.5% accuracy. Participants chose the correct insect when presented with a housefly with 87.5% accuracy and chose the dragonfly with 12.5% accuracy. Participants chose the correct insect when presented with a dragonfly with

93.75% accuracy and chose the mosquito with 6.25% accuracy. However, participants were highly accurate at making within-category discriminations for cats, the category when they were able to determine only marginally above chance (refer to the category comparison section above). More specifically, participants chose the correct cat when presented with a House cat sound 97.5% of the time and chose the bobcat 2.5% of the time. Participants chose the correct cat when presented with a bobcat sound 98.75% of the time and chose the lion 1.25% of the time. Participants were 100% accurate in choosing the lion image when presented with a lion sound. Therefore, although participants were only marginally significant at determining a cat stimulus at a categorical level, they are able to discriminate between cat stimuli using crossmodal correspondences with high levels of accuracy.

Cross-Category Discrimination

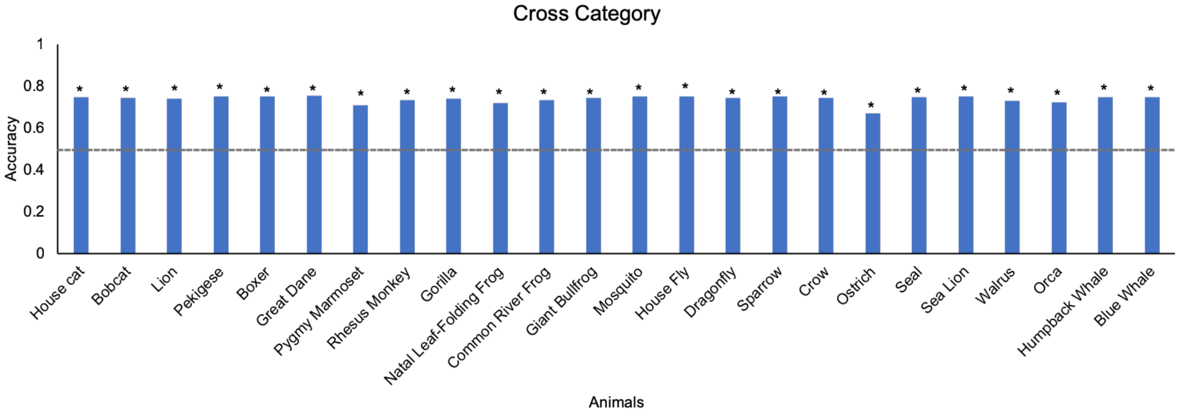
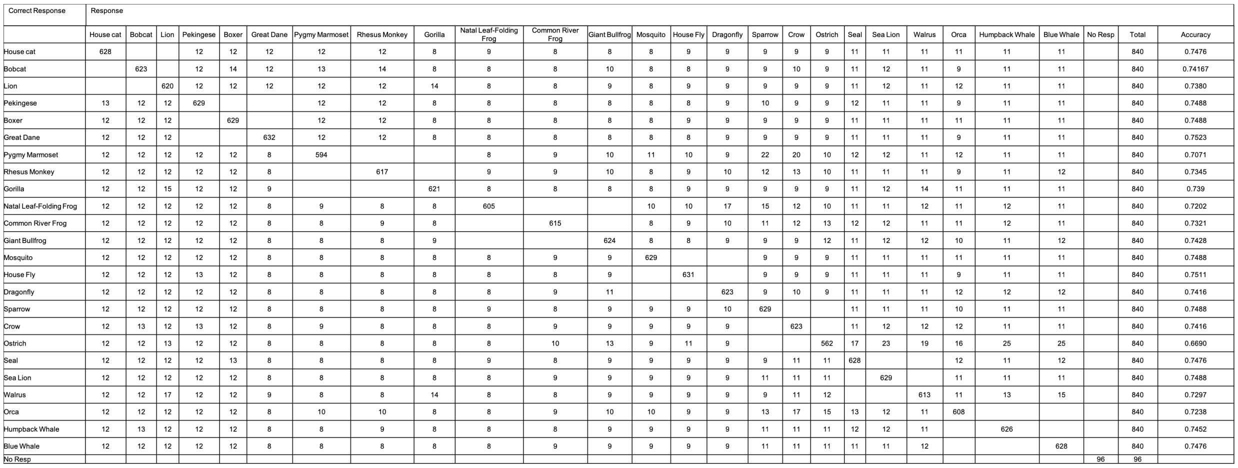


Figure 3.8. Participant accuracy when performing cross-category discriminations of animal stimuli. Asterisks indicate significantly higher accuracy compared to chance performance (0.5).

A visual inspection of Figure 3.8 shows that participants were able to discriminate between animals across categories significantly above chance performance. One sample t-tests showed that participants were able to determine that a stimulus was the appropriate animal significantly above chance (0.5) for all animals: House cat ($t(39)=66.78$, $p<0.001$), bobcat ($t(39)=46.97$, $p<0.001$), Lion ($t(39)=40.94$, $p<0.001$), pekingese ($t(39)=59.67$, $p<0.001$), boxer (t(39)=68.88, p<0.001), Great Dane (t(39)=78.06, p<0.001), pygmy marmoset (t(39)=34.19, p<0.001), Rhesus monkey (t(39)=37.26, p<0.001), gorilla (t(39)=47.33, p<0.001), natal leaf-folding frog (t(39)=25.71, p<0.001), common river frog (t(39)=28.55, p<0.001), giant bullfrog (t(39)=48.86, p<0.001), mosquito (t(39)=68.88, p<0.001), House fly (t(39)=74.40, p<0.001), dragonfly (t(39)=41.82, p<0.001), sparrow (t(39)=63.77, p<0.001), crow (t(39)=53.36, p<0.001), ostrich (t(39)=9.57, p<0.001), seal (t(39)=62.02, p<0.001), sea lion (t(39)=68.88, p<0.001), walrus (t(39)=37.64, p<0.001), orca (t(39)=33.65, p<0.001), humpback whale (t(39)=51.12, p<0.001), blue whale (t(39)=66.78, p<0.001).



Correct Response	Response																									
	House cat	Bobcat	Lion	Pekingese	Boxer	Great Dane	Pygmy Marmoset	Rhesus Monkey	Gorilla	Natal Leaf-Folding Frog	Common River Frog	Giant Bullfrog	Mosquito	House Fly	Dragonfly	Sparrow	Crow	Ostrich	Seal	Sea Lion	Walrus	Orca	Humpback Whale	Blue Whale	No Resp	Total	Accuracy
House cat	628			12	12	12	12	12	8	9	8	8	8	9	9	9	9	9	11	11	11	11	11	11		840	0.7476
Bobcat		623		12	14	12	13	14	8	8	8	10	8	8	9	9	10	9	11	12	11	9	11	11		840	0.74167
Lion			620	12	12	12	12	12	14	8	8	9	8	9	9	9	9	9	11	12	11	12	11	11		840	0.7380
Pekingese	13	12	12	629			12	12	8	8	8	8	8	8	9	10	9	9	12	11	11	9	11	11		840	0.7488
Boxer	12	12	12		629		12	12	8	8	8	8	8	9	9	9	9	9	11	11	11	11	11	11		840	0.7488
Great Dane	12	12	12			632	12	12	8	8	8	8	8	8	9	9	9	9	11	11	11	9	11	11		840	0.7523
Pygmy Marmoset	12	12	12	12	12	8	594			8	9	10	11	10	9	22	20	10	12	12	11	12	11	11		840	0.7071
Rhesus Monkey	12	12	12	12	12	8		617		9	9	10	8	9	10	12	13	10	11	11	11	9	11	12		840	0.7345
Gorilla	12	12	15	12	12	9			621	8	8	8	8	9	9	9	9	9	11	12	14	11	11	11		840	0.739
Natal Leaf-Folding Frog	12	12	12	12	12	8	9	8	8	605			10	10	17	15	12	10	11	11	12	11	12	11		840	0.7202
Common River Frog	12	12	12	12	12	8	8	9	8		615		8	9	10	11	12	13	12	12	11	11	12	11		840	0.7321
Giant Bullfrog	12	12	12	12	12	8	8	8	9			624	8	8	9	9	9	12	11	12	12	10	11	12		840	0.7428
Mosquito	12	12	12	12	12	8	8	8	8	8	9	9	629			9	9	9	11	11	11	11	11	11		840	0.7488
House Fly	12	12	12	13	12	8	8	8	8	8	8	9		631		9	9	9	11	11	11	9	11	11		840	0.7511
Dragonfly	12	12	12	12	12	8	8	8	8	8	9	11			623	9	10	9	11	11	11	12	12	12		840	0.7416
Sparrow	12	12	12	12	12	8	8	8	8	9	8	9	9	9	10	629			11	11	11	10	11	11		840	0.7488
Crow	12	13	12	13	12	8	9	8	8	8	9	9	9	9	9		623		11	12	12	12	11	11		840	0.7416
Ostrich	12	12	13	12	12	8	8	8	8	8	10	13	9	11	9			562	17	23	19	16	25	25		840	0.6690
Seal	12	12	12	12	13	8	8	8	8	9	8	9	9	9	9	9	11	11	628			12	11	12		840	0.7476
Sea Lion	12	12	12	12	12	8	8	8	8	8	9	9	9	9	9	11	11	11		629		11	11	11		840	0.7488
Walrus	12	12	17	12	12	9	8	8	14	8	8	9	9	9	9	9	11	12			613	11	13	15		840	0.7297
Orca	12	12	12	12	12	8	10	10	8	8	9	10	10	9	9	13	17	15	13	12	11	608				840	0.7238
Humpback Whale	12	13	12	12	12	8	8	9	8	8	8	9	9	9	9	11	11	11	12	12	11		626			840	0.7452
Blue Whale	12	12	12	12	12	8	8	8	8	8	9	9	9	9	9	11	11	11	11	11	12			628		840	0.7476
No Resp																									96	96	

Figure 3.9. Analysis of errors to further understand the pattern of response for each of the animals across a category. Numbers within the grid indicate the number of trials on which participants chose a particular stimulus as their response.

We conducted an analysis of errors to further understand the pattern of response for each of the animals (Figure 3.9). When looking at performance for animals in the four animal categories with the highest discriminability at the categorical level, we found that participants had the most confusion when identifying the ostrich stimulus, even though they were still significantly above chance at doing so. While participants were able to identify the ostrich stimulus with 66.90% accuracy, they were most likely to confuse it with the pinniped (seal:

2.02%; sea lion: 2.73%; walrus: 2.26%) or whale (orca: 1.90%; humpback whale: 2.97%; blue whale: 2.97%) categories. Participants had similar levels of accuracy when discriminating cats (House cat: 74.76% ; bobcat: 74.16% ; lion: 73.80%), dogs (pekingese: 74.88% ; boxer: 74.88% ; Great Dane: 75.23%) and insects (mosquito: 74.88%; House Fly: 75.11%; dragonfly:74.16%) from animals in other categories.

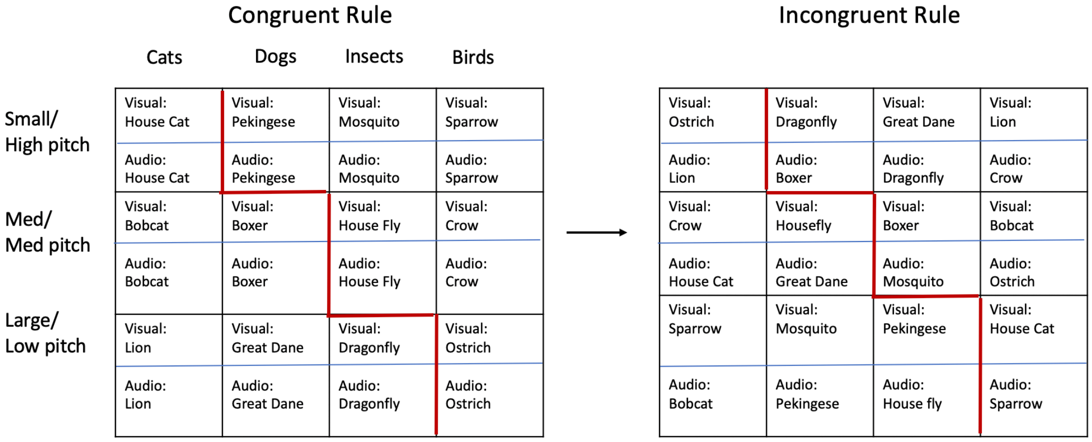
Brief Discussion

In this experiment, we tested the participants' familiarity and ability to discriminate between animal stimuli with the goal of selecting the final stimulus set to test on a working memory task. To do this, we tested three animals within the category of cats, dogs, monkeys, frogs, insects, birds, whales, and pinnipeds. In the categorical discrimination task, participants were able to identify animals in the dog, insect, and bird categories significantly above chance performance, and animals categorized as cats marginally above chance performance. However, in the within-category discrimination task, participants were able to discriminate between the three the cats (House cat, bobcat, lion) significantly above chance, and with high levels of accuracy. For this reason, we decided to the choose animals in the cat (House cat, bobcat, lion), dog (pekingese, boxer, Great Dane), insect (mosquito, House fly, dragonfly), and bird (sparrow, crow, ostrich) categories for our final stimulus set. We have used this stimulus set in Experiments 2, 3, and 4 to test the impact of complex multisensory stimuli on working memory.

Experiment 2

Methods

Forty-two adult participants were recruited for this experiment. All recruited participants had normal or corrected-to-normal vision and hearing and were naïve to the purposes of the experiment. Thirty-five participants (mean=20.31, M=5, F=30) were included in the final data set after the data cleaning analyses were conducted. For the experiment, each of the four animal categories consisted of three animals which were small, medium, or large in size and produced a high, medium, or low pitch sound respectively, resulting in a total of 12 stimuli (Figure 3.10).



Congruent Rule

	Cats	Dogs	Insects	Birds
Small/ High pitch	Visual: House Cat	Visual: Pekingese	Visual: Mosquito	Visual: Sparrow
	Audio: House Cat	Audio: Pekingese	Audio: Mosquito	Audio: Sparrow
Med/ Med pitch	Visual: Bobcat	Visual: Boxer	Visual: House Fly	Visual: Crow
	Audio: Bobcat	Audio: Boxer	Audio: House Fly	Audio: Crow
Large/ Low pitch	Visual: Lion	Visual: Great Dane	Visual: Dragonfly	Visual: Ostrich
	Audio: Lion	Audio: Great Dane	Audio: Dragonfly	Audio: Ostrich

Incongruent Rule

Visual: Ostrich	Visual: Dragonfly	Visual: Great Dane	Visual: Lion
Audio: Lion	Audio: Boxer	Audio: Dragonfly	Audio: Crow
Visual: Crow	Visual: Housefly	Visual: Boxer	Visual: Bobcat
Audio: House Cat	Audio: Great Dane	Audio: Mosquito	Audio: Ostrich
Visual: Sparrow	Visual: Mosquito	Visual: Pekingese	Visual: House Cat
Audio: Bobcat	Audio: Pekingese	Audio: House fly	Audio: Sparrow

Figure 3.10. Final set of stimuli used to test the impact of complex stimuli on working memory. The stimuli were presented as per the crossmodal correspondences in the congruent rule. The colors were flipped across the diagonal line and sounds were rotated within the columns to form the schematic to form the incongruent rule.

Congruent pairs were those in which cross-sensory correspondences were maintained (e.g. a lion roar paired with the image of a lion), and incongruent trials were those in which these crossmodal correspondences were not maintained (e.g. a lion roar paired with the visual of an ostrich). Analogous to the procedure used with the simple stimuli, animal images were flipped across the diagonal line and sounds were rotated within the columns to create the schematic that formed the incongruent rule. In this way, we were able to systematically vary the distance of each stimulus from its position in the congruent grid.

Participants were presented with a different number of to-be-remembered items (i.e. memory load) in each trial. In this experiment, participants were presented with a memory load of two, four, or eight items. In a given trial, each stimulus was presented for 750 ms with an interval of 250 ms between stimuli. After a delay period (4000 ms), participants were presented with either the same pattern or a different pattern of stimuli, and participants were asked to indicate whether they thought the test pattern contained a change during the response period (2000 ms). Only one item out of the string of stimuli changed in the change trials, and none of the stimuli changed in the no-change trials (Figure 3.11).

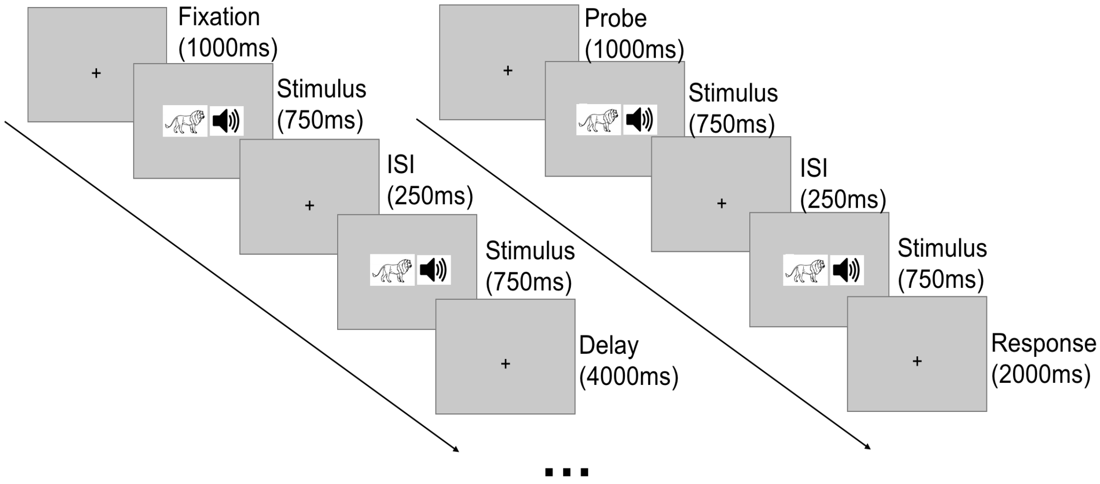


Figure 3.11. The multisensory working memory paradigm used with a set of complex stimuli. Participants will be presented with a randomly selected array of two, four, or eight animal sounds or animal pictures. After a delay, they will be presented with either the same pattern or a different pattern, and will be asked to detect a change in the pattern presented during the response duration.

In the unisensory auditory (A) and visual (V) baseline conditions, participants were presented with and tested on stimuli only from the given modality. In the multisensory (audio-visual: AV) condition, participants were presented with auditory and visual stimuli in a temporally and spatially synchronized manner, but were only tested on stimuli from one modality in each block (Conditions: AV-Test A or AV-Test V). Participants were cued to the condition (A, V, or AV) in a blocked fashion, which was counterbalanced across participants.

In the multisensory conditions, on half of the trials, participants were presented with congruent stimuli in which cross-sensory correspondences were maintained (e.g. a lion roar paired with the visual of a lion), and incongruent stimuli in which crossmodal correspondences were not maintained (e.g. a lion roar paired with the visual of an ostrich) on half of the trials. All the stimuli presented had congruent pairings in a congruent trial, and incongruent pairings in an incongruent trial. To form the incongruent grid, we systematically varied the distance of each stimulus from its position in the congruent grid by exchanging the positions of the images across the diagonal line and rotating the pitch of the tones for a given animal.

Analysis

Accuracy, capacity estimates, and reaction times were calculated, as together they provide a deeper understanding of the working memory mechanisms at play. While performance accuracy tends to decrease as the mnemonic load increases, individual differences in working memory capacity have also been noted (Linderholm et al., 2008; Luck & Vogel, 2013). Accuracy was calculated based on the proportion of correct responses on each trial. All participants with accuracies below 50% (chance performance) at lowest load in baseline A or V conditions were removed from the final analyzed data set. Reaction times were calculated as the time it took participants to respond (in milliseconds) from the offset of the last stimulus in the memory array.

Working memory capacity was calculated using Cowan's k (Cowan, 2001, 2010; Rouder et al., 2011). Under the assumption that participants have a working memory capacity that is fully employed on each trial, this formula to calculate working memory capacity (k) is detailed in the following manner: The probability that the participant makes a hit is $H= k/N + g((N-k)/N)$, where k = working memory capacity, N=total number of stimuli presented, and g is the probability of guessing correctly (based on the number of choices). That is, the probability of a hit is the sum of the probability that the participant detects the change (k/N) plus the probability that the participant guesses correctly, even when they do not detect the change ($g((N-k)/N)$). Based on this formula, the probability that the participant makes a correct rejection is $CR= k/N + (1-g)((N-k)/N)$. Combining these two formulas and solving for k, we can obtain the formula to calculate working memory capacity: $k=N(H+CR-1)$. Working memory capacity was calculated in the same manner here. Statistical results for all the pairwise comparisons of working memory performance (accuracy, reaction time, capacity) across conditions conducted for experiments in this chapter can be found in Appendix C.

Results

Effects of Modality

Accuracy

As can be seen in Figure 3.12A, accuracy of participants decreased as the memory load increased. Additionally, we found evidence of a multisensory advantage in when tested on auditory working memory, but not when tested on visual working memory. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,34)=166.51 , p<0.001, η_p^2=0.83), and condition (F(1,34)=5.42 , p<0.01, η_p^2=0.14), and a significant interaction of load x condition (F(1,34)= 3.05, p<0.01, η_p^2=0.08). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. We found significant

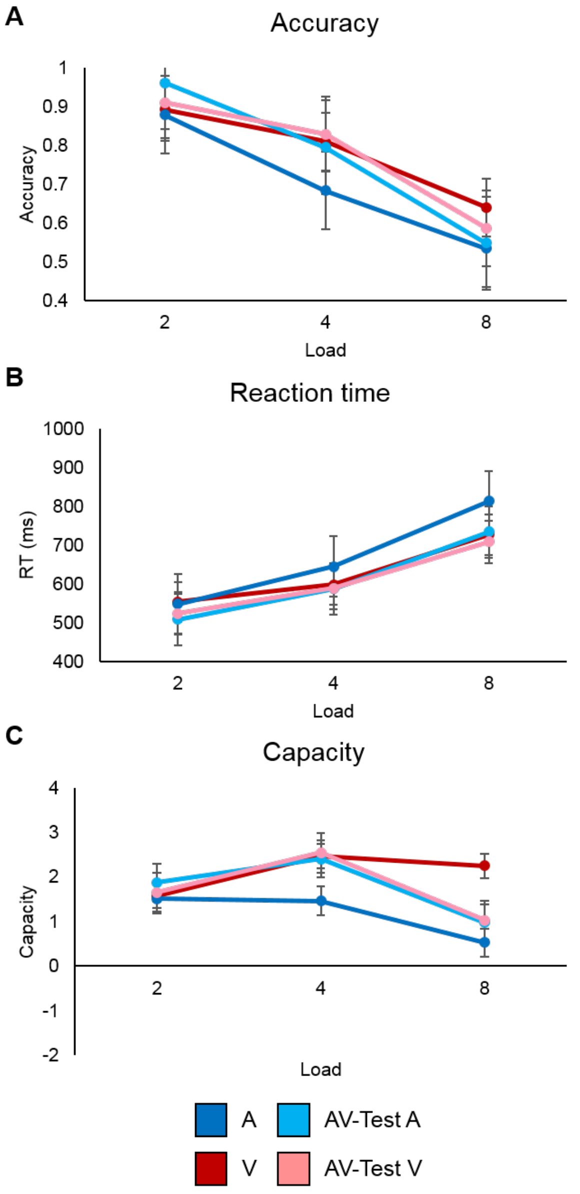


Figure 3.12. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants in the unisensory (V, A) or multisensory (AV-test A, AV-test V) conditions. Error bars represent the standard error of the mean.

differences between accuracies when multisensory stimuli (AV- Test A or AV- Test V) were compared to unisensory stimuli (A or V) in auditory (condition F(1,34)=11.76, p<0.01, η_p^2=0.26, load F(1,34)=95.16, p<0.001, η_p^2=0.73, load x condition interaction F(1,34)=2.46, p=0.09, η_p^2=0.07), but not visual (condition F(1,34)=0.02, p=0.88, η_p^2=0.001, load F(1,34)=75.21, p<0.001, η_p^2=0.69, load x condition interaction F(1,34)=1.68, p=0.19, η_p^2=0.47) working memory. Specifically, we found a significant difference in accuracy when the auditory alone and AV-Test A conditions were compared only at a load of 2 (t(34)=3.18, p<0.01), and 4 (t(34)=2.91, p<0.01) items. When comparing accuracy for unisensory auditory and visual conditions, we found a significant main effect of condition (F(1,34)=17.73, p<0.001, η_p^2=0.34), load (F(1,34)=103.51, p<0.001, η_p^2=0.75), and a significant load x condition interaction (F(1,34)=3.37, p=0.04, η_p^2=0.09). Accuracy for the visual alone was significantly higher than in the auditory alone condition at loads of 4 (t(34)=4.07, p<0.001), and 8 (t(34)=2.82, p<0.01) items.

Reaction Time

As seen in Figure 3.12B, we found higher reaction times for unisensory compared to multisensory stimuli auditory working memory, but did not find a difference in the visual working memory domain. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,34)=43.39 , p<0.001, η_p^2=0.56), but not a significant effect of condition (F(1,34)= 1.81, p=0.15, η_p^2=0.05), nor a significant interaction of load x condition (F(1,34)= 0.62, p=0.72, η_p^2=0.02).

Capacity

Overall, we found that estimated capacity for participants increased as the memory load increased but began to plateau or decline (depending on the condition) after a load of 4 items

(Figure 3.12C). The largest capacity estimate on a given trial is dictated by the number of stimuli presented on that trial. Thus, the possible capacity on a trial with a load of eight items is larger than on a trial with a load of two items. For this reason, the capacity estimate, k, is modulated by the multiplier, *N*, which accounts for the number of stimuli presented. Furthermore, visual inspection of Figure 3.12C reveals a multisensory advantage when tested in the auditory sensory domain but not the visual sensory domain. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,34)=8.24 , p<-.01, η_p^2=0.19), and condition (F(1,34)=3.54 , p=0.02, η_p^2=0.09), but not a significant interaction of load x condition (F(1,34)=1.94, p=0.08, η_p^2=0.05). Interestingly, we found enhanced visual working memory capacity estimates when presented with incongruent simuli at all three memory loads. In contrast, we found enhanced auditory working memory capacity estimates when presented with congruent simuli. To directly contrast performance between the sensory modalities tested, we ran a series of 2 (condition) x 3 (load) ANOVAs. When comparing capacity estimates for unisensory auditory and visual conditions, we found a significant main effect of condition (F(1,34)=21.53, p<0.001, η_p^2=0.39), but not a significant main effect of load (F(1,34)=2.23, p=0.12, η_p^2=0.06), and a significant load x condition interaction (F(1,34)=5.2, p<0.01, η_p^2=0.13). Capacity estimates were higher in the unisensory visual compared to unisensory auditory condition are compared at a load of 4 (t(34)=4.37, p<0.001), and 8 (t(34)=3.04, p<0.01) items. We did find evidence of an auditory multisensory enhancement, as unisensory auditory and AV-Test A were significantly different (condition F(1,33)=6.94, p=0.01, η_p^2=0.17, load F(1,33)=4.99, p=0.01, η_p^2=0.13, load x condition interaction F(1,33)=0.72, p=0.49, η_p^2=0.02). AV test-A capacity estimates were significantly higher than auditory alone (A) at loads of 2 (t(34)=3.36, p<0.01), and 4 (t(34)=3.52, p=0.001) items. Lastly, we did not find significant differences when comparing the unisensory visual and AV-Test V conditions. A 2 (condition) x 3 (load) ANOVA revealed a significant main effect of load (F(1,34)=5.4, p<0.01, η_p^2=0.14), but not of condition

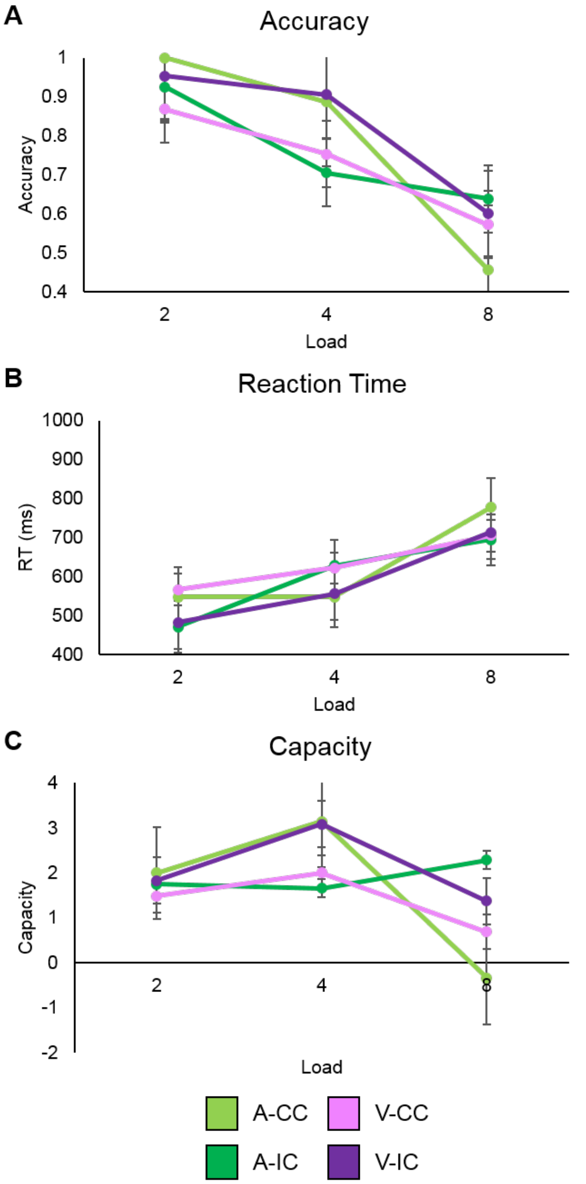


Figure 3.13. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants when presented with congruent and incongruent stimuli and tested on the auditory (A-CC & A-IC) and visual (V-CC & V-IC) working memory. Error bars represent the standard error of the mean.

(F(1,34)=1.21, p=0.28, η_p^2=0.03), and a marginal load x condition interaction (F(1,34)=2.7, p=0.07, η_p^2=0.07).

Effects of Congruency

Accuracy

Figure 3.13A shows results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). Participants performed with higher accuracy, especially at lower loads, when presented with congruent stimuli and tested on auditory working memory. However, they performed with higher accuracy when presented with incongruent stimuli in the visual working memory domain. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,34)=101.39 , p<0.001, η_p^2=0.75), but not a significant effect of condition (F(1,34)=1.95, p=0.13, η_p^2=0.05), and a significant interaction of load x condition (F(1,34)=5.13 , p<0.001, η_p^2=0.14).

Reaction Time

Visual inspection of Figure 3.13B does not show differences in reaction times when presented with congruent versus incongruent stimuli tested in either visual or auditory working memory. This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a main effect of load (F(1,34)=17.75, p<0.001, η_p^2=0.343), but not a significant effect of condition (F(1,34)=1.02, p=0.39, η_p^2=0.03), nor a significant interaction of load x condition (F(1,34)=1.79, p=0.11, η_p^2=0.05).

Capacity

Figure 3.13C shows results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). Overall, we found that participants had higher capacity estimates when presented with incongruent stimuli and tested on visual working memory. On the contrary, we found that participants had higher capacity when presented with congruent stimuli and tested on auditory working memory, but only at the lower memory loads. A 3 (load) x 4 (condition) ANOVA revealed a main effect of load ($F(1,34)=9.77$, $p<0.001$, $\eta_p^2=0.22$), not a significant effect of condition ($F(1,34)=0.69$, $p=0.56$, $\eta_p^2=0.02$), but revealed a significant interaction of load x condition ($F(1,34)=2.65$, $p=0.02$, $\eta_p^2=0.07$).

Brief Discussion

We tested the impact of these complex stimuli on working memory using a change detection task by presenting participants with a memory load of two, four, or eight items in Experiment 2. Overall, we found that participants displayed greater multisensory enhancement when tested on auditory working memory compared to visual working memory. In fact, we did not find significant modality-specific differences in visual working memory performance when presented with unisensory compared to multisensory stimuli for accuracies, reaction times, or capacity estimates. Participants showed benefits in auditory working memory when presented with congruent multisensory stimuli, but only at loads of two and four items. Contrary to our predictions, we found that participants had enhanced working memory performance when presented with incongruent multisensory stimuli in the visual working memory domain.

While these effects of congruency are contrary to the previous literature, it is important to note that enhancements for congruently presented multisensory stimuli have mostly been found in the perceptual domain (Kim, Seitz, & Shams, 2008; Mishra & Gazzaley, 2012). For this reason it could be the case that congruent stimuli do not produce the same enhancements in the cognitive, namely memory, domain (Brunel, Carvalho, & Goldstone, 2015). To better understand whether working memory advantages for multisensory stimuli, as well as whether effects of congruency were influenced by the memory load, we presented participants with a larger array of complex stimuli and higher memory load in Experiment 3.

Experiment 3

Methods

Forty adult participants were recruited for this experiment. All recruited participants had normal or corrected-to-normal vision and hearing and were naïve to the purposes of the experiment. Thirty-two participants (mean=20.22, M=6, F=26) were included in the final data set after the data cleaning analyses were conducted. Participants were presented with a randomly selected array of four, six, or eight sounds or color patches. The procedure for this study followed the experimental design and analysis as described in Experiment 2, and participants were cued on the modality (auditory or visual) that they were going to be tested on in the multisensory condition (AV-Test A or AV-Test V) (Figure 3.9). Due to the increased difficulty of the task with the increase in memory load, participants were presented with practice trials at the beginning of the experiment which were excluded from the final analyses. Participants were presented with a set of congruent stimuli on half the trials, and incongruent stimuli on the other half of the trials in a randomized manner.

Results

<u>Effects of Modality</u>

Accuracy

We found evidence of a multisensory advantage when tested on auditory working memory, most particularly at the highest loads, but not when tested on visual working memory (Figure 3.14A). A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,31)=42.77, $p<0.001$, η_p^2=0.58), condition (F(1,31)=10.20, $p<0.001$, η_p^2=0.25), and a significant interaction of load x condition (F(1,31)=4.08, $p=0.01$, η_p^2=0.12). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. When comparing accuracy for unisensory auditory and visual conditions, we found a significant main effect of condition (F(1,31)=34.55, $p<0.001$, η_p^2=0.53), load (F(1,31)=29.05, $p<0.001$, η_p^2=0.48), and a marginally significant load x condition

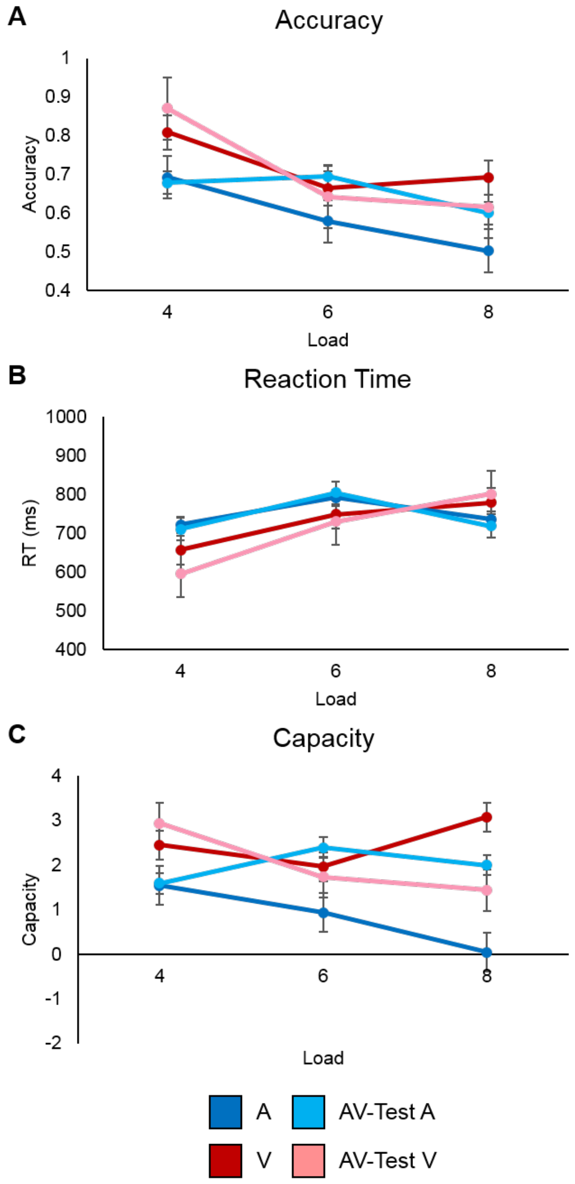


Figure 3.14. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants in the unisensory (V, A) or multisensory (AV-test A, AV-test V) conditions. Error bars represent the standard error of the mean.

interaction (F(1,31)=2.85, p=0.07, η_p^2=0.08). Specifically, we found that accuracy was higher in the visual alone condition compared to the auditory alone condition at loads of 4 (t(31)=3.88, p<0.01), 6 (t(31)=2.27, p=0.03), and 8 (t(31)=5.61, p<0.001) items. We also found evidence of an auditory multisensory enhancement, as unisensory auditory and AV-Test A were significantly different (condition F(1,34)=6.76, p=0.01, η_p^2=0.18, load F(1,34)=12.08, p<0.001, η_p^2=0.28, load x condition interaction F(1,34)=2.86, p=0.07, η_p^2=0.09). Accuracy was higher in the AV-Test A compared to the unisensory auditory condition when compared at loads of 6 (t(31)=2.84, p<0.01), and 8 (t(31)=2.10, p=0.04) items. Lastly, we did not find evidence of an visual multisensory enhancement, as unisensory visual and AV-Test V were not significantly different (condition F(1,38)=0.22, p=0.64, η_p^2<0.01, load F(1,38)=34.99, p<0.001, η_p^2=0.53, load x condition interaction F(1,38)=3.11, p=0.05, η_p^2=0.09).

Reaction Time

Figure 3.14B shows that the reaction times of participants were significantly different at different loads but did not vary significantly by condition of the stimuli presented. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,31)=21.04, p<0.001, η_p^2=0.40), but not a significant effect of condition (F(1,31)=6.90, p=0.45, η_p^2=0.03), and revealed a significant interaction of load x condition (F(1,31)=3.96, p<0.01, η_p^2=0.11).

Capacity

Visual inspection of Figure 3.14C shows a multisensory advantage when tested on the auditory working memory but not the visual working memory. Furthermore, this multisensory advantage in auditory working memory became more robust as the load increased. A 3 (load) x 4 (condition) ANOVA did not reveal a significant main effect of load (F(1,31)=2.10, p=0.14, η_p^2=0.06), and but did reveal a significant effect of condition (F(1,31)=8.39, p<0.001, η_p^2=0.21), and a significant interaction of load x condition (F(1,31)=3.27, p<0.01, η_p^2=0.10). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. We did not find evidence of a visual multisensory enhancement, as unisensory visual and AV-Test V were not significantly different (condition F(1,31)=1.51, p=0.23, η_p^2=0.05, load F(1,31)=2.95, p=0.06, η_p^2=0.09, load x condition interaction F(1,31)=3.52, p=0.04, η_p^2=0.10). However, we did find evidence of an auditory multisensory enhancement, as unisensory auditory and AV-Test A were significantly different (condition F(1,31)=13.80, p=0.001, η_p^2=0.31, load F(1,31)=1.95, p=1.15, η_p^2=0.06, load x condition interaction F(1,31)=3.36, p=0.04, η_p^2=0.10). Specifically, we found that capacity estimates were higher in the AV-Test A condition compared to the unisensory auditory condition at loads of 6 (t(31)=2.92, p<0.01), and 8 (t(31)=2.76, p=0.01) items. When comparing accuracy for unisensory auditory and visual conditions, we found a significant main effect of condition (F(1,31)=34.28, p<0.001, η_p^2=0.53), but not a significant main effect of load (F(1,31)=2.18, p=0.12, η_p^2=0.07), and a significant load x condition interaction (F(1,31)=8.99, p<0.001, η_p^2=0.23). Capacity estimates were significantly higher for the unisensory visual compared to the unisensory auditory condition at loads of 4 (t(31)=3.81, p<0.01), 6 (t(31)=2.27, p=0.03), and 8 (t(31)=5.61, p<0.001) items.

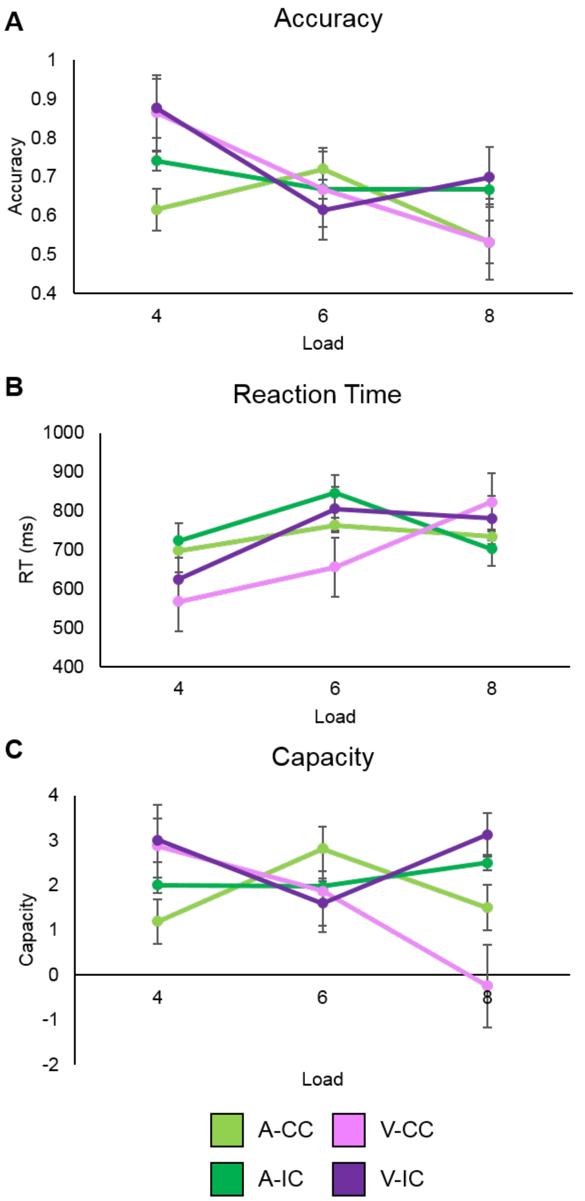


Figure 3.15. Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants when presented with congruent and incongruent stimuli and tested on the auditory (A-CC & A-IC) and visual (V-CC & V-IC) working memory. Error bars represent the standard error of the mean.

Effects of Congruency

Accuracy

Figure 3.15A shows results from trials in which participants were presented with congruent (CC) or incongruent (CC) stimuli but were only tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a main effect of load (F(1,31)=14.91, p<0.001, η_p^2=0.33), condition (F(1,31)=2.79, p=0.04, η_p^2=0.08), and a significant interaction of load x condition (F(1,31)=4.32, p<0.001, η_p^2=0.12). A 2 (condition) x 3 (load) ANOVA showed significant differences when comparing congruent and incongruent stimuli in the auditory domain by revealing a marginally significant main effect condition (F(1,31)=3.98, p=0.06, η_p^2=0.11), but not a significant main effect of load (F(1,31)=2.31, p=0.11, η_p^2=0.07), and a marginally significant load x condition interaction (F(1,31)=2.63, p=0.08, η_p^2=0.08). A 2 (condition) x 3 (load) ANOVA in the visual domain revealed a significant effect of load (F(1,31)=19.80, p<0.001, η_p^2=0.39), a significant load x condition interaction (F(1,31)=3.10, p=0.05, η_p^2=0.09), but not a significant main effect of condition (F(1,31)=1.35, p=0.25, η_p^2=0.04). The presentation of incongruent stimuli significantly enhanced auditory working memory at the loads of 4 (t(31)=2.17, p) and 8 (t(31)=2.03, p=0.05) items, but only enhanced visual working memory performance at the highest load of 8 (t(31)=-2.26, p=0.03) items.

Reaction Time

Visual inspection of Figure 3.15B shows that reaction times were longer when participants were presented with incongruent stimuli, particularly for visual working memory. This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (CC) stimuli but were only tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,31)=11.48, p<0.001, η_p^2=0.27), but not a significant effect of condition (F(1,31)=2.23, p=0.90, η_p^2=0.07), and a significant interaction of load x condition (F(1,31)=3.52, p<0.01, η_p^2=0.10).

Capacity

Figure 3.15C shows results from trials in which participants were presented with congruent (CC) or incongruent (CC) stimuli but were only tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). Overall, we found increased capacity estimates when presented with incongruent stimuli in visual and auditory working memory. A 3 (load) x 4 (condition) ANOVA did not reveal a significant effect of load (F(1,31)=0.80, p0.45, η_p^2=0.03), or condition (F(1,31)=1.77, p=0.17, η_p^2=0.05), but revealed a significant interaction of load x condition (F(1,31)=3.10, p<0.01, η_p^2=0.09).

Brief Discussion

To better understand whether working memory advantages for multisensory stimuli were affected by the memory load, in Experiment 3 we presented participants with four, six, or eight auditory, visual, or audiovisual stimuli. All other aspects of the experimental paradigm remained identical to those that were used in Experiment 2. Similar to the results in Experiment 2, we found that participants displayed greater multisensory enhancement when tested on auditory working memory compared to visual working memory. This advantage for multisensory stimuli in auditory working memory grew as the memory load increased. We did not find significant differences in visual working memory performance when presented with unisensory and multisensory stimuli. We found that the presentation of incongruent stimuli led to benefits in visual and auditory working memory at the highest load of eight items. To further understand the influence of congruency for crossmodal stimuli, we presented participants with stimuli that varied in degrees of congruency in Experiment 4.

Experiment 4

Methods

Thirty-three adult participants (mean=20.31, M=7, F=26) were recruited for this experiment. All recruited participants had normal or corrected-to-normal vision and hearing and were naïve to the purposes of the experiment. All participants were included in the final data set. Participants were presented with a randomly selected array of four, six, or eight animal sounds and animal line drawings. The procedure for this study followed the experimental design and analysis of Experiments 2 and 3 except that, participants were only presented with multisensory audiovisual stimuli and were cued on the modality (auditory or visual) that they were going to be tested on (AV-Test A or AV-Test V) (Figure 3.10). Participants were presented with a set of congruent stimuli (distance 0 ; e.g. lion visual image presented with a lion sound), stimuli that had within category incongruency (distance 1; e.g. lion visual image presented with a House cat sound) and stimuli that had cross-category incongruencies (distance 2; e.g. lion visual image presented with a pekingese sound) in a randomized manner. In this way, we were able to systematically test the effects of stimulus congruency on working memory performance.

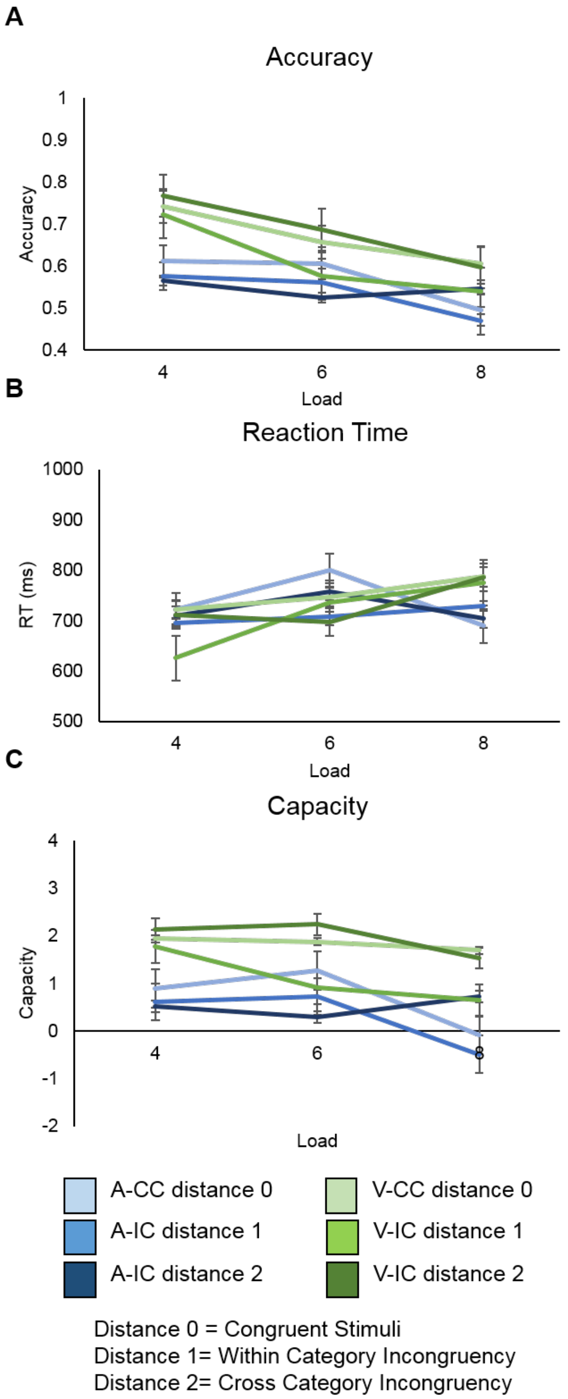


Figure 3.16. Working memory accuracies (A) reaction times (B) and capacity estimates (C) when participants were tested with congruent (distance 0), within-category incongruent (distance 1), and cross-category incongruent (distance 2) stimuli in the multisensory, AV, condition. Error bars represent the standard error of the mean.

Results

Accuracy

As can be seen in Figure 3.16A, accuracy of participants decreased as the memory load increased across all conditions. Furthermore, we found that performance was higher for visual working memory compared to auditory working memory at lower loads, but that this difference diminished at higher loads. To directly contrast performance when presented with congruent, within-category incongruent, and cross-category incongruent, we ran 3 (congruency/distance) x 3(load) ANOVAs when tested on auditory and visual working memory. We did not find evidence of an effect of congruency in the auditory domain (load F(1,32)=5.03, $p<0.001$, η_p^2=0.14), distance (F(1,32)=0.80, p=0.46, η_p^2=0.02), load x distance interaction (F(1,32)=0.99, p=0.42, η_p^2=0.03). In the visual domain, we found a significant main effect of load (F(1,32)=17.76, $p<0.001$, η_p^2=0.36), but not a significant effect of distance (F(1,32)=2.40, p=0.10, η_p^2=0.07), nor a significant interaction of load and distance (F(1,32)=0.54, p=0.71, η_p^2=0.02). Accuracy was significantly higher when presented with incongruent stimuli with a distance of 2 (cross-category incongruency) compared to incongruent stimuli with a distance of 1 (within-category incongruency) at a load of 6 (t(32)=2.46, p=0.02) items.

Reaction Time

Reaction times of participants did not differ by memory load or congruency distance in the auditory modality, but did differ by the memory load in the visual modality (Figure 3.16B). To directly contrast performance when presented with congruent, within-category incongruent, and cross-category incongruent, we ran 3 (congruency/distance) x 3(load) ANOVAs when tested on auditory and visual working memory. This analysis in the auditory modality did not reveal a significant main effect of load (F(1,32)=1.94, p=0.15, η_p^2=0.06), distance (F(1,32)=0.32, p=0.73, η_p^2=0.01), nor a significant interaction of load and distance (F(1,32)=1.21, p=0.31, η_p^2=0.04). A 3 (congruency/distance) x 3(load) ANOVA in the visual modality revealed a significant main effect of load (F(1,32)=5.43, $p<0.01$, η_p^2=0.15), but not a significant effect of distance (F(1,32)=0.64, p=0.53, η_p^2=0.02), nor a significant interaction of load and distance (F(1,32)=1.44, p=0.23, η_p^2=0.04). We did not find significant statistical differences at loads of four, six, or eight items when reaction times were compared across different distances in auditory or visual working memory.

Capacity

As can be seen in Figure 3.16C, overall, capacity estimates were highest for incongruent stimuli when tested on visual working memory, but were highest for congruent stimuli when tested on auditory working memory. To directly contrast performance when presented with congruent, within-category incongruent, and cross-category incongruent, we ran 3 (congruency/distance) x 3(load) ANOVAs when tested on auditory and visual working memory. We did not find a significant main effect of load (F(1,32)=2.83, p=0.07, η_p^2=0.08), distance (F(1,32)=0.83, p=0.44, η_p^2=0.03), nor a significant interaction of load and distance (F(1,32)=1.92, p=0.11, η_p^2=0.06) in the auditory domain. A 3 (congruency/distance) x 3(load) ANOVA in the visual modality did not reveal a significant main effect of load (F(1,32)=2.08, p=0.13, η_p^2=0.06), distance (F(1,32)=2.72, p=0.07, η_p^2=0.08), nor a significant interaction of load and distance (F(1,32)=0.69, p=0.60, η_p^2=0.02). We did not find significant statistical differences at loads of four, six, or eight items when reaction times were compared across different distances in auditory or visual working memory.

Brief Discussion

Our findings in the explicit congruency manipulation in Experiment 4 match with the existing literature in the domain of auditory working memory as we found that participants had the highest memory performance when presented with congruent stimuli. On the contrary, we found that participants have the lowest memory performance when presented with congruent stimuli and tested on visual working memory. This result indicates that congruent auditory information is not helping visual working memory, while congruent visual information is indeed helping auditory working memory. We also found that participants overall had higher working memory performance for stimuli in the visual rather than auditory working memory domain.

General Discussion

The aim of the experiments in this chapter was to test the impact of complex unisensory and multisensory stimuli on working memory. In Experiment 1, we tested participants' discrimination abilities on a variety of animal stimuli to choose the final stimulus set that we would use to test the impact of complex stimuli on working memory in Experiments 2, 3, and 4. Based on results from Experiment 1, we chose animals in the cat (House cat, bobcat, lion), dog (pekingese, boxer, Great Dane), insect (mosquito, House fly, dragonfly), and bird (sparrow, crow, ostrich) categories for our final stimulus set. In Experiment 2, we tested the impact of these complex stimuli on working memory using a change detection paradigm by presenting participants with a memory load of two, four, or eight items. To better understand whether working memory advantages for multisensory stimuli were affected by the memory load, we presented participants with four, six, or eight auditory, visual, or audiovisual stimuli in Experiment 3. All other aspects of the experimental paradigm remained identical to those of Experiment 2. In both of these experiments, we found multisensory advantage in auditory but not visual working memory stores.

Furthermore, based on previous findings, we had hypothesized that presenting crossmodally congruent stimuli would lead to higher accuracy, faster reaction times, and larger capacity estimates. However, the results from Experiment 2 and Experiment 3 were not consistent with this prediction. In Experiment 2, we found that congruent stimuli enhanced accuracy and capacity only when auditory working memory was tested at loads of two or four items, but did not impact reaction times, or visual working memory performance. Interestingly, we did not find evidence for enhanced working memory processing when presented with congruent stimuli compared to incongruent stimuli in Experiment 3 when participants were presented with a larger memory load. On the contrary, we found evidence for incongruent audio-visual stimuli improving working memory performance in both experiments. In Experiment 2, an advantage for incongruent stimuli in the visual working memory domain was greatest at a load of four items, and in Experiment 3 this effect was greatest at load of eight items.

Our findings in the explicit congruency manipulation in Experiment 4 matched with the existing literature in the domain of auditory working memory as we found that participants had the highest memory performance when presented with congruent stimuli. On the contrary, we found that participants had the lowest memory performance when presented with congruent stimuli and tested on visual working memory. These results indicate that congruent auditory information is not aiding visual working memory, while congruent visual information is indeed aiding auditory working memory. We also found that participants overall had higher working memory performance for stimuli in the visual rather than auditory working memory domain. These results are similar to the ones from Experiment 4 in Chapter 2, where we used simple stimuli, and thus may be stemming from similar cognitive processes. More specifically, since participants were cued on the modality they were going to be tested on, they could’ve found that incongruent stimuli are easier to unbind compared to congruent stimuli, which are linked by implicit crossmodal correspondences.

This finding of enhanced memory performance when presented with incongruent stimuli contradicts a large portion of the findings in the current literature which cite that congruent stimuli lead to greater cognitive benefits compared to incongruent stimuli (Matusz et al., 2017). However, it is important to note that enhancements for congruently presented multisensory stimuli have mostly been found in the perceptual domain (Kim, Seitz, & Shams, 2008; Mishra & Gazzaley, 2012). For this reason it could be the case that congruent stimuli do not produce the same enhancements in the cognitive, namely memory domain (Brunel, Carvalho, & Goldstone, 2015). It is also possible that incongruent stimuli may be more salient due to the specific stimulus set that we used. A memory benefit for incongruent stimuli may be due to the mismatch of an animal image and sound being more unusual and thus more easily remembered by participants due to the semantic, learned, knowledge that participants have about these stimuli. For example, the sound of an insect buzzing presented with an insect image may be less salient than the roar of a lion presented with an insect leading to participants being better able to recall the incongruent, mismatching stimulus. These clashes of incongruently presented stimuli may work to be more memorable, especially given that these are stimuli that people have prior knowledge with the naturalistic stimuli used in these experiments. Additionally, these findings may also be influenced by the load that is exerted on mnemonic processes in a given task (Cowan, 2010). A larger discussion of these explanations is included in Chapter 5 of this thesis.

Using crossmodal stimuli consisting of animal sound and drawings provides a robust example of the naturally occurring, complex multisensory sensory in the environment. Previous studies have used complex stimuli such as speech, tools, and objects in motion to test effects of multisensory processing (Grant & Seitz, 1998; James et al., 2011; Stevenson & James, 2009; Stevenson & Wallace, 2013). However, this study is the first to use naturalistic size-pitch correspondences with animal stimuli to test MSI and working memory. A previous study demonstrated that animals exhibit crossmodal correspondences in a similar manner to humans (Ettlinger, 1961; Ludwig, Adachi, & Matsuzawa, 2011). These results led these researchers to conclude that crossmodal correspondences, such as those between loudness and brightness, and size and pitch are innate. Furthermore, it challenges the notion that crossmodal correspondences are dependent on linguistic and semantic processes (Martino & Marks, 2001). We add to this debate by using naturalistic, complex stimuli in humans and demonstrating an advantage in auditory, but not visual working memory.

Results of the experiments detailed in this chapter bring forward a plethora questions: Does the type of multisensory stimuli determine the memory domain in which they are beneficial? What are the mechanistic differences in auditory and visual working memory? What are the dependencies of memory processes on the stimuli presented? What are the behavioral and neural bases of these mechanisms? While various aspects of these questions remain unanswered, the current and previous studies can begin to shed light on some of these questions. There has been a debate in the working memory literature regarding the storage of auditory and visual information. While some studies have supported the notion of distinct storage systems for different sensory information (Baddeley, 1986), others have supported the notion of a central storage system (Cowan, 2005). Our results add to this debate by showing that information in working memory may indeed have a central when presented with multisensory stimuli but tested on only auditory or visual working memory. This central store could allow for the dynamic interaction of sensory representations such that visual information may be helping auditory working memory, but auditory information may not be aiding visual working memory when presented with certain types of stimuli. However, recent research has suggested that it may not be as simple as crafting mechanistic explanations of working memory mnemonics by divisions of sensory modalities. More specifically, elements such as the memory load tested, stimulus properties, and how performance in unisensory or multisensory tasks is measured may affect whether information is stored in a central or modality-specific store in working memory. To further understand the role of stimulus type in MSI, we tested the impact of simple and complex stimuli on working memory within the same group of participants (Chapter 4).

Chapter 4: Comparing the Impact of Simple and Complex Stimuli on Multisensory Integration and Working Memory Within Participants

Introduction

Multisensory integration is crucial for the activities that we do on an everyday basis, from having conversations, to detecting an ambulance in a crowded city square. To date, various studies have tested the impact of simple (e.g. flashes and beeps) and complex (e.g. speech) stimuli on multisensory integration. However, many of these studies have tested only one type of stimulus set within the study. Early studies of multisensory integration used simple stimuli to understand the organization of visual and auditory cortices, and were instrumental in establishing the correspondence between properties of vision and sound (Evans & Whitfield, 1964; Hubel & Wiesel, 1962; Spence, 2011). Indeed, simple stimuli elicit well-understood neural responses, are advantageous for isolating response properties, and are considered to be well-suited for perceptual experiments due to their negligible semantic content. After researchers began understanding fundamentals of MSI, simple stimuli were used to understand the various task properties that influenced crossmodal integration (Gepshtein et al., 2005). More recently, researchers have investigated the importance of semantic properties of multisensory stimuli by using richer complex stimuli such as speech, and motion (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007).

Combined, these investigations have advanced our knowledge of multisensory processing, and have elucidated the mechanisms behind this process at the behavioral and neural levels. While these studies remain foundational to the field, there is a gap in our knowledge regarding mechanistic differences in MSI and working memory using simple and complex stimuli within the same group of individuals. One study tested multisensory temporal integration using different stimulus types in the same participants (Stevenson & Wallace, 2013). This study used simple stimuli such as flashes and beeps along with dynamic stimuli such as handheld tools and speech in the form of single syllable utterances. This study concluded that the type of multisensory stimulus presented to participants has an effect on the time window during which these stimuli to bind into a single coherent percept (i.e. temporal binding window). While this study tested MSI using a variety of stimuli in the same group of participants, the analyses and conclusions were focused on understanding the dynamics of temporal binding, and not stimulus differences. Additionally, the twelve conditions tested in this study were all basic perceptual tasks. That is, to date, no study has investigated the influence of stimulus type on multisensory integration in working memory in the same participants. In Chapters 2 and 3 of this dissertation, I compared the impact of MSI on WM using simple and complex stimuli, respectively. However, as these studies tested different participants, the effects of stimulus properties on MSI for WM are confounded with participant factors. In order to control for these differences, and also to replicate the key findings from Chapters 2 and 3, I carried out the study described here.

The Current Study

The aim of this study is to extend the findings of Chapter 2 and 3 by testing the impact of simple and complex stimuli on MSI and working memory in the same group of experimental subjects. To do this, we replicated the change detection paradigm detailed in the previous chapters with a memory load of two, four, or eight stimuli (Experiment 2 in Chapter 2 and Experiment 2 in Chapter 3). Additionally, we used the same simple stimulus set of color patches and instrument tones and the same complex stimulus set of line drawings and animal sounds used in those studies. We expected that participants would show enhanced working memory processing when presented with complex compared with simple stimuli, particularly when testing auditory working memory. We expected to find differences at the modality-specific level.

In particular, we expected participant to have an advantage in working memory when presented with multisensory compared to unisensory stimuli in the auditory (AV-Test A compared to A) and visual (AV-Test V compared to V) domains. We hypothesized that we would find effects of congruency such that participants would have higher working memory on trials that have congruently presented audiovisual stimuli, compared to those that do not, based on the principles of crossmodal correspondence (simple stimuli in Chapter 2 and complex stimuli in Chapter 3; Brunetti et al., 2017). We also expected that participants would perform with lower accuracy and have longer reaction times on trials with a higher, compared to a lower, memory load. Conversely, we hypothesized the capacity estimates would increase as the memory load increased (Rouder et al., 2011). The current study represents the first to assess the impact of simple versus complex stimuli on working memory and multisensory integration within the same set of experimental subjects.

Methods

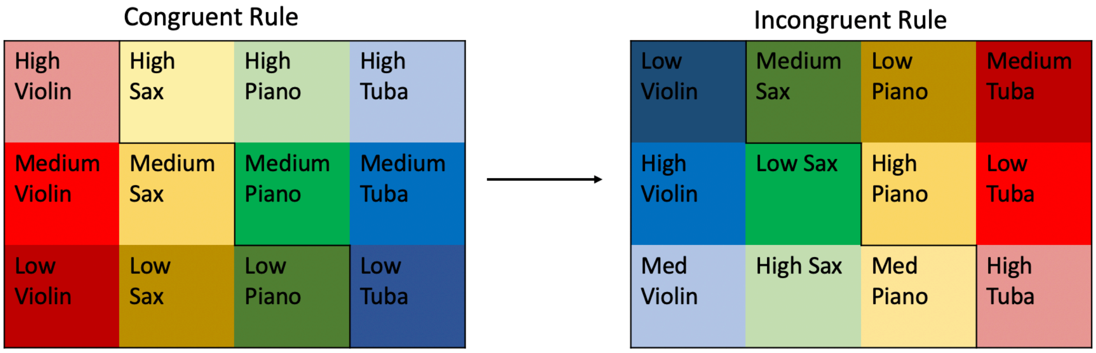
Participants & Procedure

We recruited forty-eight adult participants for this study. Participants were compensated for their participation and time with partial course credit. All components of the study were approved by the Institutional Review Board (IRB #2018-0395). Thirty-nine (mean=20.08 years, M=7, F=32) participants were included in the final data set after the data cleaning analyses were conducted (see below). The procedure was administered in a quiet room in the Educational Neuroscience Lab space. Experiments were programmed with E-prime 2.0.8.90a (Psychology Software Tools, Sharpsburg, PA) on a Dell Optiplex 390 Desktop PC (3.1 GHz, 4 GB RAM) running Windows 7.0 64-bit operating system. Visual stimuli were presented on a Dell UltraSharp U2212H 21.5” flat-screen monitor at a resolution of 1024 x 768 and a refresh rate of

60 Hz. Auditory stimuli were presented using headphones (Sony MDR-ZX110 Black). We tested each participant on two working memory tasks, using simple (Experiment 2 in Chapter 2) and complex stimuli (Experiment 2 in Chapter 3), across two experimental sessions separated by seven days. The stimulus type tested was counterbalanced across participants such that half the participants were tested on working memory with simple stimuli in the first session and complex stimuli in the second session and the other half of the participants were tested in the opposite order.

Simple Stimuli & Paradigm

To understand the influence of simple auditory-visual correspondence on MSI of WM, we presented participants with one of four different instrumental timbres: violin, piano, tuba, and saxophone, and four colors corresponding to these sounds: red, green, blue, yellow, respectively. We have found evidence for these timbre-color associations a study of synesthetes and non-synesthetes previously conducted in the Educational Neuroscience Lab, and similar associations have been also documented in previous research by other groups (Gosavi, Bade, & Hubbard, in prep; Ward, Huckstep, & Tsakanikos, 2006). Following principles of color-pitch



Congruent Rule

High Violin	High Sax	High Piano	High Tuba
Medium Violin	Medium Sax	Medium Piano	Medium Tuba
Low Violin	Low Sax	Low Piano	Low Tuba

Incongruent Rule

Low Violin	Medium Sax	Low Piano	Medium Tuba
High Violin	Low Sax	High Piano	Low Tuba
Med Violin	High Sax	Med Piano	High Tuba

Figure 4.1. Stimulus set used to test the impact of simple stimuli on working memory. The stimuli were presented as per the cross-modal correspondences in the congruent rule. The colors were flipped across the diagonal line and sounds were rotated within the columns to create the schematic that formed the incongruent rule.

correspondences (Martino & Marks, 1999; Melara, 1989), each of these sounds were presented in either a high, medium, or low pitch, and each color was presented in a light, medium or dark value (shade), resulting in a total of 12 stimuli (Figure 4.1). The stimuli were presented consistently with the crossmodal correspondences in congruent trials (e.g. high-pitched violin tone paired with the light shade of red). Crossmodal correspondence were systematically disrupted in incongruent trials. To form the incongruent grid, we systematically varied the distance of each stimulus from its position in the congruent grid by exchanging the positions of the colors across the diagonal line and rotating the pitch of the tones for a given instrument.

On each trial, participants were presented with a different number of to-be-remembered items (i.e. memory load, 2, 4 or 8 items). In a given trial, each stimulus was presented for 750 ms with a 250 ms interstimulus interval (ISI). After a delay period (4000 ms), participants were

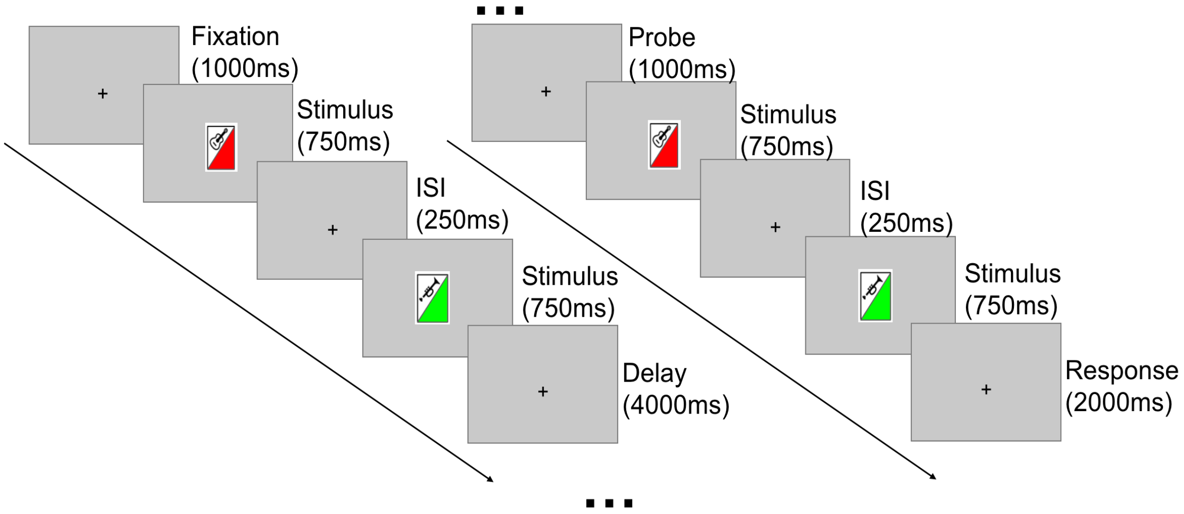


Figure 4.2. The multisensory working memory paradigm used with a set of simple stimuli. Participants were presented with a randomly selected array of two, four, or eight sounds or color patches. After a delay, they were presented with either the same pattern or a different pattern, and were asked to detect a change in the pattern presented during the response duration.

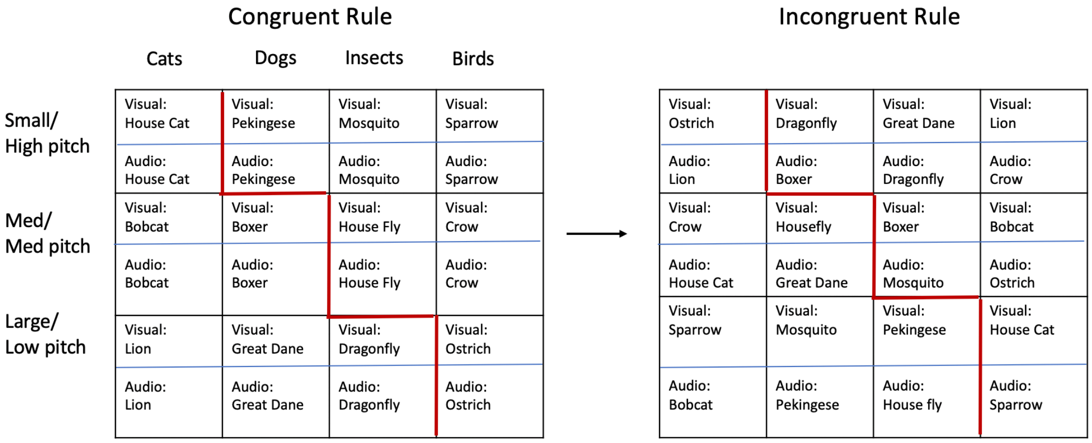
presented with either the same pattern or a different pattern of stimuli, and participants were asked to detect a change in the pattern presented in the response duration (2000 ms) (Figure 4.2). Only one item out of the string of stimuli changed in the change trials, and none of the stimuli changed in the no-change trials.

In the unisensory auditory (A) and visual (V) baseline conditions, participants were presented with and tested only on stimuli from the given modality. In the multisensory (audio-visual: AV) condition, participants were presented with auditory and visual stimuli in a temporally and spatially synchronized manner, but were only tested on stimuli from one modality (Conditions: AV-Test A or AV-Test V). Participants were cued to the condition (A, V, or AV) in a blocked fashion, which was counterbalanced across participants. We cued participants on this task as we found effects of attention comparing we compared results from Experiments 1 and 2 in Chapter 2.

In the multisensory conditions, participants were presented with congruent trials in which cross-sensory correspondences were maintained on half the trials, and incongruent trials in which auditory and visual stimuli were presented as per the varied grid on half the trials (Figure 4.1). All the stimuli were presented with congruent pairings in congruent trials, and incongruent pairings in incongruent trial.

Complex Stimuli & Paradigm

We used naturalistic size-pitch crossmodal correspondences found in animals as our complex stimulus set. Visual stimuli comprised of black-and-white line drawings of the animals, and auditory stimuli were clips of the corresponding sound that the animal produced (Figure 3.1). We extracted pictures of animals from the Snodgrass image database, a standardized set of 260 black and white line drawings (Snodgrass & Vanderwart, 1980). Auditory clips of animal sounds were downloaded from Cornell Lab of Ornithology's Macaulay Library, which has a freely available database of over 150,000 animal sounds. Each of the four animal categories consisted of three animals which were small, medium, or large in size and produced a high, medium, or low pitch sound respectively, resulting in a total of 12 stimuli (Figure 4.3).



Congruent Rule

	Cats	Dogs	Insects	Birds
Small/ High pitch	Visual: House Cat Audio: House Cat	Visual: Pekingese Audio: Pekingese	Visual: Mosquito Audio: Mosquito	Visual: Sparrow Audio: Sparrow
Med/ Med pitch	Visual: Bobcat Audio: Bobcat	Visual: Boxer Audio: Boxer	Visual: House Fly Audio: House Fly	Visual: Crow Audio: Crow
Large/ Low pitch	Visual: Lion Audio: Lion	Visual: Great Dane Audio: Great Dane	Visual: Dragonfly Audio: Dragonfly	Visual: Ostrich Audio: Ostrich

Incongruent Rule

Visual: Ostrich Audio: Lion	Visual: Dragonfly Audio: Boxer	Visual: Great Dane Audio: Dragonfly	Visual: Lion Audio: Crow
Visual: Crow Audio: House Cat	Visual: Housefly Audio: Great Dane	Visual: Boxer Audio: Mosquito	Visual: Bobcat Audio: Ostrich
Visual: Sparrow Audio: Bobcat	Visual: Mosquito Audio: Pekingese	Visual: Pekingese Audio: House fly	Visual: House Cat Audio: Sparrow

Figure 4.3. The set of stimuli used to test the impact of complex stimuli on working memory. The stimuli were presented as per the cross-modal correspondences in the congruent rule. The colors were flipped across the diagonal line and sounds were rotated within the columns to create the schematic to form the incongruent rule.

Congruent pairs were those in which cross-sensory correspondences were maintained (e.g. a lion roar paired with the image of a lion), and incongruent trials were those in which these crossmodal correspondences were not maintained (e.g. a lion roar paired with the visual of an ostrich). Analogous to the procedure used with the simple stimuli, animal images were flipped across the diagonal line and sounds were rotated within the columns to create the schematic that formed the incongruent rule. In this way, we were able to systematically vary the distance of each stimulus from its position in the congruent grid.

Participants were presented with a different number of to-be-remembered items (i.e. memory load) in each trial. In this experiment, participants were presented with a memory load of two, four, or eight items. In a given trial, each stimulus was presented for 750 ms with an interval of 250 ms between stimuli. After a delay period (4000 ms), participants were presented with either the same pattern or a different pattern of stimuli, and participants were asked to indicate whether they thought the test pattern contained a change during the response period (2000 ms). Only one item out of the string of stimuli changed in the change trials, and none of the stimuli changed in the no-change trials (Figure 4.4).

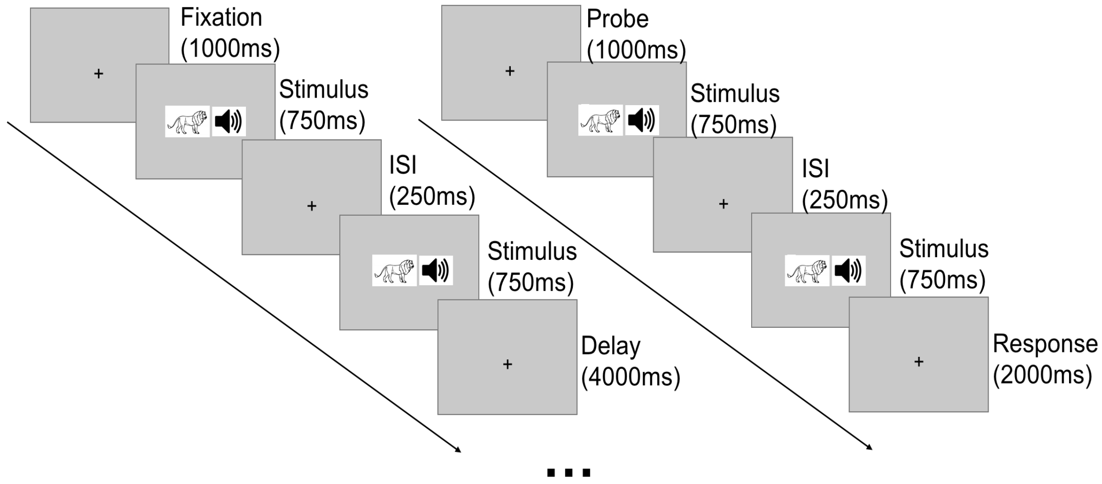


Figure 4.4. The multisensory working memory paradigm used with a set of complex stimuli. Participants will be presented with a randomly selected array of two, four, or eight animal sounds or animal pictures. After a delay, they will be presented with either the same pattern or a different pattern, and will be asked to detect a change in the pattern presented during the response duration.

In the unisensory auditory (A) and visual (V) baseline conditions, participants were presented with and tested on stimuli only from the given modality. In the multisensory (audio-visual: AV) condition, participants were presented with auditory and visual stimuli in a temporally and spatially synchronized manner, but were only tested on stimuli from one modality in each block (Conditions: AV-Test A or AV-Test V). Participants were cued to the condition (A, V, or AV) in a blocked fashion, which was counterbalanced across participants.

In the multisensory conditions, on half of the trials, participants were presented with congruent stimuli in which cross-sensory correspondences were maintained, and incongruent stimuli in which crossmodal correspondences were not maintained on half of the trials. All the stimuli presented had congruent pairings in a congruent trial, and incongruent pairings in an incongruent trial. To form the incongruent grid, we systematically varied the distance of each stimulus from its position in the congruent grid by exchanging the positions of the images across the diagonal line and rotating the pitch of the tones for a given animal.

Analysis

Accuracy, capacity estimates, and reaction times were calculated, as together they provide a deeper understanding of the working memory mechanisms at play. While performance accuracy tends to decrease as the mnemonic load increases, individual differences in working memory capacity have also been noted (Linderholm et al., 2008; Luck & Vogel, 2013). Accuracy was calculated based on the proportion of correct responses on each trial. All participants with accuracies below 50% (chance performance) at lowest load in baseline A or V conditions were removed from the final analyzed data set. Reaction times were calculated as the time it took participants to respond (in milliseconds) from the offset of the last stimulus in the memory array.

Working memory capacity was calculated using Cowan's k (Cowan, 2001, 2010; Rouder et al., 2011). Under the assumption that participants have a working memory capacity that is fully employed on each trial, this formula to calculate working memory capacity (k) is detailed in the following manner: The probability that the participant makes a hit is $H= k/N + g((N-k)/N)$, where k = working memory capacity, N=total number of stimuli presented, and g is the probability of guessing correctly (based on the number of choices). That is, the probability of a hit is the sum of the probability that the participant detects the change (k/N) plus the probability that the participant guesses correctly, even when they do not detect the change ($g((N-k)/N)$). Based on this formula, the probability that the participant makes a correct rejection is $CR= k/N + (1-g)((N-k)/N)$. Combining these two formulas and solving for k, we can obtain the formula to calculate working memory capacity: *k=N(H+CR-1).* Working memory capacity was calculated in the same manner here. Statistical results for all the pairwise comparisons of working memory performance (accuracy, reaction time, capacity) across conditions conducted for experiments in this chapter can be found in Appendix E.

Results

To determine if results from the two sessions of testing could be collapsed across participants, we conducted a repeated measures ANOVA with session as a between-subject factor, with load and condition as the within-subject factors for accuracies, reaction times, and capacity estimates. Results with simple stimuli did not reveal main effects of session on accuracy (F(1,38)=0.07, p=0.79, η_p^2<0.001), reaction times (F(1,38)=0.59, p=0.45, η_p^2=0.02), or capacity estimates (F(1,38)=0.07, p=0.80, η_p^2<0.001). Results with complex stimuli did not reveal main effects of session with accuracies (F(1,38)=0.31, p=0.58, η_p^2=0.01), and capacity estimates (F(1,38)=0.78, p=0.38, η_p^2=0.02). Analysis of reaction times with complex did show a significant effect of session (F(1,38)=32.19, p<0.001, η_p^2=0.47), but no significant condition x session (F(1,38)=0.26, p=0.85, η_p^2=0.01), nor significant load x session (F(1,38)=2.31, p=0.13, η_p^2=0.06) interactions. Since reaction time measures were not our main analytical criteria of interest, and differences between sessions were not found with both stimuli sets, we decided to collapse participant data across the two testing sessions. Furthermore, we did not find an effect of session on accuracy or capacity estimates, our main analyses of interest, when presented with simple or complex stimuli. For a full listing of results from this analysis, see Appendix D.

Simple Stimuli

Effects of Modality

Accuracy

As can be seen in Figure 4.5A, overall accuracy decreased as memory load increased. Additionally, we found evidence of a multisensory advantage when participants were tested on visual working memory, but not when tested on auditory working memory. A 3 (load) x 4 (condition) ANOVA revealed significant main effects of load (F(1,38)=159.26, p<0.001, η_p^2=0.81) and condition (F(1,38)=13.23, p<0.001, η_p^2=0.26), and a significant load x condition interaction (F(1,38)=7.34, p<0.001, η_p^2=0.16). As shown in Figure 4.5A, accuracy was highest in the AV-test V condition, followed by the visual alone, and then AV- test A and auditory alone condition, but these effects were modulated by interactions with load. To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. When comparing accuracy for unisensory auditory and visual conditions, we found significant main effects of condition (F(1,38)=9.78, p<0.01, η_p^2=0.21) and load (F(1,38)=74.44,

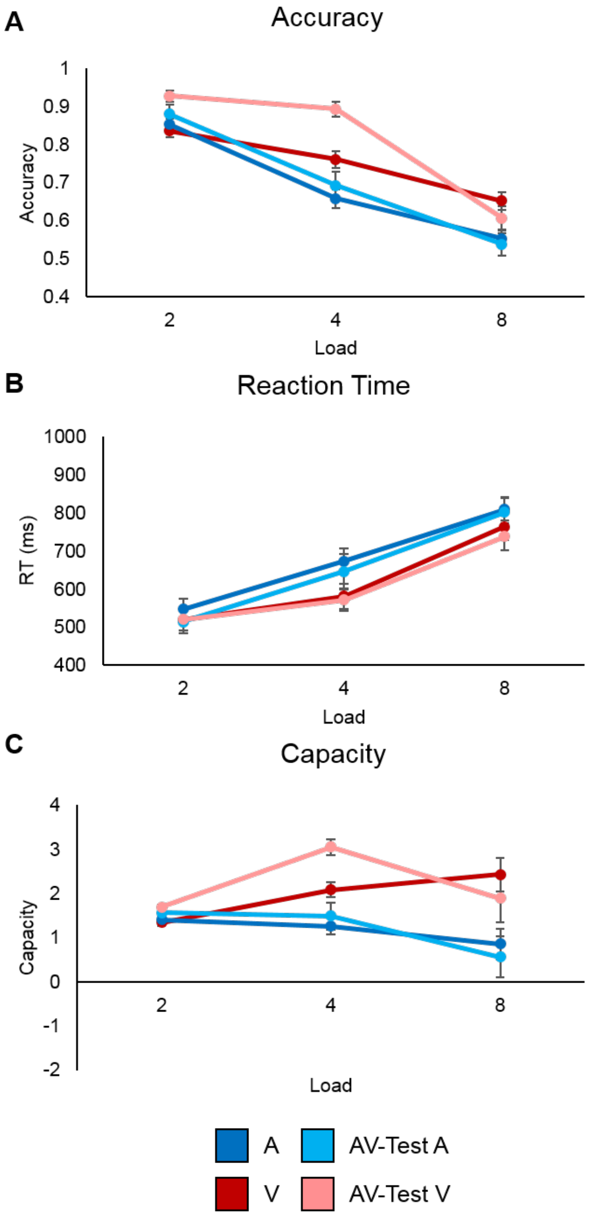


Figure 4.5 Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants in the unisensory (V, A) or multisensory (AV-test A, AV-test V) conditions. Error bars represent the standard error of the mean.

$p<0.001$, $\eta_p^2=0.66$), and a significant load x condition interaction ($F(1,38)=8.42$, $p<0.001$, $\eta_p^2=0.18$). Specifically, accuracy was higher in the visual alone condition than in the auditory alone condition at loads of 4 ($t(38)=3.00$, $p<0.01$), and 8 ($t(38)=3.00$, $p<0.01$) items. Comparing accuracy for unisensory visual and AV-Test V conditions, we found significant main effects of condition ($F(1,38)=10.43$, $p<0.01$, $\eta_p^2=0.22$) and load ($F(1,38)=119.56$, $p<0.001$, $\eta_p^2=0.76$), and a significant load x condition interaction ($F(1,38)=12.71$, $p<0.001$, $\eta_p^2=0.25$). AV test-V accuracy was significantly higher than visual alone (V) at loads of 2 ($t(38)=4.69$, $p<0.001$), and 4 ($t(38)=5.00$, $p<0.001$) items. We found no evidence of an auditory multisensory enhancement, as unisensory auditory and AV-Test A were not significantly different (load $F(1,38)=89.58$, $p<0.001$, $\eta_p^2=0.70$, condition $F(1,38)=0.42$, $p=0.52$, $\eta_p^2=0.01$, load x condition interaction $F(1,38)=0.65$, $p=0.52$, $\eta_p^2=0.02$).

Reaction Time

Our results show that the reaction times of participants were significantly different at different loads and vary by condition of the stimuli presented (Figure 4.5B). A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load ($F(1,38)=76.67$, $p<0.001$, $\eta_p^2=0.67$), condition ($F(1,38)=2.90$, $p=0.04$, $\eta_p^2=0.07$), but not a significant interaction of load x condition ($F(1,38)=0.84$, $p=0.54$, $\eta_p^2=0.02$). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. Reaction times were significantly longer in the unisensory auditory condition compared to the unisensory visual condition. A 2 (condition) x 3 (load) ANOVA revealed a significant main effect of condition ($F(1,38)=5.15$, $p=0.03$, $\eta_p^2=0.12$), load ($F(1,38)=51.55$, $p<0.001$, $\eta_p^2=0.58$), but did not show a significant load x condition interaction ($F(1,38)=1.50$, $p=0.23$, $\eta_p^2=0.04$). Specifically, a significant difference was found at a load of 4 ($t(38)=2.86$, $p<0.01$) items. We did not find evidence of a multisensory enhancement, as unisensory visual and AV-Test V were not significantly different (condition F(1,38)=0.20, p=0.66, η_p^2<0.01, load F(1,38)=49.12, p<0.001, η_p^2=0.57, load x condition interaction F(1,38)=0.25, p=0.78, η_p^2<0.01), and unisensory auditory and AV-Test A were not significantly different (condition F(1,38)=0.87, p=0.36, η_p^2=0.02, load F(1,38)=42.75, p<0.001, η_p^2=0.53, load x condition interaction F(1,38)=0.11, p=0.90, η_p^2<0.01).

Capacity

Visual inspection of Figure 4.5C reveals a multisensory advantage when tested in the visual working memory domain but not the auditory working memory domain. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,38)=3.70, p=0.03, η_p^2=0.09), condition (F(1,38)=8.85, p<0.001, η_p^2=0.19), and a significant interaction of load x condition (F(1,38)=4.82, p<0.001, η_p^2=0.11). Capacity estimates were highest in the AV-test V, followed by the unisensory visual, AV-Test A, and unisensory auditory conditions. To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. When comparing accuracy for unisensory auditory and visual conditions, we found a significant main effect of condition (F(1,38)=12.52, p=0.001, η_p^2=0.25), and a significant load x condition interaction (F(1,38)=8.76, p<0.001, η_p^2=0.19) but not a significant load (F(1,38)=0.97=, p=0.39, η_p^2=0.03). Furthermore, participants had significantly higher capacity estimates in the unisensory visual compared to the unisensory auditory conditions at loads of 4 (t(38)=3.15, p<0.01), and 8 (t(38)=3.31, p<0.01) items. We also found a multisensory advantage when tested on visual working memory. When comparing accuracy for unisensory visual and AV-Test V conditions, we did not a significant main effect of condition (F(1,38)=1.14, p=0.29, η_p^2=0.03), but did find a significant main effect of load (F(1,38)=8.16, p=0.001, η_p^2=0.18), and a significant load x condition interaction (F(1,38)=4.04, p=0.02, η_p^2=0.10). Capacity estimates were significantly higher for the multisensory audiovisual (test V) compared to the unisensory visual condition at loads of 2 (t(38)=4.08, p<0.001), and 4 (t(38)=4.45, p<0.001) items. We did not find evidence of an auditory multisensory enhancement, as unisensory auditory and AV-Test A were not significantly different (condition F(1,38)=0.01, p=0.91, η_p^2<0.001, load F(1,38)=4.37, p=0.02, η_p^2=0.10, load x condition interaction F(1,38)=0.67, p=0.52, η_p^2=0.02).

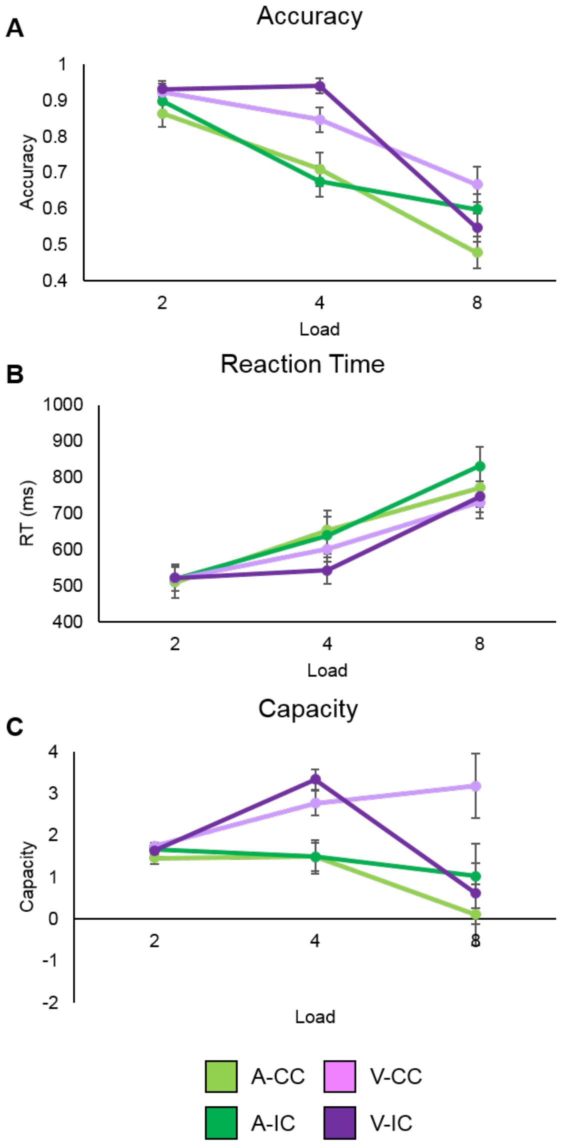


Figure 4.6 Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants when presented with congruent and incongruent stimuli and tested on the auditory (A-CC & A-IC) and visual (V-CC & V-IC) modalities. Error bars represent the standard error of the mean.

<u>Effects of Congruency</u>

Accuracy

Figure 4.6A shows participant performance when presented with congruent versus incongruent stimuli in the multisensory condition. This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). By visually inspecting Figure 4.6A, it is seen that participants performed with higher accuracy when presented with incongruent stimuli in the visual domain. Additionally, participants did not show a difference in performance when presented with congruent versus incongruent stimuli in the auditory domain. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,38)=102.08, $p<0.001$, η_p^2=0.73), condition (F(1,38)=8.93, $p<0.001$, η_p^2=0.19), and a significant interaction of load x condition (F(1,38)=3.89, p=0.001, η_p^2=0.09). To directly contrast performance between congruent and incongruent stimuli, we ran a series of 2 (condition) x 3 (load) ANOVAs. When comparing accuracy for congruent and incongruent stimuli in the visual domain, we found a significant main effect of load (F(1,38)=83.02, $p<0.001$, η_p^2=0.69), and a significant load x condition interaction (F(1,38)=5.40, $p<0.01$, η_p^2=0.12), but not a significant main effect of condition (F(1,38)=0.04, p=0.85, $\eta_p^2<0.01$). Accuracy was higher when presented with incongruent stimuli compared to congruent stimuli and tested on visual working memory at a load of 4 (t(38)=2.00, p=0.03) items. However, there was no difference between performance when presented with congruent and incongruent stimuli and tested on auditory working memory (condition F(1,38)=1.89, p=0.18, η_p^2=0.05, load F(1,38)=38.52, $p<0.001$, η_p^2=0.50, load x condition interaction F(1,38)=1.66, p=0.20, η_p^2=0.04).

Reaction Time

Figure 4.6B shows participant performance when presented with congruent versus incongruent stimuli in the multisensory condition. This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,38)=36.01, $p<0.001$,

η_p^2=0.49), but not a significant effect condition (F(1,38)=1.67, p=0.18, η_p^2=0.04), nor a significant interaction of load x condition (F(1,38)=1.67, p=0.57, η_p^2=0.02).

Capacity

Figure 4.6C shows participant performance when presented with congruent versus incongruent stimuli in the multisensory condition. This analysis showed results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC).A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,38)=5.52, p<0.01, η_p^2=0.13), condition (F(1,38)=6.01, p=0.001, η_p^2=0.14), and a significant interaction of load x condition (F(1,38)=3.09, p<0.01, η_p^2=0.08). To directly contrast performance between congruent and incongruent stimuli, we ran a series of 2 (condition) x 3 (load) ANOVAs. We did not find differences at loads of two, four, or eight items in auditory working memory performance when presented with congruent versus incongruent stimuli (condition F(1,38)=1.00, p=0.32, η_p^2=0.03, load F(1,38)=3.41, p=0.04, η_p^2=0.08, load x condition interaction F(1,38)=0.39, p=0.68, η_p^2=0.01). Participants had greater capacity estimates for visual working memory when presented with congruent compared to incongruent stimuli, but only at a load of eight items (t(38)=2.48, p=0.02) items. We found a marginal effect of condition (F(1,38)=3.39, p=0.07, η_p^2=0.08), a significant main effect of load (F(1,38)=5.76, p<0.01, η_p^2=0.13), a significant load x condition interaction (F(1,38)=6.81, p<0.01, η_p^2=0.15) when tested on visual working memory.

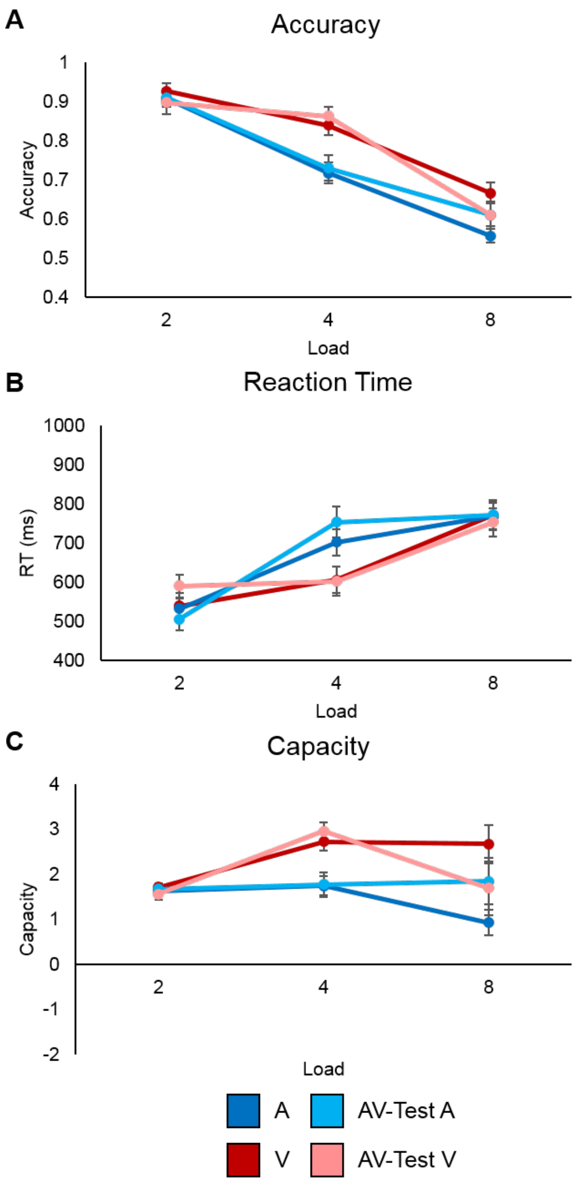


Figure 4.7 Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants in the unisensory (V, A) or multisensory (AV-test A, AV-test V) conditions. Error bars represent the standard error of the mean.

Complex Stimuli

<u>Effects of Modality</u>

Accuracy

As can be seen in Figure 4.7A, overall, accuracy of participants decreased as the memory load increased. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,38)=196.99, p<0.001, η_p^2=0.84), condition (F(1,38)=9.47, p<0.001, η_p^2=0.20), and a significant interaction of load x condition (F(1,38)=3.80, p=0.001, η_p^2=0.09). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. We did not find a multisensory enhancement when tested on visual (condition F(1,38)=1.67, p=0.20, η_p^2=0.04, load F(1,38)=87.40, p<0.001, η_p^2=0.70, load x condition interaction F(1,38)=2.42, p=0.11, η_p^2=0.06) or auditory working memory (condition F(1,38)=2.10, p=0.16, η_p^2=0.05, load F(1,38)=100.35, p<0.001, η_p^2=0.73, load x condition interaction F(1,38)=0.88, p=0.42, η_p^2=0.02). However, we did find that participants had significanly higher accuracy in the unisensory visual compared to the unisensory auditory conditions (condition F(1,38)=27.80, p<0.001, η_p^2=0.42, load F(1,38)=171.89, p<0.001, η_p^2=0.82, load x condition interaction F(1,38)=5.30, p<0.01, η_p^2=0.12). This difference was specifically found at a load of 4 (t(38)=4.34, p<0.001), and 8 (t(38)=3.88, p<0.001) items.

Reaction Time

Our results show that the reaction times of participants were significantly different at different loads and varied by condition of the stimuli presented (Figure 4.7B). A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,38)=57.27, p<0.001, η_p^2=0.60), but not of condition (F(1,38)=1.05, p=0.37, η_p^2=0.03), and a significant interaction of load x condition (F(1,38)=5.55, p<0.001, η_p^2=0.13). Overall, participants had the longest reaction times in the AV-Test A condition followed by the unisensory auditory contion.

Capacity

Visual inspection of Figure 4.7C reveals that capacity estimates for the multisensory and unisensory conditions plateaued or decreased as the memory load increased. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,38)=4.99, p<0.01, η_p^2=0.12), condition (F(1,38)=6.83, p<0.001, η_p^2=0.15), and a significant interaction of load x condition (F(1,38)=2.98, p<0.01, η_p^2=0.07). To directly contrast performance between modalities, we ran a series of 2 (condition) x 3 (load) ANOVAs. We found that capacity estimates for the unisensory visual condition were significantly higher than the unisensory auditory condition were compared (condition F(1,38)=26.91, p<0.001, η_p^2=0.42, load F(1,38)=3.73, p=0.03, η_p^2=0.09, load x condition interaction F(1,38)=8.48, p<0.001, η_p^2=0.18) at loads of 4 (t(38)=4.32, p<0.01), and 8 (t(38)=3.86, p<0.01) items. We did not find evidence of an auditory multisensory enhancement, as unisensory auditory and AV-Test A were not significantly different (condition F(1,38)=2.72, p=0.11, η_p^2=0.07, load F(1,38)=0.95, p=0.39, η_p^2=0.02, load x condition interaction F(1,38)=2.08, p=0.13, η_p^2=0.05). We also did not find evidence of a visual multisensory enhancement, as unisensory visual and AV-Test V were not significantly different (condition F(1,38)=1.92, p=0.17, η_p^2=0.05, load F(1,38)=6.78, p<0.01, η_p^2=0.15, load x condition interaction F(1,38)=3.07, p=0.05, η_p^2=0.08).

Effects of Congruency

Accuracy

Figure 4.8A shows results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,38)=72.06, p<0.001, η_p^2=0.66), but not a significant effect of condition (F(1,38)=2.01, p=0.12, η_p^2=0.05), and revealed a significant interaction of load x condition (F(1,38)=2.15, p=0.05, η_p^2=0.05).

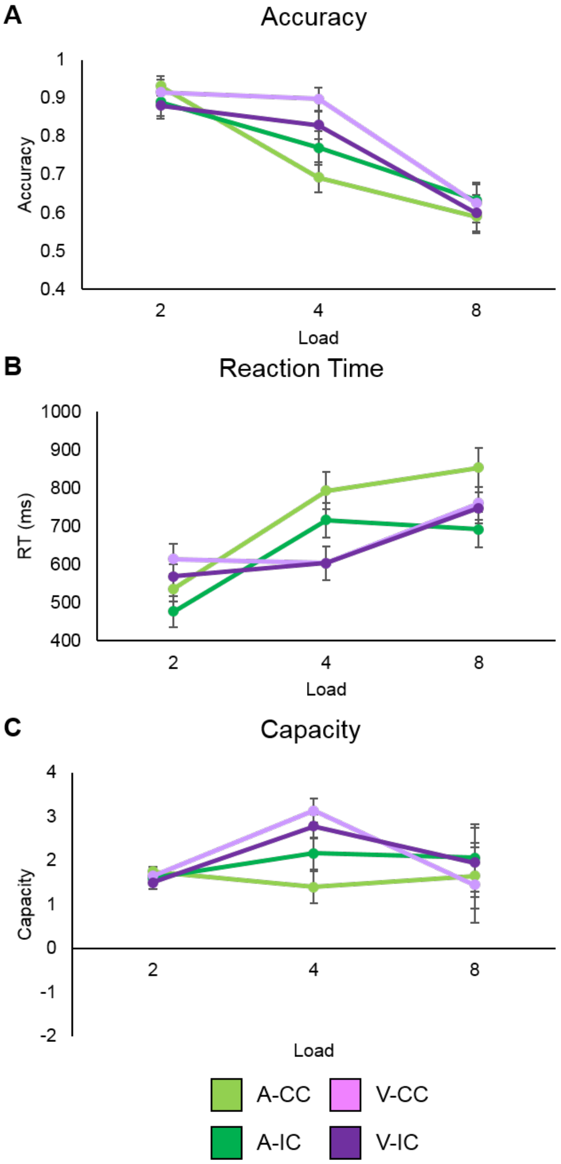


Figure 4.8 Working memory accuracies (A) reaction times (B) and capacity estimates (C) for participants when presented with congruent and incongruent stimuli and tested on the auditory (A-CC & A-IC) and visual (V-CC & V-IC) modalities. Error bars represent the standard error of the mean.

Reaction Time

Figure 4.8B shows participant performance when presented with congruent versus incongruent stimuli and tested on visual and auditory working memory. A 3 (load) x 4 (condition) ANOVA revealed a significant main effect of load (F(1,38)=24.02, p<0.001, η_p^2=0.38), condition (F(1,38)=4.52, p<0.001, η_p^2=0.11), and a significant interaction of load x condition (F(1,38)=4.18, p=0.001, η_p^2=0.10). To directly contrast performance between congruent and incongruent stimuli, we ran a series of 2 (condition) x 3 (load) ANOVAs. Participants had longer reaction times when presented with congruent stimuli in the auditory domain as we found a significant main effect of condition (F(1,38)=12.10, p<0.01, η_p^2=0.24), and load (F(1,38)=25.65, p<0.001, η_p^2=40), but not a significant load x condition interaction (F(1,38)=1.00, p=0.37, η_p^2=0.3). We did not find differences in reaction times when congruent and incongruent stimuli were compared in the visual modality (condition F(1,38)=0.60, p=0.44, η_p^2=0.02, load F(1,38)=10.29, p<0.001, η_p^2=0.21, load x condition interaction F(1,38)=0.30, p=0.74, η_p^2<0.01).

Capacity

Figure 4.8C shows results from trials in which participants were presented with congruent (CC) or incongruent (IC) multisensory stimuli and were tested on auditory (A-CC, A-IC) or visual working memory (V-CC, V-IC). A 3 (load) x 4 (condition) ANOVA revealed a marginally significant main effect of load (F(1,38)=2.77, p=0.07, η_p^2=), but a non-significant effect of condition (F(1,38)=0.66, p=0.58, η_p^2=0.02), and a non-significant interaction of load x condition (F(1,38)=1.1, p=0.36, η_p^2=0.03).

Effects of Stimulus

To directly contrast the results from the two stimulus types (simple and complex), we conducted an omnibus repeated measures ANOVA with stimulus, condition, and load as within-subject factors. We first conducted this analysis when participants were presented with unisensory (Conditions: A and V) stimuli compared to multisensory stimuli (Conditions: AV-Test A and AV-Test V). Results with accuracies revealed significant main effects of stimulus (F(1,38)=5.72, p=0.02, η_p^2=0.13), condition (F(1,38)=17.96, p<0.001, η_p^2=0.32), and load (F(1,38)=305.36, p<0.001, η_p^2=0.89). Additionally, we found that the condition tested is affected by the stimulus (F(1,38)=3.78, p=0.01, η_p^2=0.09), and load (F(1,38)=10.51, p<0.001, η_p^2=0.22). However, we did not find a significant stimulus x load (F(1,38)=0.29, p=0.74, η_p^2=0.01), nor a significant three-way stimulus x condition x load (F(1,38)=1.22, p=0.30, η_p^2=0.03) interaction. Results with reaction times revealed a significant main effect of load (F(1,38)=101.42, p<0.001, η_p^2=0.73), a marginal effect of condition (F(1,38)=2.48, p=0.06, η_p^2=0.06), but not a significant effect of stimulus (F(1,38)=0.55, p=0.46, η_p^2=0.01). We found the load presented had a significant effect on the condition tested (F(1,38)=5.34, p<0.001, η_p^2=0.12), and a marginally significant effect on the stimulus presented (F(1,38)=2.96, p=0.06, η_p^2=0.07). However, we did not find that the stimulus type presented and condition tested have a significant relationship (F(1,38)=1.04, p=0.38, η_p^2=0.03). Furthermore, we did not a three-way stimulus x condition x load interaction (F(1,38)=1.09, p=0.37, η_p^2=0.03). Results with capacity estimates were similar to those of accuracy. We found a significant main effect of load (F(1,38)=6.62, p<0.01, η_p^2=0.15), and condition (F(1,38)=12.49, p<0.001, η_p^2=0.25), but a marginally significant effect of stimulus (F(1,38)=3.86, p=0.06, η_p^2=0.09). Additionally, we found that load has a significant effect on the condition tested (F(1,38)=6.54, p<0.001, η_p^2=0.15), but not on the stimulus presented (F(1,38)=0.37, p=0.69, η_p^2=0.01). We did not find that the stimulus presented has an effect on the condition tested (F(1,38)=1.96, p=0.13, η_p^2=0.05). Lastly, we did find a three-way stimulus x condition x load interaction for capacity estimates (F(1,38)=1.10, p=0.36, η_p^2=0.03).

Next, we conducted an analogous repeated measures ANOVA with stimulus (simple or complex), condition (congruent: V-CC, A-CC or incongruent V-IC, A-IC), and load (two, four, or eight) as within subject factors. We did not find significant differences in accuracy when participants were presented with simple or complex stimuli (F(1,38)=0.52, p=0.47, η_p^2=0.01). However, we did find a significant effect of memory load presented (F(1,38)=187.80, p<0.001, η_p^2=0.83), and condition tested (F(1,38)=8.80, p<0.001, η_p^2=0.19). We found that the load presented has a significant interaction with the condition (F(1,38)=4.62, p=0.001, η_p^2=0.11), but not with the stimulus type (F(1,38)=0.72, p=0.48, η_p^2=0.02). We did not find a significant interaction between the stimulus presented and the condition tested (F(1,38)=1.91, p=0.13,

η_p^2=0.05). Lastly, we did not find a significant three-way stimulus x condition x load (F(1,38)=1.65, p=0.14, η_p^2=0.04) interaction. Results with reaction times revealed significant main effects of condition (F(1,38)=2.98, p=0.03, η_p^2=0.07), and load (F(1,38)=56.11, p<0.001, η_p^2=0.60), but not of stimulus (F(1,38)=1.50, p=0.23, η_p^2=0.04). Additionally, we found that condition has a significant effect on the stimulus type (F(1,38)=3.25, p=0.03, η_p^2=0.08), and the memory load (F(1,38)=3.07, p<0.01, η_p^2=0.08). However, the load tested did not have a significant interaction with the stimulus type presented (F(1,38)=1.61, p=0.21, η_p^2=0.04). We also did not find evidence for a significant three way stimulus x condition x load (F(1,38)=1.63, p=0.14, η_p^2=0.04) interaction. Results with capacity estimates revealed significant main effect of load (F(1,38)=6.10, p<0.01, η_p^2=0.14), and condition (F(1,38)=4.83, p<0.01, η_p^2=0.11), but not of stimulus (F(1,38)=1.01, p=0.32, η_p^2=0.03). We found that the load tested has a significant interaction with the condition (F(1,38)=2.16, p=0.05, η_p^2=0.05), but not with the stimulus type (F(1,38)=1.02, p=0.37, η_p^2=0.03). Furthermore, we did not find a significant relationship between the stimulus type presented and the condition tested (F(1,38)=1.67, p=0.18, η_p^2=0.04). Lastly, we found a marginally significant three-way stimulus x condition x load (F(1,38)=1.98, p=0.07, η_p^2=0.05) interaction.

Discussion

In this study, we investigated the impact of simple and complex stimuli on working memory within the same group of participants. Overall, our results demonstrate that multisensory auditory and visual working memory differ when tested on simple versus complex stimuli. When presented with simple stimuli, our results demonstrated a robust enhancement for visual, but not auditory working memory. This finding implies that while auditory information aids visual working memory, visual information does not aid auditory working memory. When presented with complex stimuli, we do not find significant effects of multisensory enhancement in either the visual or auditory working memory domain. However, we do find significant overall difference in visual and auditory working memory performance across the unisensory and multisensory conditions. Furthermore, we did not find strong effects of congruency when tested with simple or complex stimuli.

To directly contrast the results from the two stimulus types (simple and complex), we conducted an omnibus repeated measures ANOVA. We first conducted this analysis when participants were presented with unisensory (Conditions: A and V) stimuli compared to multisensory stimuli (Conditions: AV-Test A and AV-Test V). In this analysis, we only found a significant effect of stimulus for accuracies, but not reaction times or capacity estimates. Furthermore, we did not find significant three-way stimulus x condition x load interactions for accuracies (F(1,38)=1.22, p=0.30, η_p^2=0.03), reaction times (F(1,38)=1.09, p=0.37, η_p^2=0.03), or capacity estimates (F(1,38)=1.10, p=0.36, η_p^2=0.03). When testing modality-specific effects for accuracy, a significant main effect of stimulus but a lack of a significant three-way interaction can be explained the statistical results of the two-way interactions. First, a non-significant three-way interaction implies that two of the factors in this analysis do not depend on the third factor. Statistical results from the two-way interactions reveal that the effect of stimulus depends on condition, and effect of condition depends on load, but the effect of the stimulus does not depend on load. Most importantly, this lack of a significant three-way interaction implies that the multisensory advantage across the two stimulus types is present in a consistent manner. While we do find differences in performance across visual and auditory working memory, it could be that enhancements are comparable when presented with multisensory stimuli in both of these domains.

The experimental design and stimuli in this study were the same as those used in Experiment 2 of Chapter 2 with simple stimuli and Experiment 2 of Chapter 3 with complex stimuli. We replicated our findings from the previous studies when testing working memory with simple, but not complex stimuli. We found an enhancement in visual but not auditory working memory when testing simple stimuli in both instances. However, we did find an advantage when presented with congruent multisensory stimuli and tested in auditory working memory in Chapter 2, which was absent in the current study. When testing complex stimuli, we found a robust advantage in the auditory working memory domain in the previous study (Chapter 3, Experiment 2), which we did not find in this round of testing. More specifically, this advantage was previously reported in the accuracy domain, but was largely absent in the reaction times and capacity estimates analyses. Furthermore, we did not find the previously noted congruency effects in this round of testing. We had previously found an advantage in for congruent multisensory stimuli in auditory working memory and an advantage for incongruent stimuli in visual working memory.

Replicability may differ when using the two different (simple versus complex) stimulus sets for a few reasons. First, these differences may be due to the more basic associations, with little context, between simple multisensory stimuli, but more semantic, context-rich, associations between complex multisensory stimuli. Simple stimuli such as those used in the current studies and previous studies tend to be fairly artificial in nature and use low-level sensory integration mechanisms (Barutchu et al., 2010; Evans & Whitfield, 1964; Hillock et al., 2011; Hubel & Wiesel, 1962; Stevenson & Wallace, 2013). Second, research has reported results with higher variability when presented with complex stimuli, which are based semantic associations, than when presented with simple stimuli between participants (Laurienti et al., 2004). One reason that that complex stimuli may have more variability in behavioral performance is that they tend to represent multisensory associations, which are naturally present in life. These natural associations are highly dependent on context and are learned throughout an individual's lifetime. For this reason, these stimuli may not only be dependent on semantics, but rather a variety of factors such as an individual's cultural background. Lastly, our set of complex stimuli may have more variability than desired in the familiarly of animals that we selected across our participant set. While we selected the complex stimuli based on the result that participants were able to discriminate and categorize the stimuli selected well above chance performance levels (refer to Chapter 3 Experiment 1), not all participants may have the same level of familiarity between the stimuli. For this reason, moving forward, it would be desirable to test the participant's ability to discriminate between stimuli in the given experimental group.

Chapter 5: Discussion, Conclusion, and Future Directions

The experiments described in this thesis craft a rich description of the impact that multisensory integration has on working memory. Taken together, these studies answer a variety of questions such as: Do multisensory stimuli provide a benefit for working memory processing? Are the auditory and visual modalities linked differently based on stimulus properties? Does an investigation in the same group of participants reveal that simple and complex stimuli affect working memory differently? What are the effects of crossmodal congruency on MSI and working memory? The experiments contained in this dissertation demonstrate that multisensory stimuli benefit working memory faculties. However, these advantages differ depending on the stimuli presented and sensory modality tested (visual or auditory working memory). Furthermore, I found that both attentional demands and the memory load tested mediate the influence of multisensory stimuli on memory. Lastly, I can begin to draw inferences regarding the extent to which the presentation of congruent crossmodal correspondences benefits working memory based on the results from the current experiments with simple and complex stimuli.

In Chapter 1, I reviewed the mechanisms of MSI through a cognitive neuroscience perspective and highlighted the principles of MSI in perception, working memory, and the shared behavioral and neural correlated of MSI and working memory. Chapters 2, 3, and 4 summarize a series of studies examining the impact of audiovisual MSI on working memory using simple (Chapter 2) and complex stimuli (Chapter 3). In Chapter 2, I found that multisensory stimuli enhance auditory working memory, but this enhancement depended on the memory load and attentional demands of the task. In Chapter 3, I found a benefit for auditory but not visual working memory when presented with complex multisensory stimuli. I tested how stimulus complexity impacts MSI and working memory in the same group of participants in

Chapter 4 and found a multisensory advantage visual working memory when presented with simple stimuli, but did not find a multisensory advantage in either visual or auditory working memory when presented with complex stimuli. In this concluding chapter, I will discuss some limitations of the current studies, the main take home points from the work detailed in this thesis, and how this line of work can be expanded by future investigations in the field of mind, brain, and education.

Limitations

While the studies presented in this thesis provide novel insights into the impact of multisensory integration on memory, there are some limitations. First, I piloted the complex stimuli used in this dissertation with other groups participants, but did not test each participant's prior familiarity with the stimuli used for the studies outlined in this thesis. While numerous studies have investigated the existence, nature, and generalizability of crossmodal correspondences (Evans & Treisman, 2010; Gallace & Spence, 2006; Marks, 1987), it has also been noted that there are individual differences in the associations that people make across sensory modalities (Greenwald, McGhee, & Schwartz, 1998; Marks, 2011). For example, Marks (1974) found that half of the participants matched loud sounds to a dark gray color, while the other half matched loud sounds to a light gray color. These individual differences may exist due to the fact that crossmodal correspondences are made on a variety of different dimensions and are statistically, structurally, and semantically mediated. For this reason, it is important to understand the degree to which each individual in the participant group is attuned to the correspondences being tested.

Second, while I found that participants performed with higher working memory accuracy and capacity when presented with multisensory compared with unisensory stimuli, they were not at ceiling performance in all the tested conditions (V, A, AV-TestV, AV-TestA). Thus, it is unclear what the upper bound for memory capacity is, where this capacity starts to decline for sensory stimuli. To address this issue, future studies testing memory capacity of multisensory stimuli should use higher memory loads than ones that I used here. Using a higher item loads would help determine what the memory capacity is for these stimuli and where it starts to decline.

Third, the groups of participants that I tested was relatively homogeneous. While this is a positive for reducing variability in the sample, testing a wider array of participants would provide a more holistic view of the impact of multisensory processing on working memory. The current sample consisted of neurotypical UW-Madison undergraduate students (mean= 20.19 years), who were taking an educational psychology course and participated in these studies as part of their class requirement. Future studies should expand these investigations to test participants in a similar age range from other universities across the country or world, along with participants in different age ranges, and special populations. Furthermore, the current sample consisted mostly of participants who self-identified as female. Even though the ratios of male to female participants in these studies largely matched the sex ratios in educational psychology courses, moving forward, it would be important to take the ratio of males and females into account if that is a variable of interest in the study at hand. Additionally, studies in special populations have found sex differences in multisensory processing (Ross, Del Bene, Molholm, Frey, & Foxe, 2015). Understanding how MSI processes have gone awry in special populations could reveal additional mechanisms for typical crossmodal processing. Additionally, the current literature makes the assumption that multisensory stimuli are an asset to sensory processing. Gaining a perspective about when reliance on multisensory cues hinders decision making or leads to maladaptive choices could also shed light on the mechanisms of MSI. Lastly, including a battery of standardized assessments of cognitive skills may be necessary to better understand how individual differences in broad cognitive skills mediate the impact of MSI on WM, specifically when studying typically and atypically developing individuals.

Main Conclusions & Theoretical Implications

Considerations for Attention

Experiments 1 and 2 of Chapter found evidence for a mediating role of attention on the impact of MSI on WM. In Experiment 1, I analyzed participants' working memory performance without cueing the tested modality in the multisensory condition. I did not find evidence for multisensory enhancement in the auditory or visual domain in this experiment. In Experiment 2, I cued participants to which sensory modality would be tested in the multisensory condition, while keeping all other parameters identical to those in Experiment 1. Unlike Experiment 1, Experiment 2 demonstrated a significant multisensory advantage in for auditory and visual working memory stores. As detailed in the discussion portion of Chapter 2, the discrepancies between these findings can be explained by the influence of attention on multisensory integration (Alsius et al., 2014; Koelewijn et al., 2010; Matusz et al., 2015; Senkowski et al., 2007). Although the task used for this study uses multisensory stimuli, it is a classic change detection task of working memory. For this reason, I can attribute the differences in findings between Experiments 1 and 2 to the involvement of attention in working memory mechanisms, along with the involvement of attention in MSI as discussed in Chapter 2. More specifically, when participated are cued to the modality they will be tested in for the change detection task, they can allocate their attention to the appropriate sensory dimension.

Many studies have shown that working memory and attentional processes are intertwined due to the fact that both are instrumental in encoding and maintaining information to guide subsequent behavioral output (Awh, Vogel, & Oh, 2006; Cowan, 1998; Treisman, 1969). One view linking these processes suggests that information held in working memory is simply in the focus of attention in the long-term memory store, thus making it "active" (Cowan, 1998). One metaphor for thinking about attentional mechanisms is that of a spotlight, especially in the visual domain. This spotlight highlights spatial locations, objects, or features of those objects that should be prioritized for processing (Posner, 1980). Interestingly, this spotlight can be allocated to a spatial location in a visual field covertly, without making an eye movement (Chen & Choi, 1998). Recent studies have extended this line of thinking to suggest that this attentional spotlight can impact internally held representations as well (Kiyonaga & Egner, 2013). Emrich, Lockhart, & Al-Aidroos (2017) found that the allocation of attention determines the internal resources that are used for the storage and selection of information in working memory. They also noted that these resources are stronger predictors of working memory performance than the memory load used, suggesting a critical role for attention in the selection/readout of information.

Neural evidence linking attention and working memory comes from a study in which participants were instructed to either shift their focus of attention or retain it at the same location while maintaining an internal memory representation (Munneke, Belopolsky, & Theeuwes, 2012). This study had three key findings: First, early visual areas (e.g. V1) were involved in shifting attention within memory representations. Second, BOLD activation changed when the participants were instructed to shift their focus of attention in an analogous manner. Lastly, BOLD activation increased when participants were instructed to retain their focus of attention in the same location. Furthermore, this relationship is not altered by the degree to which the internal representation and the external task are consistent or inconsistent (Downing, 2000; Olivers, Meijer, & Theeuwes, 2006). Taken together, emerging evidence suggests that attention and working memory mechanisms may have a bidirectional relationship such that they are constantly impacting each other due to their goals and shared cognitive resources (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Kiyonaga & Egner, 2013). Thus, attention and working memory are linked because the focus of attention can be shifted to focus on either internal or external working memory representations.

Studies have also shown that working memory and attention share neural correlates in the sensory cortices (Emrich et al., 2013; Luck, Chelazzi, Hillyard, & Desimone, 2017; Woldorff et al., 1993). In an early study using event-related electrical potentials (ERPs), Woldorff and colleagues found activation in the auditory cortex on the supratemporal plane when attending to tones in one ear, but ignoring them in the other (Woldorff et al., 1993). Another study found evidence of top-down attention in the visual cortex (areas V1, V2, and V4) of the macaque monkey when presented with visual stimuli sequentially in spatially separated locations (Ciaramitaro, Buračas, & Boynton, 2007; Luck et al., 2017). Additionally, studies have also noted the involvement of the inferior temporal (IT) cortex (Chelazzi, Duncan, Miller, & Desimone, 1998), the middle temporal visual area (MT), and the medial superior temporal area (MST) (Recanzone, Wurtz, & Schwarz, 1997) when attending to stimuli in a single neuron’s receptive field. Lastly, attention has been found to facilitate the transfer of information between two regions of the brain. One possible mechanism for this transfer is the synchronized spiking activity between neurons encoding the attended information (Salinas & Sejnowski, 2001; Tiesinga, Fellous, & Sejnowski, 2008). Future studies should more explicitly investigate the involvement of attention in MSI and working memory due to the intertwined nature of the behavioral and neural correlates of these processes.

Differential Effects for Visual and Auditory Working Memory

The most important results in this dissertation reveal that visual and auditory working memory are enhanced when participants are presented with bimodal stimuli. However, I also found that participants overall have higher working memory performance in the visual domain compared to the auditory domain, even in the unimodal conditions. These differences could exist for several reasons. First, there could be a difference in familiarity between the auditory and visual stimuli used in these studies. For example, a participant may more readily recognize a dark blue color patch or an image of lion, compared with a low tuba sound or a “roar” animal call. While I did test participants’ ability to discriminate between complex animal stimuli (Chapter 3, Experiment 1), their ability to discriminate between visual and auditory stimuli was not explicitly tested. However, participants were able to match animal images with sounds significantly above chance within and across animal categories (Chapter 3, Experiment 1). I did not test the participant’s familiarity with visual stimuli in the simple stimulus set used (used in Chapter 2) as the stimuli selected were color patches and were based on previous studies (Gosavi et al., in prep; Ward et al., 2006). Thus, being more familiar with stimuli in the visual modality could have caused performance to be higher overall in this modality. Second, performance in the auditory modality may be more constrained by perceptual than mnemonic abilities. In particular, studies have demonstrated that the auditory stimuli have relatively lower perceptual resolution than visual stimuli (Storms & Zyda, 2000). As proposed by the Modality Appropriateness Hypothesis, the influence of each sensory modality depends how appropriate that modality is for the given task (Welch & Warren, 1980). It has been found that the auditory modality has higher temporal precision than the visual modality, but lower spatial resolution (Repp & Penel, 2002, 2004). Given the spatial nature of my change detection task, memory for auditory stimuli maybe less effective due to difficulties in perceiving them.

Finally, it is important to consider non-attention-based memory resources, such as sensory memory. While this bank of memory processing is not a part of Baddeley's model of working memory (Baddeley, 1986), sensory memory is the first bank of memory processing, and can have downstream influences on the processing of visual and auditory working memory. Visual (iconic) and auditory (echoic) sensory memory are classically thought of as transient sensory stores, which decay quickly after stimulus offset (Chow, 1985; Sakitt, 1976; Sergent, Ruff, Barbot, Driver, & Rees, 2011; Sergent et al., 2013; Sperling, 1960, 1963). Previous studies have suggested that sensory memory may lie between perception and memory processes, and has interactions with multisensory processing mechanisms (Gosavi & Hubbard, 2019; Sergent et al., 2011). While it has been suggested that iconic and echoic memory share a sensory memory buffer, they have been found to differ in both capacity and decay duration (Crowder, 1978). Iconic memory has a large capacity and a duration of about 1000 ms while echoic memory has a smaller capacity but has a longer duration of about 4000 ms (Sperling, 1960; Watkins & Watkins, 1980). Thus, differences in iconic and echoic memory capacity may be contributing to the differences in visual and auditory working memory observed in the current experiments.

A debate has existed in the working memory literature surrounding the question: "is information in WM stored in a central or modality-specific storage system?" While some studies have supported the notion of distinct storage systems for sensory information from different modalities (Baddeley, 1986), others have supported the notion of a central storage system (Cowan, 2005). The series of experiments in this thesis are more consistent with the notion of a central storage system rather than a distributed, modality-specific storage system in working memory. The data also revealed that the storage of information in this central system is dependent on modality and complexity of the stimuli being tested. Specifically, I found more robust enhancement for visual working memory (compared to auditory working memory) when participants were presented with simple multisensory stimuli (Chapter 2, Experiments 2 & 3). Conversely, I found significantly more enhancement for auditory working memory compared to visual working memory when presented with complex multisensory stimuli (Chapter 3, Experiments 2 & 3).

This support for a more central memory store is evident from a number of perspectives. First, while I do find greater working memory capacity estimates when participants were presented with multisensory stimuli, these capacity estimates never exceed the approximate four-item limit, even when combined across the tested modalities. A model with a separate, modality-specific, storage system would predict that capacity estimates for audiovisual stimuli should be well above four items. Second, a centralized store for visual and auditory information would allow for the dynamic and flexible interactions between modality-specific representations. Moreover, based on the differential enhancements found, it is possible that the complexity and naturalistic quality of the stimuli presented mediate the interaction between visual and auditory working memory faculties. More specifically, I found that these semantically meaningful visual stimuli aid auditory working memory, while auditory stimuli do not aid visual working memory (Chapter 3). On the contrary, simple auditory stimuli aid visual working memory, while simple visual stimuli do not aid auditory working memory (Chapter 2). Stimuli may be able to produce this differential enhancement in visual and auditory working memory as these representations are able to interact in this central storage. Thus, the stimulus-dependent, and divergent enhancements could have been found between the two sensory stores based on the reasons outlined above. Lastly, I found enhancements in visual working memory in the presence of auditory information, and enhancements in auditory working memory in the presence of visual information. A central store would allow for the dynamic interaction between sensory stimuli from different modalities to create these crossmodal enhancements. On the contrary, a modality-specific storage system would only allow for enhancement in the same sensory modality that the stimuli were presented in (e.g. a benefit for visual working memory when presented with visual stimuli).

While the aforementioned reasons outline the evidence in support of a centralized working memory storage, there could be other mechanistic explanations for how sensory information is stored in working memory. It could be possible that there are sensory specific working memory stores, but that the recruitment of this information from these independent stores is controlled and coordinated by a resource-limited central executive. Neural evidence for this idea comes from the sensory recruitment hypothesis which proposes that stimulus information is stored in sensory regions during working memory tasks, and higher order regions like the PFC may act the central executive (Emrich et al., 2013; Sprague et al., 2014). A theory of MSI proposes that sensory information from multiple modalities is integrated in higher order regions while it is stored independently in sensory regions (Cohen et al., 2004).

Additionally, it is possible that, as participants were reporting some items from working memory, the remainder were fading from memory. As Sperling (1963) described, “more is seen than can be remembered” meaning that there is a capacity to memory and that after the presentation of the stimulus, there is more information available than can be reported. To overcome this hurdle, Sperling developed the partial report paradigm, which presented participants with a stimulus for a very short duration, enough for an iconic memory trace to remain afterwards, and then asked for a cued response on a part of stimulus array (Sperling, 1960). Similarly, it could be possible that participants are losing information from their working memory while reporting the presented stimuli. While previous studies have limited the information held in working memory by guiding the focus of attention, it would be valuable to test this prediction explicitly in a working memory task. The partial report paradigm could be adapted to working memory tasks (Sperling, 1960). Participants could be presented with a stimulus array followed, after a variable delay, by a cue to guide participants to recall a portion of the stimulus array. Comparing performance on this task with performance on a task in which participants are instructed to remember the entire stimulus array would shed light on these questions. Differences in participant performance, namely that participants recall more information when only asked to report a portion of the stimulus array, would indicate that while participants were reporting some items from working memory, the remainder were fading away.

I found that presenting complex audiovisual stimuli improved working memory only when participants were tested in the auditory domain (Chapter 3, Experiments 2 & 3). Moreover, I found that the addition of auditory stimuli decreased visual working memory performance (Chapter 3, Experiments 2 & 3). This result is in line with the principle of inverse effectiveness (Kayser et al., 2007; Stanford, 2005; Stein & Meredith, 1993; Stevenson & James, 2009). This principle states that multisensory stimuli are integrated with the highest success when the salience of the stimuli is weak (Kayser et al., 2007; Stanford, 2005; Stein & Meredith, 1993; Stevenson & James, 2009). Evidence for this principle has primarily come from animal studies, which show that unimodal stimuli which are minimally effective in evoking a response produce the most robust multisensory enhancement (Stein & Meredith, 1993). Relevant to the findings at hand, based on the theory of inverse effectiveness, I hypothesize that the addition of auditory stimuli increased the salience of the audiovisual stimulus above the threshold that it could be effectively integrated to produce a multisensory advantage in visual working memory. Perhaps the salience of the complex stimuli is higher than the simple stimuli due to the fact that they are semantically meaningful and naturalistic, as I found a multisensory advantage in visual working memory in Experiments 2 and 3 of Chapter 2, but not Experiments 2 and 3 of Chapter 3. For this reason, when combined, these audiovisual stimuli surpass the optimal threshold to produce a multisensory advantage much more readily than the less salient simple stimuli.

Effects of Crossmodal Congruency

The current studies make an important contribution to understanding the effects of congruently versus incongruently presented audiovisual stimuli on working memory. Previous studies have found that stimuli presented congruently with crossmodal correspondences led to higher working memory performance than incongruently presented crossmodal correspondences (Brunetti et al., 2017). While I hypothesized that the pattern of results would be consistent with previous literature, I found that participants had higher working memory performance when presented with incongruent audiovisual stimuli in some instances. More specifically, I found the greatest memory advantage when presented with incongruent multisensory complex stimuli at memory loads of four, six, or eight items (Chapter 3, Experiment 3). Furthermore, I found these effects more prominently when presented with complex stimuli in the visual compared with the auditory working memory domain when presented with a memory load of two, four or eight items (Chapter 3, Experiment 2) and a memory load of four, six or eight items (Chapter 3, Experiment 3). Interestingly, I did find prominent enhancements when participants were presented with congruent stimuli in the auditory modality.

I hypothesize that incongruent stimuli may be producing memory benefits as they are easier to unbind or disentangle than congruent stimuli. It might be easier to create these integrated representations for congruent stimuli as they are inherently bound together based on the principles of crossmodal correspondences (Spence, 2011). Previous research has found that stimuli are "tagged" neurally based on the context of their presentation within 100 ms after stimulus onset (Thelen, Cappe, & Murray, 2012). In this way, the brain can categorize an object based on its representation, relevance to the task, and effectiveness of the multisensory stimulus. Matusz and colleagues (2017) found that task-irrelevant stimuli that were presented crossmodally led to memory enhancements in an object discrimination task. Combining these two pieces of evidence, I postulate that participants had enhanced working memory performance for incongruent stimuli for two reasons: First, the brain was able to tag these as mismatching stimuli and therefore was able to determine that they irrelevant to the task at hand. Second, in line with previous research, perhaps the participants were able to represent the stimuli separately due to their incongruency and thus unbind the stimuli. For this reason, the effective memory load may be greater for congruent stimuli. In contrast, when presented with incongruent stimuli participants might be able to separate out the appropriate information in the to-be-remembered sensory modality, and thereby lower their effective memory load.

This proposal could be aligned with either the central storage or the modality-specific storage models. A central working memory storage model would allow for the dynamic and flexible interaction of sensory information in order to unbind incongruent stimuli, as seen here. However, the modality-specific storage account could also account for the present findings if information is recruited from these stores by a resource-limited central executive. Once the stimulus information is encoded in the separate modality-specific stores, and is neurally tagged based on its relevance, the central executive could differentially recruit information from these stores based on the effectiveness of the sensory stimulus for the modality being tested (e.g. auditory information when tested in auditory working memory. Additional studies that only test congruent crossmodal correspondences but vary the stimulus presentation on other dimensions (e.g. spatial location) may be useful for testing this hypothesis.

A Link Between Perception and Memory

Studies investigating the perceptual underpinnings of multisensory processing have demonstrated several principles of how MSI affects perceptual processing. First, the spatial and temporal proximity of the stimuli presented are instrumental in determining whether two stimuli will be integrated (Stevenson & Wallace, 2013). Multisensory stimuli are most likely to be integrated if they are presented in a spatially and temporally synchronized manner (Cecere et al., 2016; Gepshtein et al., 2005). Second, the ability to integrate information from multiple senses is also highly dependent on the properties of the stimulus presented. While simple stimuli such as flashes and beeps have formed the basis of many experiments, more recently, studies have started advocating for the use of more naturalistic stimuli. (Barutchu et al., 2010; Beauchamp, 2005). Complex stimuli are beneficial as they are semantically rich and mimic the complexities present in the environment. Lastly, the studies have established that MSI is facilitated by crossmodal correspondences between properties of vision and audition (Spence, 2011). Evidence from the experiments detailed in this thesis suggests that the properties that of MSI that influence perception are the same as those that influence working memory. I presented participants with spatially and temporally synchronized stimuli in all of the experiments and found an advantage for visual and auditory working memory when presented with multisensory stimuli. Additionally, I found that the complexity of stimuli presented and the congruency of crossmodal correspondences mediates the influence of the multisensory stimuli on working memory. Thus, the properties of MSI that have an effect on low-level perceptual processes could be the same as those which influence more conceptual, memory-based processes.

A Dual Coding and Recoding Framework

The experiments conducted in Chapters 2 and 3 showed evidence of a working memory enhancement when participants were presented with multisensory stimuli. In Chapter 2, this enhancement was found in the visual and auditory domains in Experiment 2, but was reduced when tested with a higher memory load in Experiment 3. In Chapter 3, Experiments 2 and 3 found a multisensory enhancement only for auditory working memory processing. Overall, I found that multisensory stimuli enhanced working memory abilities compared to unisensory stimuli. These findings are consistent with the dual coding account of MSI and memory. The dual coding theory (Paivio, 1969) proposes that having two codes (e.g., verbal and visual codes) for information leads to memory enhancements. By this account, the dual coding mechanism produces enhanced memory recall for multisensory stimuli due to the ability to access a greater amount of information about the stimulus via the two sensory traces.

Furthermore, it is possible that this overall advantage for multisensory stimuli in memory is driven not only by this automatic, yet passive, dual code, but rather an active process which involves the recoding of sensory information (Gosavi & Hubbard, in press). This modified version builds on the original theory and accounts for the results in this thesis by explaining *how* a multisensory advantage can be observed across stimulus types. Considering recoding as the active component in the retention of information can begin to explain how the presence of an automatic dual code can lead to a sustained memory advantage. Recoding is a process by which one form of information is translated to another form (Baddeley, 1986; Cruse & Clifton, 1973;

Loftus & Loftus, 1976). Delogu and Raffone (2009) found that participants verbally encoded object pictures, sounds or words, even when they were instructed not to verbally articulate the stimulus information, which led the authors to conclude that participants were able to manipulate the representation of the stimulus during the encoding stage.

In the current studies, participants may be recoding the sensory information that they were tested on in the multisensory condition (visual or auditory) into the other, untested, sensory modality to help them retrieve the correct stimulus representation. For example, a participant may be recoding the visual image of a lion as an auditory representation of lion's roar while being tested on visual information. For this reason, the participant may have two sensory traces to rely on during the testing round, even though he/she was only being tested on one sensory modality. The Articulatory Control Process has been postulated to play a role in the recoding of visual information to auditory information (Baddeley, 1986; Henson, Burgess, & Frith, 2000). The Articulatory Control Process is thought to be a part of the phonological loop as proposed by Baddeley, and is responsible for the rehearsal of phonological information (Baddeley, Lewis, & Vallar, 1984; Longoni, Richardson, & Aiello, 1993). The phonological similarity effect supports the existence of the Articulatory Control Process by demonstrating that stimuli that are similar sounding are more difficult to remember, even when they are presented visually (e.g. letters) (Conrad & Hull, 1964). Support for the idea that the Articulatory Control Process is involved in the recoding of visual information to auditory information comes from the fact that the suppression of articulation removes the phonological similarity effect for visual, but not auditory stimuli, which are directly linked to the phonological store.

To test this theory using the complex set of animal stimuli used in this thesis, I could present participants with stimuli in either the visual or auditory sensory modality (e.g. an image of a lion), but test them using the other sensory modality which was not presented (e.g. a lion "roar"). If participants are indeed recoding the stimuli, they would be able to perform significantly above chance levels in this task. Furthermore, I could present participants with similar (e.g. animals from within a category) or different (e.g. animals from across two categories) visual or auditory stimuli and could instruct participants to suppress the articulation of these stimuli. The absence of a phonological similarity effect for visual stimuli (i.e. participants are able to remember similar visual stimuli just as well as visual stimuli that are different), would support the claim that participants are recoding sensory information.

Furthermore, Delogu & Raffone (2009) concluded that the recoding process may require semantic information from long-term memory to be used in the working memory store. The use of complex stimuli in the current studies would support this idea, as participants would need the semantic knowledge that a lion produces a "roar" sound in order to recode that sound as an image of a lion. Furthermore, based on the result that participants have enhanced auditory working memory performance when presented with complex multisensory stimuli, I postulate that recoding from the auditory to the visual domains may be easier than recoding from the visual to the auditory domain. It is easier generate a picture (e.g., a mental image of a lion) when presented with a sound (e.g., a lion "roar") than vice versa. This may further imply that the visual representations of multisensory audiovisual information are more flexible, and therefore easier to call upon, than the auditory representations.

Future Directions

The evidence from this thesis has implications both for basic cognitive psychology research and for more applied work in the educational setting. Moving forward, research should investigate the involvement of multisensory integration mechanisms in working memory and attention at the behavioral and neural levels. To do this, studies can vary the working memory load presented to participants to understand the capacity of working memory capacity and where it starts to decline. Furthermore, studies can explicitly test the centralized versus sensory-modality specific theories of working memory and understand how the focus of attention affects MSI and working memory. At a neural level, studies can investigate how patterns of brain activity diverge for unisensory and multisensory stimuli in the sensory cortices and the PFC during working memory tasks. More specifically, researchers can understand the nature of this activity by studying neural oscillations as previous studies have noted their importance for working memory, and multisensory processing (Bonnefond, Kastner, & Jensen, 2017; Roux & Uhlhaas, 2014).

Moving forward, the impact of MSI on cognition and education needs to be further investigated. One avenue to address this is by investigating MSI abilities in school-aged children as they are actively learning new concepts that use multisensory processing (Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009; Matusz et al., 2015). Another method is to investigate the relationship between MSI cognitive skills such as attention (Cowan, Li, Glass, & Scott Saults, 2017). Particularly, future investigations need to address questions such as: how does multisensory training in students affect perceptual, and memory mechanisms? Do the types of associations that are used for training have an effect on learning? Furthermore, is the degree of MSI in pedagogy related to a sensory representation, and does it change based on the change in the representation?

Another way to understand the role of MSI in education is to build a complete trajectory of its development. Numerous studies detail the emergence of MSI abilities in neonates and into infancy and the toddler years (Bahrick, 1983; Bahrick, Lickliter, Castellanos, Todd, & Bahrick, 2015; Walton & Bower, 1993). Another branch of studies discusses the intricacies of MSI in the adolescent years (Hillock-Dunn & Wallace, 2012) and into adulthood (Barutchu et al., 2010). Moving forward, it is crucial to analyze multisensory integration in children who are five to fifteen years old. Children in this age group are in their early school and middle school years, which is a time when they begin learning many new concepts. This cognitive engagement is instrumental in shaping their cognitive and neural circuits. Additionally, investigating children will help us further understand the bidirectional relationship between education on multisensory processing (Shams & Seitz, 2008). Crafting a richer developmental trajectory of MSI will provide insight into the innate versus learned nature of key stimuli involved, and this process as a whole. Lastly, to have a day-to-day impact on the student, teachers can employ multisensory learning methods which involve engaging two or more sensory modalities in one activity (Birsh, 2009; Vickery, Reynolds, & Cochran, 1987). These activities can be more engaging to the students, can help them make new connections within the content, and can help children with learning difficulties involving the integration of auditory (e.g. speech sound) and visual (e.g. letters) information, such children with dyslexia (Blau et al., 2009).

The studies detailed in this thesis investigate the impact of multisensory integration on working memory performance. To do this, I used a variety of stimuli, crossmodal correspondence, and memory loads. Together, this work extends the current literature by demonstrating that multisensory, audiovisual, stimuli enhance working memory accuracy, and capacity, and that this enhancement in visual and auditory working memory is dependent on the type of stimulus presented. Furthering our understanding of how MSI affects cognition, from perception to memory and learning, is an important piece of the puzzle to build a complete understanding of this process. Evolutionarily, the role of multisensory processing has been to combine relevant sensory input from the environment to create a representation to appropriately allocate attention, retrieve a memory, or make a decision. Investigating the relationship between MSI and memory will help us further understand the mechanistic underpinnings of this process and will allow us to leverage this information to impact educational practices. Research investigating multisensory processing and memory has gained substantial momentum in the last two decades. As more information is uncovered, the future of investigations aiming to understand the behavioral and neural correlates of multisensory integration promises to be an exciting one.

References

Alsius, A., Möttönen, R., Sams, M. E., Soto-Faraco, S., & Tiippana, K. (2014). Effect of attentional load on audiovisual speech perception: Evidence from ERPs. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2014.00727

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*. https://doi.org/10.1016/j.cub.2005.03.046

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*. https://doi.org/10.1007/s00221-007-1110-1

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2005.08.023

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation - Advances in Research and Theory*. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. D., & Hitch, G. J. (2000). Development of Working Memory: Should the Pascual-Leone and the Baddeley and Hitch Models Be Merged? *Journal of Experimental Child Psychology*. https://doi.org/10.1006/jecp.2000.2592

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the Articulatory Loop. *The Quarterly Journal of Experimental Psychology Section A*. https://doi.org/10.1080/14640748408402157

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development*, *6*(4), 429–451. https://doi.org/10.1016/S0163-6383(83)90241-2

Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology*, *41*(4), 352–363. https://doi.org/10.1002/dev.10049

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*(2), 190–201. https://doi.org/10.1037//0012-1649.36.2

Bahrick, L. E., & Lickliter, R. (2003). Intersensory redundancy guides early perceptual and cognitive development. *Advances in Child Development and Behavior*, *30*(C), 153–187. https://doi.org/10.1016/S0065-2407(02)80041-6

Bahrick, L. E., Lickliter, R., Castellanos, I., Todd, J. T., & Bahrick, L. (2015). Intrasensory redundancy facilitates infant detection of tempo: Extending predictions of the intersensory redundancy hypothesis. *Infancy*, 1–28. https://doi.org/10.1111/infa.12081

Bahrick, L. E., Lickliter, R., Castellanos, I., & Vaillant-Molina, M. (2010). Increasing task difficulty enhances effects of intersensory redundancy: Testing a new prediction of the Intersensory Redundancy Hypothesis. *Developmental Science*, *13*(5), 731–737.

https://doi.org/10.1111/j.1467-7687.2009.00928.x

Bahrick, L. E., & Pickens, J. N. (1994). Amodal relations: The basis for intermodal perception and learning. In *The development of intersensory perception: Comparative perspectives* (pp. 204–233).

Bao, M., Li, Z. H., & Zhang, D. R. (2007). Binding Facilitates Attention Switching Within Working Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.33.5.959

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology*, *105*(1–2), 38–50. https://doi.org/10.1016/j.jecp.2009.08.005

Beauchamp, M., Nath, A. R., & Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the mcGurk effect. *Journal of Neuroscience*, *30*(7), 2414–2417. https://doi.org/10.1523/JNEUROSCI.4865-09.2010

Beauchamp, M. S. (2005). See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*. https://doi.org/10.1016/j.conb.2005.03.011

Beauchamp, M. S., Lee, K., Argall, B., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*, 809–823. https://doi.org/10.1016/S0896-6273(04)00070-4

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482–489. https://doi.org/10.3758/BF03208826

Bertelson, P., & Radeau, M. (1981). Erratum to: Cross-modal bias and perceptual fusion with auditoryvisual spatial discordance. *Perception & Psychophysics*, *30*(3), 227–227. https://doi.org/10.3758/BF03214277

Birsh, J. R. (2009). Multisensory teaching of basic language skills. *Exceptional Parent*, *39*(1), 143. Retrieved from http://search.proquest.com/docview/62365191?accountid=142023%5Cnhttp://bd4uz2kj6y.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/ERIC&rft_val_fmt=info:ofi/fmt:kev:mtx:book&rft.genre=unknown&rft.jtitle=&rft.atit

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced Neural Integration of Letters and Speech Sounds Links Phonological and Reading Deficits in Adult Dyslexia. *Current Biology*. https://doi.org/10.1016/j.cub.2009.01.065

Boltz, M. G. (2011). Illusory Tempo Changes Due to Musical Characteristics. *Music Perception: An Interdisciplinary Journal*. https://doi.org/10.1525/mp.2011.28.4.367

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H. J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology*, *17*(19), 1697–1703. https://doi.org/10.1016/j.cub.2007.08.050

Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics*, *6*(6), 337–339. https://doi.org/10.3758/BF03212787

Bonnefond, M., Kastner, S., & Jensen, O. (2017). Communication between Brain Areas Based on Nested Oscillations. *Eneuro*. https://doi.org/10.1523/ENEURO.0153-16.2017

Botta, F., Santangelo, V., Raffone, A., Sanabria, D., Lupiáñez, J., & Belardinelli, M. O. (2011). Multisensory integration affects visuo-spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1099–1109. https://doi.org/10.1037/a0023513

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*. https://doi.org/10.1006/nimg.1996.0247

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K. P., … Fink, G. R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron*, *29*(1), 287–296. https://doi.org/10.1016/S0896-6273(01)00198-2

Brunel, L., Carvalho, P. F., & Goldstone, R. L. (2015). It does belong together: Cross-modal correspondences influence cross-modal integration during perceptual learning. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.00358

Brunetti, R., Indraccolo, A., Mastroberardino, S., Spence, C., & Santangelo, V. (2017). The impact of cross-modal correspondences on working memory performance. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000348

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn2459

Calvert, G. A. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596. https://doi.org/10.1126/science.276.5312.593

Calvert, G., Spence, C., & Stein, B. E. (2004). The handbook of multisensory processes. *The Handbook of Multisensory Processes*, 933. https://doi.org/nicht verfügbar?

Campo, P., Maestú, F., Capilla, A., Morales, M., Fernández, S., Río, D. del, & Ortiz, T. (2008). Temporal Dynamics of Parietal Activity During Word-Location Binding. *Neuropsychology*. https://doi.org/10.1037/0894-4105.22.1.85

Campo, P., Poch, C., Parmentier, F. B. R., Moratti, S., Elsley, J. V., Castellanos, N. P., … Maestú, F. (2010). Oscillatory activity in prefrontal and posterior regions during implicit letter-location binding. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2009.10.024

Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience*, *22*(11), 2886–2902. https://doi.org/10.1111/j.1460-9568.2005.04462.x

Cecere, R., Gross, J., & Thut, G. (2016). Behavioural evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality. *European Journal of Neuroscience*, *43*(12), 1561–1568. https://doi.org/10.1111/ejn.13242

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1998.80.6.2918

Chen, K., & Choi, H. J. (1998). Visual attention and eye movements. *Attention*. https://doi.org/S1415-790X2013000400953 [pii]

Chow, S. L. (1985). Iconic store and partial report. *Memory & Cognition*, *13*(3), 256–264. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1865196

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the Contents of Visual Short-Term Memory from Human Visual and Parietal Cortex. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.0184-12.2012

Ciaramitaro, V. M., Buračas, G. T., & Boynton, G. M. (2007). Spatial and Cross-Modal Attention Alter Responses to Unattended Sensory Information in Early Visual and Auditory Human Cortex. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00580.2007

Cichy, R. M., & Teng, S. (2017). Resolving the neural dynamics of visual and auditory scene processing in the human brain: a methodological approach. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160108. https://doi.org/10.1098/rstb.2016.0108

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory and Cognition*. https://doi.org/10.3758/BF03194326

Cohen, Y. E., Cohen, I. S., & Gifford, G. W. (2004). Modulation of LIP activity by predictive auditory and visual cues. *Cerebral Cortex*, *14*(12), 1287–1301. https://doi.org/10.1093/cercor/bhh090

Collier, W. G., & Hubbard, T. L. (2001). Judgments of happiness, brightness, speed and tempo change of auditory stimuli varying in pitch and tempo. *Psychomusicology: A Journal of Research in Music Cognition*. https://doi.org/10.1037/h0094060

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology (London, England : 1953)*. https://doi.org/10.1111/j.2044-8295.1964.tb00928.x

Cowan, N. (1998). *Attention and Memory: An Integrated Framework. Attention and Memory: An Integrated Framework*. https://doi.org/10.1093/acprof:oso/9780195119107.001.0001

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences.*, *24*(1), 154–176. https://doi.org/10.1017/S0140525X0161392X

Cowan, N. (2005). *Working Memory Capacity. Working Memory Capacity*. https://doi.org/10.4324/9780203342398

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics*. https://doi.org/10.3758/BF03203031

Cowan, N., Li, Y., Glass, B. A., & Scott Saults, J. (2017). Development of the ability to combine visual and acoustic information in working memory. *Developmental Science*. https://doi.org/10.1111/desc.12635

Crowder, R. (1978). Sensory Memory Systems. In *Percetual Coding* (pp. 343–373). https://doi.org/https://doi.org/10.1016/B978-0-12-161908-4.50016-4

Cruse, D., & Clifton, C. (1973). Recoding strategies and the retrieval of information from memory. *Cognitive Psychology*, *4*(2), 157–193. https://doi.org/10.1016/0010-0285(73)90010-8

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

de Nó, L. (1933). Studies on the structure of the cerebral cortex. *Trab. Lab. Invest. Biol. Univ. Madr*, *20*, 41–78.

Delogu, F., & Raffone, A. (2009). Semantic encoding in working memory : Is there a (multi) modality effect ? *Memory*, *17*(6), 655–664. https://doi.org/10.1080/09658210902998054

Deroy, O., & Spence, C. (2013). Why we are not all synesthetes (not even weakly so). *Psychonomic Bulletin & Review*, *20*(4), 643–664. https://doi.org/10.3758/s13423-013-0387-2

Dionne, J. K., Meehan, S. K., Legon, W., & Staines, W. R. (2010). Crossmodal influences in somatosensory cortex: Interaction of vision and touch. *Human Brain Mapping*. https://doi.org/10.1002/hbm.20841

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*. https://doi.org/10.1111/1467-9280.00290

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*. https://doi.org/10.1016/j.cognition.2009.10.013

Emrich, S. M., Lockhart, H. A., & Al-Aidroos, N. (2017). Attention mediates the flexible allocation of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000398

Emrich, S. M., Riggall, A. C., LaRocque, J. J., & Postle, B. R. (2013). Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.5732-12.2013

Erickson, L. C., Heeg, E., Rauschecker, J. P., & Turkeltaub, P. E. (2014). An ALE meta-analysis on the audiovisual integration of speech signals. *Human Brain Mapping*, *35*(11), 5587–5605. https://doi.org/10.1002/hbm.22572

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*. https://doi.org/10.1167/7.5.7

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn_a_00357

Ettlinger, G. (1961). Learning in two sense-modalities. *Nature*, *191*, 308.

Evans, E. F., & Whitfield, I. C. (1964). Classifcation of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat. *The Journal of Physiology*, *171*, 476–493. https://doi.org/10.1113/jphysiol.1964.sp007391

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*. https://doi.org/10.1167/10.1.6

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, *8*(12), 1643–1646. https://doi.org/10.1038/nn1608

Finney, E. M., Fine, I., & Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nature Neuroscience*, *4*(12), 1171–1173. https://doi.org/10.1038/nn763

Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2006.01739.x

Fougnie, D., & Marois, R. (2011). What Limits Working Memory Capacity? Evidence for Modality-Specific Sources to the Simultaneous Storage of Visual and Auditory Arrays. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/a0024834

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., … Murray, M. M. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.2002.88.1.540

Funahashi, S., Inoue, M., & Kubota, K. (1993). Delay-related activity in the primate prefrontal cortex during sequential reaching tasks with delay. *Neuroscience Research*. https://doi.org/10.1016/0168-0102(93)90019-M

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*. https://doi.org/10.1126/science.173.3997.652

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception and Psychophysics*. https://doi.org/10.3758/BF03193720

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *The American Journal of Psychology*, *72*, 521–529. https://doi.org/10.2307/1419493

Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision*, *5*(11), 1013–1023. https://doi.org/10.1167/5.11.7

Ghazanfar, A. A., & Logothetis, N. K. (2003). Facial expressions linked to monkey calls. *Nature*. https://doi.org/10.1038/423937a

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2006.04.008

Gnadt, J. W., & Andersen, R. A. (1988). Memory related motor planning activity in posterior parietal cortex of macaque. *Experimental Brain Research*. https://doi.org/10.1007/BF00271862

Goldman-Rakic, P. S. (1987). Circuitry of the frontal association cortex and its relevance to dementia. *Archives of Gerontology and Geriatrics*. https://doi.org/10.1016/0167-4943(87)90029-X

Goolkasian, P., & Foos, P. W. (2005). Bimodal format effects in working memory. *American Journal of Psychology*.

Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time. *Frontiers in Integrative Neuroscience*, *6*. https://doi.org/10.3389/fnint.2012.00077

Gosavi, R. S., Bade, R., & Hubbard, E. M. (n.d.). Timbre Impacts the Consistency of Music-Color Synesthesia.

Gosavi, R. S., & Hubbard, E. M. (n.d.). How Synesthesia May Lead to Enhanced Memory. In K. Sathian & V. S. Ramachandran (Eds.), *Multisensory Perception: From Laboratory to Clinic*. Academic Press/Elsevier.

Gosavi, R. S., & Hubbard, E. M. (2019). A Color Advantage in Iconic Memory. *Cognition*, *187*, 32–37. https://doi.org/https://doi.org/10.1016/j.cognition.2019.02.009

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory–visual integration in nonsense syllables and sentences. *The Journal of the Acoustical Society of America*, *104*(4), 2438–2450. https://doi.org/10.1121/1.423751

Grantyn, A., & Grantyn, R. (1982). Axonal patterns and sites of termination of cat superior colliculus neurons projecting in the tecto-bulbo-spinal tract. *Experimental Brain Research*, *46*(2), 243–256. https://doi.org/10.1007/BF00237182

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.74.6.1464

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*. https://doi.org/10.1038/nature07832

Hartline, H. K. (1940). The receptive fields of optic nerve fibers. *Am J. Physiol.*, *130*, 690–699.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2012.03.016

Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—it's my area: Or is it? *Journal of Cognitive Neuroscience*, *20*(12), 2125–2136. https://doi.org/10.1162/jocn.2008.20148

Henschke, J. U., Oelschlegel, A. M., Angenstein, F., Ohl, F. W., Goldschmidt, J., Kanold, P. O., & Budinger, E. (2017). Early sensory experience influences the development of multisensory thalamocortical and intracortical connections of primary sensory cortices. *Brain Structure and Function*, pp. 1–26. https://doi.org/10.1007/s00429-017-1549-1

Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: An fMRI study. *Neuropsychologia*. https://doi.org/10.1016/S0028-3932(99)00098-6

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*, *15*(5), 688–696. https://doi.org/10.1111/j.1467-7687.2012.01171.x

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia*, *49*(3), 461–467. https://doi.org/10.1016/j.neuropsychologia.2010.11.041

Hirsh, I. J., & Sherrick Jr., C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, *62*(5), 423–432. https://doi.org/10.1037/h0045283

Hocking, J., & Price, C. J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing. *Cerebral Cortex*, *18*(10), 2439–2449. https://doi.org/10.1093/cercor/bhn007

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Ishai, A., & Sagi, D. (1997). Visual imagery facilitates visual perception: Psychophysical evidence. *Journal of Cognitive Neuroscience*, *9*(4), 476–489.

Izumi, A., & Kojima, S. (2004). Matching vocalizations to vocalizing faces in a chimpanzee (Pan troglodytes). *Animal Cognition*. https://doi.org/10.1007/s10071-004-0212-4

James, T. W., VanDerKlok, R. M., Stevenson, R. A., & James, K. H. (2011). Multisensory perception of action in posterior temporal and parietal cortices. *Neuropsychologia*, *49*(1), 108–114. https://doi.org/10.1016/j.neuropsychologia.2010.10.030

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of Visual Short-Term Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.26.3.683

Jones, J. a, & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, *14*(8), 1129–1133. https://doi.org/10.1097/01.wnr.0000074343.81633.2a

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196323

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structure and Function*. https://doi.org/10.1007/s00429-007-0154-0

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. *Journal of Neuroscience*, *27*(8), 1824–1835. https://doi.org/10.1523/JNEUROSCI.4737-06.2007

Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0001532

King, A. J. (2004). The superior colliculus. *Current Biology : CB*, *14*(9), R335–R338. https://doi.org/10.1016/j.cub.2004.04.018

Kiyonaga, A., & Egner, T. (2013). Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-012-0359-y

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

Krauzlis, R. J., Lovejoy, L. P., & Zénon, A. (2013). Superior colliculus and visual spatial attention. *Annual Review of Neuroscience*, *36*(1), 165–182. https://doi.org/10.1146/annurev-neuro-062012-170249

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*. https://doi.org/10.1016/j.neuron.2008.10.043

Kubota, K. (2003). One approach to understanding the function of the frontal pole. *International Congress Series*. https://doi.org/10.1016/S0531-5131(03)00974-9

Kubota, K., & Niki, H. (2017). Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1971.34.3.337

Kujala, T., Huotilainen, M., Sinkkonen, J., Ahonen, A. I., Alho, K., Hämälä:inen, M. S., … Näätänen, R. (1995). Visual cortex activation in blind humans during sound discrimination. *Neuroscience Letters*, *183*(1–2), 143–146. https://doi.org/10.1016/0304-3940(94)11135-6

Lacey, S., Martinez, M., McCormick, K., & Sathian, K. (2016). Synesthesia strengthens sound-symbolic cross-modal correspondences. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13381

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*. https://doi.org/10.1007/s00221-004-1913-2

Lehnert, G., & Zimmer, H. D. (2006). Auditory and visual spatial working memory. *Memory and Cognition*. https://doi.org/10.3758/BF03193254

Leibenluft, E., Saad, Z. S., Cox, R. W., Adleman, N. E., & Chen, G. (2014). Applications of multivariate modeling to neuroimaging group analysis: A comprehensive alternative to univariate general linear model. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2014.06.027

Lewis-Peacock, J. A., & Norman, K. A. (2014). *Multi-voxel pattern analysis of fMRI data. The Cognitive Neurosciences*.

Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology Human Perception & Performance*, *22*(5), 1094–1106. https://doi.org/10.1037/0096-1523.22.5.1094

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Developmental Psychology*, *16*(6), 597–607. https://doi.org/10.1037/0012-1649.16.6.597

Li, D., Cowan, N., & Saults, J. S. (2012). Estimating working memory capacity for lists of nonverbal sounds. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-012-0383-z

Linderholm, T., Cong, X., & Zhao, Q. (2008). Differences in low and high working-memory capacity readers' cognitive and metacognitive processing patterns as a function of reading for different purposes. *Reading Psychology*. https://doi.org/10.1080/02702710701568587

Loftus, G. R., & Loftus, E. F. (1976). Human memory: The processing of information. *Psychology Press*.

Longoni, A. M., Richardson, J. T. E., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory & Cognition*. https://doi.org/10.3758/BF03211160

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (2017). Neural Mechanisms of Spatial Selective Attention in Areas V1, V2, and V4 of Macaque Visual Cortex. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1997.77.1.24

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–284. https://doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.06.006

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1112605108

Magee, J. C. (2000). Dendritic integration of excitatory synaptic input. *Nature Reviews Neuroscience*. https://doi.org/10.1038/35044552

Makovac, E., Kwok, S. C., & Gerbino, W. (2014). Attentional cueing by cross-modal congruency produces both facilitation and inhibition on short-term visual recognition. *Acta Psychologica*. https://doi.org/10.1016/j.actpsy.2014.07.008

Marks, L. E. (1974). On associations of light and sound: the mediation of brightness, pitch, and loudness. *The American Journal of Psychology*, *87*(1–2), 173–188. https://doi.org/10.2307/1422011

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384–394. https://doi.org/10.1037/0096-1523.13.3.384

Marks, L. E. (2011). For Hedgehogs and Foxes: Individual Differences in the Perception of Cross-Modal Similarity. In *Psychophysics in Action*. https://doi.org/10.1007/978-3-642-74382-5_5

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*. https://doi.org/10.1068/p2866

Martino, G., & Marks, L. E. (2001). Synesthesia : strong and Weak. *Current Directions in Psychological Science*, *10*(2), 61–65. https://doi.org/https://doi.org/10.1111/1467-8721.00116

Matusz, P. J., Broadbent, H., Ferrari, J., Forrest, B., Merkley, R., & Scerif, G. (2015). Multi-modal distraction: Insights from children's limited attention. *Cognition*, *136*, 156–165. https://doi.org/https://doi.org/10.1016/j.cognition.2014.11.031

Matusz, P. J., Wallace, M. T., & Murray, M. M. (2017). A multisensory perspective on object memory. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2017.04.008

Maurer, D., & Mondloch, C. J. (2005). Neonatal synesthesia: A reevaluation. *Synesthesia: Perspectives from Cognitive Neuroscience*. Retrieved from https://login.proxy.library.emory.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2005-09449-010&site=ehost-live

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316–322. https://doi.org/10.1111/j.1467-7687.2006.00495.x

May, P. J. (2005). The mammalian superior colliculus: Laminar structure and connections. *Progress in Brain Research*. https://doi.org/10.1016/S0079-6123(05)51011-2

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

Meier, T. B., Nair, V. A., Meyerand, M. E., Birn, R. M., & Prabhakaran, V. (2014). The neural correlates of age effects on verbal-spatial binding in working memory. *Behavioural Brain Research*. https://doi.org/10.1016/j.bbr.2014.03.005

Melara, R. D. (1989). Dimensional Interaction Between Color and Pitch. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.15.1.69

Miller, E., Li, L., & Desimone, R. (2018). Activity of neurons in anterior inferior temporal cortex during a short- term memory task. *The Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.13-04-01460.1993

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*. https://doi.org/10.1037/h0043158

Miller, L. M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *Journal of Neuroscience*, *25*(25), 5884–5893. https://doi.org/10.1523/JNEUROSCI.0896-05.2005

Milner, B. (1964). Some effects of frontal lobectomay in man. In *The frontal granular cortex and behaviour*.

Mishra, J., & Gazzaley, A. (2012). Attention Distributed across Sensory Modalities Enhances Perceptual Performance. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.0867-12.2012

Molholm, S. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of Neurophysiology*, *96*(2), 721–729. https://doi.org/10.1152/jn.00285.2006

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory Visual-Auditory Object Recognition in Humans: A High-density Electrical Mapping Study. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhh007

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. In *Cognitive, Affective and Behavioral Neuroscience*. https://doi.org/10.3758/CABN.4.2.133

Morrongiello, B. A., Fenwick, K. D., & Nutley, T. (1998). Developmental changes in associations between auditory-visual events. *Infant Behavior and Development*, *21*(4), 613–626. https://doi.org/10.1016/S0163-6383(98)90033-9

Munneke, J., Belopolsky, A. V., & Theeuwes, J. (2012). Shifting attention within memory representations involves early visual areas. *PloS One*. https://doi.org/10.1371/journal.pone.0035528

Neisser, U. (1967). *Cognitive Psychology*. *The American Journal of Psychology* (Vol. 105). https://doi.org/10.2307/1422917

Newell, F., Mamassian, P., & Alais, D. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving*, *23*(1), 3–38. https://doi.org/10.1163/187847510X488603

Newell, F. N., & Mitchell, K. J. (2016). Multisensory integration and cross-modal learning in synaesthesia: A unifying model. *Neuropsychologia*, *88*, 140–150. https://doi.org/10.1016/j.neuropsychologia.2015.07.026

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2006.07.005

Ohala, J. J. (1983). Cross-Language Use of Pitch: An Ethological View. *Phonetica*. https://doi.org/10.1159/000261678

Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I. P., Joensuu, R., Autti, T., & Sams, M. (2005). Processing of audiovisual speech in Broca's area. *NeuroImage*, *25*(2), 333–338. https://doi.org/10.1016/j.neuroimage.2004.12.001

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.32.5.1243

Olson, I. R., & Marshuetz, C. (2005). Remembering "what" brings along "where" in visual working memory. *Perception and Psychophysics*. https://doi.org/10.3758/BF03206483

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*(3), 241–263. https://doi.org/10.1037/h0027272

Paivio, A. (2008). *Mental Representations: A dual coding approach*. *Mental Representations: A Dual Coding Approach*. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2008.11.007

Plant, G. (1999). Perceiving talking faces: from speech perception to a behavioral principle. *Ear and Hearing*. https://doi.org/10.1097/00003446-199904000-00011

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38. https://doi.org/10.1016/j.neuroscience.2005.06.005

Postle, B. R. (2015). The cognitive neuroscience of visual short-term memory. *Current Opinion in Behavioral Sciences*. https://doi.org/10.1016/j.cobeha.2014.08.004

Postle, B. R., & Pasternak, T. (2010). Short term and working memory. In *Encyclopedia of Neuroscience*. https://doi.org/10.1016/B978-008045046-9.00787-7

Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabriel, J. D. E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience*. https://doi.org/10.1038/71156

Pribram, K., Ahumada, A., Hartog, J., & Roos, L. (1964). A progress report on the neurological processes distributed by frontal lesions in primates. In *The frontal granular cortex and behaviour* (pp. 28–55). New York: McGraw-Hill book Company.

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00197

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, *28*(2), 617–625. https://doi.org/10.1016/S0896-6273(00)00138-0

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Frontiers in Evolutionary Neuroscience*, *4*(MAY). https://doi.org/10.3389/fnevo.2012.00007

Recanzone, G. H., Wurtz, R. H., & Schwarz, U. (1997). Responses of MT and MST neurons to one and two moving objects in the receptive field. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1997.78.6.2904

Repp, B. H., & Penel, A. (2002). Auditory Dominance in Temporal Processing: New Evidence from Synchronization with Simultaneous Visual and Auditory Sequences. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.28.5.1085

Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*. https://doi.org/10.1007/s00426-003-0143-8

Riggall, A. C., & Postle, B. R. (2012). The Relationship between Working Memory Storage and Elevated Activity as Measured with Functional Magnetic Resonance Imaging. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.1892-12.2012

Robinson, C. W., & Sloutsky, V. M. (2010). Development of cross-modal processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 135–141. https://doi.org/10.1002/wcs.12

Rock, I., & Victor, J. (1964). Vision and touch: an experimentally created conflict between the two senses. *Science (New York, N.Y.)*, *143*(3606), 594–596. https://doi.org/10.1126/science.143.3606.594

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*. https://doi.org/10.1126/science.143.3606.594

Root, R. T., & Ross, S. (1965). Further validation of subjective scales for loudness and brightness by means of cross-modality matching. *The American Journal of Psychology*, *78*, 285–289. https://doi.org/10.2307/1420502

Ross, L. A., Del Bene, V. A., Molholm, S., Frey, H. P., & Foxe, J. J. (2015). Sex differences in multisensory speech processing in both typically developing children and those on the autism spectrum. *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2015.00185

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhl024

Rothen, N., Bartl, G., Franklin, A., & Ward, J. (2017). Electrophysiological correlates and psychoacoustic characteristics of hearing-motion synaesthesia. *Neuropsychologia*, *106*, 280–288. https://doi.org/10.1016/j.neuropsychologia.2017.08.031

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-011-0055-3

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.10.010

Sakitt, B. (1976). Iconic memory. *Psychological Review*, *83*(4), 257–276. https://doi.org/10.1037/0033-295X.83.4.257

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*. https://doi.org/10.1016/S0022-5371(82)90521-7

Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience*. https://doi.org/10.1038/35086012

Sams, M., Manninen, P., Surakka, V., Helin, P., & Kättö, R. (1998). McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. *Speech Communication*, *26*(1–2), 75–87. https://doi.org/10.1016/S0167-6393(98)00051-X

Saults, J. S., & Cowan, N. (2007). A Central Capacity Limit to the Simultaneous Storage of Visual and Auditory Arrays in Working Memory. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.136.4.663

Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science*, *6*(3), 233–241. https://doi.org/10.1111/1467-7687.00276

Schneegans, S., & Bays, P. M. (2016). No fixed item limit in visuospatial working memory. *Cortex*, *83*, 181–193. https://doi.org/10.1016/j.cortex.2016.07.021

Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology*, *16*(14), 1422–1427. https://doi.org/10.1016/j.cub.2006.05.048

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, *47*(3), 277–287. https://doi.org/10.1016/S0168-0102(03)00214-1

Senkowski, D., Saint-Amour, D., Kelly, S. P., & Foxe, J. J. (2007). Multisensory processing of naturalistic objects in motion: A high-density electrical mapping and source estimation study. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2007.01.053

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Sereno, M. I., & Tootell, R. B. H. (2005). From monkeys to humans: What do we now know about brain homologies? *Current Opinion in Neurobiology*. https://doi.org/10.1016/j.conb.2005.03.014

Sergent, C., Ruff, C. C., Barbot, A., Driver, J., & Rees, G. (2011). Top–Down Modulation of Human Early Visual Cortex after Stimulus Offset Supports Successful Postcued Report. *Journal of Cognitive Neuroscience*, *23*(8), 1921–1934.

https://doi.org/10.1162/jocn.2010.21553

Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., & Tallon-Baudry, C. (2013). Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Current Biology*, *23*(2), 150–155. https://doi.org/10.1016/j.cub.2012.11.047

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*(6814), 788. https://doi.org/10.1038/35048669

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science (New York, N.Y.)*, *145*, 1328–1330. https://doi.org/10.1126/science.145.3638.1328

Simpson, R. H., Quinn, M., & Ausubel, D. P. (1956). Synesthesia in children: Association of colors with pure tone frequencies. *Journal of Genetic Psychology*, *89*(1), 95–103. https://doi.org/10.1080/00221325.1956.10532990

Slater, A., Quinn, P. C., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of arbitrary auditory-visual pairings. *Developmental Science*, *2*(3), 333–338. https://doi.org/10.1111/1467-7687.00079

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, *24*(1), 99–142. https://doi.org/10.1016/0010-0285(92)90004-L

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*. https://doi.org/10.1037/0278-7393.6.2.174

Sparks, D. L., & Jay, M. F. (1986). The functional organization of the primate superior colliculus: A motor perspective. *Progress in Brain Research*, *64*(C), 235–241. https://doi.org/10.1016/S0079-6123(08)63418-4

Spelke, E. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, *15*(6), 626–636. https://doi.org/10.1037/0012-1649.15.6.626

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*(2), 61–70. https://doi.org/10.1250/ast.28.61

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1.

Sperling, G. (1963). A Model for Visual Memory Tasks. *Human Factors*. https://doi.org/10.1177/001872086300500103

Sprague, T. C., Ester, E. F., & Serences, J. T. (2014). Reconstructions of information in visual spatial working memory degrade with memory load. *Current Biology*. https://doi.org/10.1016/j.cub.2014.07.066

Stanford, T. R. (2005). Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.5095-04.2005

Stein, B. E., Jiang, W., & Stanford, T. R. (2004). Multisensory Integration in Single Neurons of the Midbrain. *The Handbook of Multisensory Processes.* Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2004-17469-015

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

Stein, B. E., & Wallace, M. T. (1996). Comparisons of cross-modality integration in midbrain and cortex. *Progress in Brain Research*, *112*, 289–299. https://doi.org/10.1016/S0079-6123(08)63336-1

Stein, B. E., Wallace, M. W., Stanford, T. R., & Jiang, W. (2002). Book review: Cortex governs multisensory integration in the midbrain. *The Neuroscientist*, *8*(4), 306–314. https://doi.org/10.1177/107385840200800406

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: fundamental issues and a general model. *Attention and Performance IV*, 629–685. https://doi.org/10.1016/j.biosystems.2004.12.002

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–1223. https://doi.org/10.1016/j.neuroimage.2008.09.034

Stevenson, R. A., VanDerKlok, R. M., Pisoni, D. B., & James, T. W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *NeuroImage*, *55*(3), 1339–1345. https://doi.org/10.1016/j.neuroimage.2010.12.063

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: Task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249–261. https://doi.org/10.1007/s00221-013-3507-3

Storms, R. L., & Zyda, M. J. (2000). Interactions in perceived quality of auditory-visual displays. *Presence: Teleoperators and Virtual Environments*. https://doi.org/10.1162/105474600300040385

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

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

Thelen, A., Cappe, C., & Murray, M. M. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *NeuroImage*.

https://doi.org/10.1016/j.neuroimage.2012.05.027

Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expércompréhension d'un Texte Scientifiquerimentale*, *48*(3), 380–398. https://doi.org/10.1037/1196-1961.48.3.380

Tiesinga, P., Fellous, J. M., & Sejnowski, T. J. (2008). Regulation of spike timing in visual cortical circuits. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn2315

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*. https://doi.org/10.1037/h0027242

Van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, *17*(4), 962–974. https://doi.org/10.1093/cercor/bhl007

van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2007). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2007.03.065

Van Engen, K. J., Dey, A., Sommers, M., & Peelle, J. E. (n.d.). Audiovisual speech perception: Moving beyond McGurk. https://doi.org/https://doi.org/10.31234/osf.io/6y8qw

Van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention, Perception, & Psychophysics*, *79*(2), 396–403. https://doi.org/10.3758/s13414-016-1238-9

Vickery, K. S., Reynolds, V. A., & Cochran, S. W. (1987). Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. *Annals of Dyslexia*, *37*(1), 189–200. https://doi.org/10.1007/BF02648066

Von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, *4*(10), 1809–1820. https://doi.org/10.1371/journal.pbio.0040326

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

Vroomen, J., & Keetels, M. (2009). Perception of intersensory synchrony: A tutorial review. *Attention, Perception & Psychophysics*, *71*(3), 481–489. https://doi.org/10.3758/APP

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*(1), 21–25. https://doi.org/10.1177/0956797609354734

Walker, P., Scallon, G., & Francis, B. (2017). Cross-Sensory Correspondences: Heaviness is Dark and Low-Pitched. *Perception*. https://doi.org/10.1177/0301006616684369

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*(2), 252–258. https://doi.org/10.1007/s00221-004-1899-9

Wallace, M. T., & Stein, B. E. (1996). Chapter 21 Sensory organization of the superior colliculus in cat and monkey. *Progress in Brain Research*, *112*(C), 301–311. https://doi.org/10.1016/S0079-6123(08)63337-3

Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(7), 2429–2444.

Walton, G. E., & Bower, T. G. R. (1993). Amodal representation of speech in infants. *Infant Behavior and Development*, *16*(2), 233–243. https://doi.org/10.1016/0163-6383(93)80019-5

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, *42*(2), 264–280. https://doi.org/10.1016/S0010-9452(08)70352-6

Watkins, O. C., & Watkins, M. J. (1980). The modality effect and echoic persistence. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.109.3.251

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In *Handbook of perception and human performance: Sensory processes and perception* (Vol. I, p. 25.1-25.36).

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, *97*(20), 11125–11129. https://doi.org/10.1073/pnas.97.20.11125

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11. https://doi.org/10.1167/4.12.11

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences of the United States of America*.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2012). Flexibility in visual working memory: Accurate change detection in the face of irrelevant variations in position. *Visual Cognition*. https://doi.org/10.1080/13506285.2011.630694

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What You See Is Not (Always) What You Hear: Induced Gamma Band Responses Reflect Cross-Modal Interactions in Familiar Object Recognition. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.4828-06.2007

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, *67*(3), 531–544. https://doi.org/10.3758/BF03193329

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233-U13. https://doi.org/Doi 10.1038/Nature06860

Zwislocki, J. (2005). Theory of Temporal Auditory Summation. *The Journal of the Acoustical*

Society of America. https://doi.org/10.1121/1.1908276

Zwislocki, J. J. (2005). Temporal Summation of Loudness: An Analysis. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1911708

List of Appendices

Appendix A. List of definitions used for each animal category in Chapter 2 Experiment 1

Appendix B. Pairwise comparisons of working memory performance (accuracy, reaction time, capacity) across conditions for experiments in Chapter 2

Appendix C. Pairwise comparisons of working memory performance (accuracy, reaction time, capacity) across conditions for experiments in Chapter 3

Appendix D. Statistical results of performance comparisons across experimental sessions in Chapter 4

Appendix E. Pairwise comparisons of working memory performance (accuracy, reaction time, capacity) across conditions for tests with simple and complex stimuli in Chapter 4

Appendix A. List of definitions used for each animal category in Chapter 2 Experiment 1

Cat: A small animal with fur, four legs a tail, and claws, usually kept as a pet or for catching mice, or other large wild animal from the cat family

Dog: A domesticated carnivorous mammal that typically has a long snout, an acute sense of smell, non-retractable claws, and a barking, howling, or whining voice.

Monkey: Any of a group of mammals that usually have flat faces and long tails, especially any of the smaller mammals in this group

Frog: A tailless amphibian with a short squat body, moist smooth skin, and very long hind legs for leaping.

Insect: A type of small animal with six legs, a body divided into three parts, and often two pairs of wings, for example, an ant, beetle, or butterfly.

Bird: A warm-blooded egg-laying vertebrate animal distinguished by the possession of feathers, wings, a beak, and typically by being able to fly.

Whale: A very large marine mammal with a streamlined hairless body, a horizontal tail fin, and a blowhole on top of the head for breathing.

Pinniped: Any of an order or suborder (Pinnipedia) of aquatic carnivorous mammals (such as a seal or walrus) with all four limbs modified into flippers.

Appendix B. Pairwise comparisons of working memory performance (accuracy, reaction time, capacity) across conditions for experiments in Chapter 2

Experiment 1

Effects of Modality

Table 1: Accuracy

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(34)=1.17, p=0.25 4: t(34)=2.69, p=0.01 8: t(34)=7.29, p<0.001			
AV-Test V	2: t(34)=0.66, p=0.51 4: t(34)=2.59, p=0.01 8: t(34)=3.78, p=0.001	2: t(34)<0.001, p=1.00 4: t(34)<0.001, p=1.00 8: t(34)=0.07, p=0.95		
AV-Test A	2: t(34)=0.21, p=0.83 4: t(34)=3.33, p<0.01 8: t(34)=4.73, p<0.001	2: t(34)=0.56, p=0.58 4: t(34)=1.26, p=0.22 8: t(34)=0.31, p=0.76	2: t(34)=0.53, p=0.60 4: t(34)=1.13, p=0.27 8: t(34)=0.39, p=0.70	

Table 2: Reaction Time

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(34)=1.92, p=0.06 4: t(34)=0.88, p=0.38 8: t(34)=0.38, p=0.70			
AV-Test V	2: t(34)=0.22, p=0.83 4: t(34)=0.04, p=0.97 8: t(34)=0.77, p=0.45	2: t(34)=1.67, p=0.11 4: t(34)=0.89, p=0.38 8: t(34)=0.30, p=0.77		
AV-Test A	2: t(34)=2.69, p=0.01 4: t(34)=0.09, p=0.93 8: t(34)=1.10, p=0.28	2: t(34)=1.51, p=0.14 4: t(34)=1.09, p=0.28 8: t(34)=1.66, p=0.11	2: t(34)=3.10, p<0.01 4: t(34)=0.19, p=0.85 8: t(34)=1.85, p=0.07	

Table 3: Capacity

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(34)=1.17, p=0.25 4: t(34)=2.69, p=0.01 8: t(34)=7.29, p<0.001			
AV-Test V	2: t(34)=0.79, p=0.44 4: t(34)=2.67, p=0.01 8: t(34)=3.69, p=0.001	2: t(34)=0.14, p=0.89 4: t(34)=0.35, p=0.73 8: t(34)=0.24, p=0.81		
AV-Test A	2: t(34)=0.50, p=0.62 4: t(34)=3.11, p<0.01 8: t(34)=5.41, p<0.001	2: t(34)=0.29, p=0.78 4: t(34)=1.19, p=0.24 8: t(34)=0.66, p=0.52	2: t(34)=0.40, p=0.69 4: t(34)=0.77, p=0.45 8: t(34)=0.81, p=0.42	

Effects of Congruency

Table 4: Accuracy

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(34)=3.22, p<0.01 4: t(34)=1.14, p=0.26 8: t(34)=0.16, p=0.87			
V-CC	2: t(34)=1.00, p=0.32 4: t(34)=0.50, p=0.62 8: t(34)=0.50, p=0.62	2: t(34)=1.79, p=0.08 4: t(34)=1.16, p=0.26 8: t(34)=0.68, p=0.50		
V-IC	2: t(34)=0.84, p=0.41 4: t(34)=0.82, p=0.42 8: t(34)=0.48, p=0.64	2: t(34)=1.79, p=0.08 4: t(34)=1.68, p=0.10 8: t(34)=0.39, p=0.70	2: t(34)<0.001, p=1.00 4: t(34)=0.27, p=0.79 8: t(34)=0.90, p=0.37	

Table 5: Reaction Time

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(34)=2.53, p=0.02 4: t(34)=0.61, p=0.55 8: t(34)=0.72, p=0.48			
V-CC	2: t(34)=0.48, p=0.63 4: t(34)=0.04, p=0.97 8: t(34)=0.77, p=0.45	2: t(34)=2.09, p=0.04 4: t(34)=0.86, p=0.40 8: t(34)=1.74, p=0.09		
V-IC	2: t(34)=1.84, p=0.07 4: t(34)=0.74, p=0.46 8: t(34)=1.13, p=0.27	2: t(34)=3.83, p=0.001 4: t(34)=0.02, p=0.99 8: t(34)=1.68, p=0.10	2: t(34)=1.94, p=0.06 4: t(34)=0.94, p=0.35 8: t(34)=0.17, p=0.86	

Table 6: Capacity

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(34)=3.45, p=0.001 4: t(34)=0.62, p=0.54 8: t(34)=1.28, p=0.21			
V-CC	2: t(34)= 1.00, p=0.32 4: t(34)=0.67, p=0.51 8: t(34)<0.001, p=1.00	2: t(34)=2.07, p=0.05 4: t(34)=1.04, p=0.31 8: t(34)=1.48, p=0.51		
V-IC	2: t(34)=1.26, p=0.22 4: t(34)=0.46, p=0.65 8: t(34)=0.67, p=0.51	2: t(34)=1.30, p=0.20 4: t(34)=0.91, p=0.37 8: t(34)=0.72, p=0.48	2: t(34)=0.48, p=0.64 4: t(34)=0.20, p=0.84 8: t(34)=0.59, p=0.56	

Experiment 2

Effects of Modality

Table 7: Accuracy

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(33)=0.08, p=0.93 4: t(33)=3.73, p=0.001			

	8: t(33)=3.41, p<0.01			
AV-Test V	2: t(33)=3.14, p<0.01 4: t(33)=3.69, p=0.001 8: t(33)=0.37, p=0.71	2: t(33)=2.06, p=0.05 4: t(33)=7.10, p<0.001 8: t(33)=2.38, p=0.02		
AV-Test A	2: t(33)=1.46, p=0.16 4: t(33)=1.26, p=0.22 8: t(33)=2.06, p=0.05	2: t(33)=1.17, p=0.25 4: t(33)=1.69, p=0.10 8: t(33)=0.42, p=0.68	2: t(33)=0.84, p=0.41 4: t(33)=3.83, p=0.001 8: t(33)=1.91, p=0.07	

Table 8: Reaction Time

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(33)=0.13, p=0.90 4: t(33)=0.71, p=0.49 8: t(33)=1.25, p=0.22			
AV-Test V	2: t(33)=1.00, p=0.33 4: t(33)=1.61, p=0.12 8: t(33)=0.92, p=0.37	2: t(33)=0.94, p=0.35 4: t(33)=2.49, p=0.02 8: t(33)=0.31, p=0.76		
AV-Test A	2: t(33)=0.09, p=0.93 4: t(33)=1.24, p=0.23 8: t(33)=1.02, p=0.31	2: t(33)=0.02, p=0.98 4: t(33)=0.82, p=0.42 8: t(33)=0.30, p=0.77	2: t(33)=1.10, p=0.28 4: t(33)=3.01, p<0.01 8: t(33)=0.03, p=0.97	

Table 9: Capacity

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(33)=0.08, p=0.93 4: t(33)=3.73, p=0.001 8: t(33)=3.41, p<0.01			
AV-Test V	2: t(33)=3.24, p<0.01 4: t(33)=2.44, p=0.02 8: t(33)=0.09, p=0.93	2: t(33)=2.08, p=0.05 4: t(33)=6.05, p<0.001 8: t(33)=2.49, p=0.02		
AV-Test A	2: t(33)=1.72, p=0.09 4: t(33)=1.00, p=0.32 8: t(33)=2.78, p<0.01	2: t(33)=1.37, p=0.18 4: t(33)=1.81, p=0.08 8: t(33)=0.26, p=0.80	2: t(33)=0.81, p=0.42 4: t(33)=2.67, p=0.01 8: t(33)=2.79, p<0.01	

Effects of Congruency

Table 10: Accuracy

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(33)=2.15, p=0.04 4: t(33)=2.51, p=0.02 8: t(33)=1.03, p=0.31			
V-CC	2: t(33)= 1.16, p=0.25 4: t(33)=1.97, p=0.06 8: t(33)=0.65, p=0.52	2: t(33)=1.66, p=0.11 4: t(33)=3.78, p=0.001 8: t(33)=1.81, p=0.08		
V-IC	2: t(33)= 0.30, p=0.77 4: t(33)=1.42, p=0.17 8: t(33)=0.63, p=0.54	2: t(33)=1.23, p=0.23 4: t(33)=4.06, p<0.001 8: t(33)=1.48, p=0.15	2: t(33)=0.68, p=0.50 4: t(33)=0.59, p=0.56 8: t(33)=0.11, p=0.92	

Table 11: Reaction Time

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(33)=1.58, p=0.12 4: t(33)=1.40, p=0.17 8: t(33)=0.79, p=0.43			
V-CC	2: t(33)=2.12, p=0.04 4: t(33)=3.52, p=0.001 8: t(33)=0.28, p=0.78	2: t(33)=0.37, p=0.71 4: t(33)=2.49, p=0.02 8: t(33)=0.57, p=0.58		
V-IC	2: t(33)=0.73, p=0.47 4: t(33)=2.14, p=0.04 8: t(33)=0.68, p=0.50	2: t(33)=0.62, p=0.54 4: t(33)=1.38, p=0.18 8: t(33)=0.20, p=0.85	2: t(33)=0.93, p=0.36 4: t(33)=0.99, p=0.33 8: t(33)=0.41, p=0.68	

Table 12: Capacity

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(33)=2.17, p=0.04 4: t(33)=2.04, p=0.05 8: t(33)=1.86, p=0.07			
V-CC	2: t(33)<0.001, p=1.00 4: t(33)=1.58, p=0.12 8: t(33)=0.38, p=0.70	2: t(33)=2.07, p=0.05 4: t(33)=3.10, p<0.01 8: t(33)=2.45, p=0.02		
V-IC	2: t(33)=1.49, p=0.15 4: t(33)=0.78, p=0.44 8: t(33)=1.12, p=0.27	2: t(33)=0.88, p=0.38 4: t(33)=2.63, p=0.01 8: t(33)=2.64, p=0.01	2: t(33)=1.31, p=0.20 4: t(33)=0.73, p=0.47 8: t(33)=0.47, p=0.64	

Experiment 3

Effects of Modality

Table 13: Accuracy

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(32)=3.46, p<0.01 6: t(32)=3.57, p=0.001 8: t(32)=2.91, p<0.01			
AV-Test V	2: t(32)=3.31, p<0.01 6: t(32)=1.57, p=0.13 8: t(32)=1.22, p=0.23	2: t(32)=6.65, p<0.001 6: t(32)=1.61, p=0.12 8: t(32)=1.20, p=0.24		
AV-Test A	2: t(32)=1.36, p=0.18 6: t(32)=1.49, p=0.15 8: t(32)=4.96, p<0.001	2: t(32)=1.01, p=0.32 6: t(32)=1.38, p=0.18 8: t(32)=3.13, p<0.01	2: t(32)=4.24, p<0.001 6: t(32)=0.20, p=0.84 8: t(32)=3.48, p=0.001	

Table 14: Reaction Time

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				

Auditory alone	2: t(32)=2.22, p=0.03 6: t(32)=2.48, p=0.02 8: t(32)=1.31, p=0.20			
AV-Test V	2: t(32)=0.98, p=0.34 6: t(32)=1.91, p=0.07 8: t(32)=0.78, p=0.44	2: t(32)=4.63, p<0.001 6: t(32)=1.09, p=0.28 8: t(32)=0.85, p=0.40		
AV-Test A	2: t(32)=2.64, p=0.01 6: t(32)=2.32, p=0.03 8: t(32)=3.42, p<0.01	2: t(32)=0.37, p=0.71 6: t(32)=0.43, p=0.67 8: t(32)=2.51, p=0.04	2: t(32)=3.97, p<0.001 6: t(32)=0.83, p=0.41 8: t(32)=2.95, p<0.01	

Table 15: Capacity

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(32)=3.46, p<0.01 6: t(32)=3.57, p=0.001 8: t(32)=2.91, p<0.01			
AV-Test V	2: t(32)=2.96, p<0.01 6: t(32)=1.50, p=0.14 8: t(32)=1.45, p=0.16	2: t(32)=5.82, p<0.001 6: t(32)=1.63, p=0.11 8: t(32)=0.94, p=0.36		
AV-Test A	2: t(32)=1.14, p=0.26 6: t(32)=1.42, p=0.17 8: t(32)=5.23, p<0.001	2: t(32)=1.05, p=0.30 6: t(32)=1.35, p=0.19 8: t(32)=3.45, p<0.01	2: t(32)=3.56, p=0.001 6: t(32)=0.23, p=0.82 8: t(32)=3.49, p=0.001	

Effects of Congruency

Table 16: Accuracy

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(32)=0.15, p=0.88 6: t(32)=1.61, p=0.12 8: t(32)=0.81, p=0.42			
V-CC	2: t(32)=2.60, p=0.01 6: t(32)=0.68, p=0.50 8: t(32)=1.61, p=0.12	2: t(32)=2.10, p=0.04 6: t(32)=1.00, p=0.33 8: t(32)=2.48, p=0.018		
V-IC	2: t(32)=3.2, p<0.01 6: t(32)=1.25, p=0.22 8: t(32)=3.30, p<0.01	2: t(32)=2.95, p<0.01 6: t(32)=0.28, p=0.78 8: t(32)=3.73, p=0.001	2: t(32)=1.09, p=0.28 6: t(32)=0.80, p=0.43 8: t(32)=1.07, p=0.29	

Table 17: Reaction Time

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(32)=0.33, p=0.75 6: t(32)=0.57, p=0.57 8: t(32)=1.12, p=0.27			
V-CC	2: t(32)=2.58, p=0.015 6: t(32)=0.88, p=0.39 8: t(32)=1.19, p=0.24	2: t(32)=3.43, p<0.01 6: t(32)=0.26, p=0.79 8: t(32)=2.34, p=0.03		
V-IC	2: t(32)=3.15, p<0.01 6: t(32)=0.78, p=0.44	2: t(32)=3.48, p=0.001 6: t(32)=0.09, p=0.93	2: t(32)=0.66, p=0.51 6: t(32)=0.16, p=0.88	

	8: t(32)=1.97, p=0.06	8: t(32)=2.68, p=0.01	8: t(32)=0.76, p=0.45	

Table 18: Capacity

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(32)=0.22, p=0.83 6: t(32)=1.62, p=0.11 8: t(32)=1.68, p=0.10			
V-CC	2: t(32)=2.95, p<0.01 6: t(32)=1.41, p=0.17 8: t(32)=1.39, p=0.17	2: t(32)=1.94, p=0.06 6: t(32)=0.56, p=0.58 8: t(32)=2.91, p<0.01		
V-IC	2: t(32)=2.46, p=0.02 6: t(32)=0.79, p=0.44 8: t(32)=2.48, p=0.02	2: t(32)=2.15, p=0.04 6: t(32)=0.80, p=0.43 8: t(32)=3.83, p=0.001	2: t(32)=0.42, p=0.68 6: t(32)=0.35, p=0.73 8: t(32)=0.72, p=0.48	

Experiment 4

Table 19: Accuracy

Table 19A:

	A-CC Distance 0	A-IC Distance 1	A-IC Distance 2
A-CC Distance 0			
A-IC Distance 1	4: t(38)=1.94, p=0.06 6: t(38)=0.98, p=0.33 8: t(38)=1.03, p=0.31		
A-IC Distance 2	4: t(38)=3.05, p<0.01 6: t(38)=2.61, p=0.01 8: t(38)=0.99, p=0.33	4: t(38)=0.50, p=0.62 6: t(38)=1.58, p=0.12 8: t(38)=0.08, p=0.94	

Table 19B:

	V-CC Distance 0	V-IC Distance 1	V-IC Distance 2
V-CC Distance 0			
V-IC Distance 1	4: t(38)=1.43, p=0.16 6: t(38)=2.61, p=0.01 8: t(38)=0.78, p=0.44		
V-IC Distance 2	4: t(38)=1.98, p=0.06 6: t(38)=1.92, p=0.06 8: t(38)=0.84, p=0.41	4: t(38)=0.77, p=0.44 6: t(38)=0.52, p=0.61 8: t(38)=1.74, p=0.09	

Table 20: Reaction Time

Table 20A:

	A-CC Distance 0	A-IC Distance 1	A-IC Distance 2
A-CC Distance 0			
A-IC Distance 1	4: t(38)=1.94, p=0.06 6: t(38)=0.98, p=0.33 8: t(38)=1.03, p=0.31		

A-IC Distance 2	4: t(38)=3.05, p<0.01 6: t(38)=2.61, p=0.01 8: t(38)=0.99, p=0.33	4: t(38)=0.50, p=0.62 6: t(38)=1.58, p=0.12 8: t(38)=0.08, p=0.94	

Table 20B:

	V-CC Distance 0	V-IC Distance 1	V-IC Distance 2
V-CC Distance 0			
V-IC Distance 1	4: t(38)=1.43, p=0.16 6: t(38)=2.61, p=0.01 8: t(38)=0.78, p=0.44		
V-IC Distance 2	4: t(38)=1.98, p=0.06 6: t(38)=1.92, p=0.06 8: t(38)=0.84, p=0.41	4: t(38)=0.77, p=0.44 6: t(38)=0.52, p=0.61 8: t(38)=1.74, p=0.09	

Table 21: Capacity

Table 21A:

	A-CC Distance 0	A-IC Distance 1	A-IC Distance 2
A-CC Distance 0			
A-IC Distance 1	4: t(38)=0.88, p=0.39 6: t(38)=<0.01, p=0.10 8: t(38)=0.95, p=0.35		
A-IC Distance 2	4: t(38)=0.06, p=0.95 6: t(38)=0.68, p=0.50 8: t(38)=0.60, p=0.55	4: t(38)=1.08, p=0.29 6: t(38)=0.59, p=0.56 8: t(38)=1.40, p=0.17	

Table 21B:

	V-CC Distance 0	V-IC Distance 1	V-IC Distance 2
V-CC Distance 0			
V-IC Distance 1	4: t(38)=0.03, p=0.98 6: t(38)=1.25, p=0.22 8: t(38)=2.17, p=0.04		
V-IC Distance 2	4: t(38)=0.38, p=0.71 6: t(38)=2.10, p=0.04 8: t(38)=0.27, p=0.79	4: t(38)=0.43, p=0.67 6: t(38)=0.85, p=0.40 8: t(38)=1.25, p=0.22	

Appendix C. Pairwise comparisons of working memory performance (accuracy, reaction time, capacity) across conditions for experiments in Chapter 3

Experiment 2

Effects of Modality

Table 1: Accuracy

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(34)=0.48, p=0.64 4: t(34)=4.07, p<0.001 8: t(34)=2.82, p<0.01			
AV-Test V	2: t(34)=0.68, p=0.50 4: t(34)=0.67, p=0.51 8: t(34)=1.07, p=0.29	2: t(34)=1.06, p=0.30 4: t(34)=4.29, p<0.001 8: t(34)=1.23, p=0.23		
AV-Test A	2: t(34)=2.96, p<0.01 4: t(34)=0.36, p=0.72 8: t(34)=2.17, p=0.04	2: t(34)=3.18, p<0.01 4: t(34)=2.91, p<0.01 8: t(34)=0.48, p=0.63	2: t(34)=1.77, p=0.09 4: t(34)=0.93, p=0.36 8: t(34)=0.69, p=0.49	

Table 2: Reaction Time

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(34)=0.12, p=0.91 4: t(34)=1.16, p=0.25 8: t(34)=2.09, p=0.04			
AV-Test V	2: t(34)=0.75, p=0.46 4: t(34)=0.23, p=0.82 8: t(34)=0.48, p=0.64	2: t(34)=0.54, p=0.59 4: t(34)=1.22, p=0.23 8: t(34)=2.30, p=0.03		
AV-Test A	2: t(34)=1.40, p=0.17 4: t(34)=0.27, p=0.79 8: t(34)=0.19, p=0.85	2: t(34)=0.88, p=0.39 4: t(34)=1.27, p=0.21 8: t(34)=1.58, p=0.13	2: t(34)=0.31, p=0.76 4: t(34)=0.03, p=0.97 8: t(34)=0.67, p=0.51	

Table 3: Capacity

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(34)=0.48, p=0.64 4: t(34)=4.37, p<0.001 8: t(34)=3.04, p<0.01			
AV-Test V	2: t(34)=0.81, p=0.42 4: t(34)=0.25, p=0.80 8: t(34)=1.53, p=0.14	2: t(34)=0.92, p=0.36 4: t(34)=3.58, p=0.001 8: t(34)=0.66, p=0.51		
AV-Test A	2: t(34)=3.12, p<0.01 4: t(34)=0.25, p0.80 8: t(34)=1.79, p=0.08	2: t(34)=3.36, p<0.01 4: t(34)=3.52, p=0.001 8: t(34)=0.77, p=0.45	2: t(34)=1.87, p=0.07 4: t(34)=0.44, p=0.67 8: t(34)=0.06, p=0.95	

Effects of Congruency

Table 4: Accuracy

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(34)=3.17, p<0.01 4: t(34)=3.19, p<0.01 8: t(34)=2.68, p=0.01			
V-CC	2: t(34)=3.05, p<0.01 4: t(34)=2.18, p=0.04 8: t(34)=1.51, p=0.14	2: t(34)=1.10, p=0.28 4: t(34)=0.80, p=0.43 8: t(34)=0.83, p=0.41		
V-IC	2: t(34)=2.38, p=0.02 4: t(34)=0.52, p=0.61 8: t(34)=2.13, p=0.04	2: t(34)=0.83, p=0.41 4: t(34)=3.52, p=0.001 8: t(34)=0.51, p=0.61	2: t(34)=1.86, p=0.07 4: t(34)=2.60, p=0.01 8: t(34)=0.39, p=0.70	

Table 5: Reaction Time

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(34)=1.93, p=0.06 4: t(34)=1.8, p=0.08 8: t(34)=1.46, p=0.15			
V-CC	2: t(34)=0.08, p=0.94 4: t(34)=1.35, p=0.19 8: t(34)=2.07, p=0.05	2: t(34)=1.66, p=0.11 4: t(34)=0.09, p=0.93 8: t(34)=0.09, p=0.93		
V-IC	2: t(34)=1.24, p=0.23 4: t(34)=0.18, p=0.86 8: t(34)=1.49, p=0.15	2: t(34)=0.32, p=0.75 4: t(34)=1.42, p=0.16 8: t(34)=0.39, p=0.70	2: t(34)=2.01, p=0.05 4: t(34)=1.48, p=0.15 8: t(34)=0.53, p=0.60	

Table 6: Capacity

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(34)=3.01, p<0.01 4: t(34)=3.14, p<0.01 8: t(34)=2.15, p=0.04			
V-CC	2: t(34)=2.93, p<0.01 4: t(34)=1.94, p=0.06 8: t(34)=0.89, p=0.38	2: t(34)=1.25, p=0.22 4: t(34)=0.93, p=0.36 8: t(34)=1.03, p=0.31		
V-IC	2: t(34)=2.24, p=0.03 4: t(34)=0.16, p=0.87 8: t(34)=1.45, p=0.16	2: t(34)=0.68, p=0.50 4: t(34)=2.96, p<0.01 8: t(34)=0.52, p=0.61	2: t(34)=1.83, p=0.08 4: t(34)=1.99, p=0.05 8: t(34)=0.59, p=0.56	

Experiment 3

Effects of Modality

Table 7: Accuracy

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				

Auditory alone	4: t(31)=3.88, p=0.001 6: t(31)=2.27, p=0.03 8: t(31)=5.61, p<0.001			
AV-Test V	4: t(31)=1.74, p=0.09 6: t(31)=0.63, p=0.53 8: t(31)=1.45, p=0.16	4: t(31)=4.40, p<0.001 6: t(31)=1.48, p=0.15 8: t(31)=2.68, p=0.01		
AV-Test A	4: t(31)=3.29, p<0.01 6: t(31)=0.79, p=0.44 8: t(31)=1.98, p=0.06	4: t(31)=0.35, p=0.73 6: t(31)=2.84, p<0.01 8: t(31)=2.10, p=0.04	4: t(31)=5.44, p<0.001 6: t(31)=1.05, p=0.30 8: t(31)=0.26, p=0.79	

Table 8: Reaction Time

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	4: t(31)=1.76, p=0.09 6: t(31)=1.17, p=0.25 8: t(31)=1.03, p=0.31			
AV-Test V	4: t(31)=1.80, p=0.08 6: t(31)=0.47, p=0.64 8: t(31)=0.65, p=0.52	4: t(31)=3.39, p<0.01 6: t(31)=1.62, p=0.12 8: t(31)=1.53, p=0.14		
AV-Test A	4: t(31)=1.59, p=0.12 6: t(31)=1.10, p=0.28 8: t(31)=1.30, p=0.21	4: t(31)=0.30, p=0.77 6: t(31)=0.27, p=0.79 8: t(31)=0.47, p=0.65	4: t(31)=2.67, p=0.01 6: t(31)=2.12, p=0.04 8: t(31)=1.96, p=0.06	

Table 9: Capacity

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	4: t(31)=3.81, p=0.001 6: t(31)=2.27, p=0.03 8: t(31)=5.61, p<0.001			
AV-Test V	4: t(31)=1.68, p=0.10 6: t(31)=0.52, p=0.61 8: t(31)=1.82, p=0.08	4: t(31)=4.27, p<0.001 6: t(31)=1.48, p=0.15 8: t(31)=1.96, p=0.06		
AV-Test A	4: t(31)=2.56, p=0.02 6: t(31)=0.10, p=0.33 8: t(31)=1.47, p=0.15	4: t(31)=0.15, p=0.88 6: t(31)=2.92, p<0.01 8: t(31)=2.76, p=0.01	4: t(31)=4.43, p<0.001 6: t(31)=1.14, p=0.26 8: t(31)=0.59, p=0.56	

Effects of Congruency

Table 10: Accuracy

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	4: t(31)=2.17, p=0.04 6: t(31)=0.79, p=0.44 8: t(31)=2.03, p=0.05			
V-CC	4: t(31)=3.91, p<0.001 6: t(31)=0.70, p=0.49 8: t(31)<0.001, p=1.00	4: t(31)=2.64, p=0.01 6: t(31)<0.01, p=0.10 8: t(31)=1.89, p=0.07		
V-IC	4: t(31)=5.28, p<0.001 6: t(31)=1.51, p=0.14	4: t(31)=3.04, p<0.01 6: t(31)=0.75, p=0.46	4: t(31)=0.23, p=0.82 6: t(31)=0.80, p=0.43	

	8: t(31)=2.27, p=0.03	8: t(31)=0.39, p=0.70	8: t(31)=2.26, p=0.03	

Table 11: Reaction Time

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	4: t(31)=0.5, p=0.62 6: t(31)=1.20, p=0.24 8: t(31)=0.57, p=0.57			
V-CC	4: t(31)=2.22, p=0.03 6: t(31)=1.93, p=0.06 8: t(31)=1.44, p=0.16	4: t(31)=2.95, p<0.01 6: t(31)=3.26, p<0.01 8: t(31)=2.15, p=0.04		
V-IC	4: t(31)=1.38, p=0.18 6: t(31)=0.75, p=0.46 8: t(31)=0.81, p=0.42	4: t(31)=1.97, p=0.06 6: t(31)=0.76, p=0.45 8: t(31)=1.38, p=0.18	4: t(31)=1.47, p=0.15 6: t(31)=2.78, p<0.01 8: t(31)=0.72, p=0.48	

Table 12: Capacity

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	4: t(31)=1.89, p=0.07 6: t(31)=1.10, p=0.28 8: t(31)=0.83, p=0.41			
V-CC	4: t(31)=3.04, p<0.01 6: t(31)=1.07, p=0.29 8: t(31)=1.25, p=0.22	4: t(31)=2.44, p=0.02 6: t(31)=0.14, p=0.89 8: t(31)=2.20, p=0.04		
V-IC	4: t(31)=4.37, p<0.001 6: t(31)=1.45, p=0.16 8: t(31)=1.37, p=0.18	4: t(31)=2.56, p=0.02 6: t(31)=0.51, p=0.61 8: t(31)=0.50, p=0.62	4: t(31)=0.27, p=0.79 6: t(31)=0.37, p=0.71 8: t(31)=2.77, p<0.01	

Experiment 4

Table 13: Accuracy
Table 13 A:

	A-CC Distance 0	A-IC Distance 1	A-IC Distance 2
A-CC Distance 0			
A-IC Distance 1	4: t(32)=0.68, p=0.50 6: t(32)=0.94, p=0.35 8: t(32)=0.49, p=0.63		
A-IC Distance 2	4: t(32)=1.01, p=0.32 6: t(32)=1.12, p=0.27 8: t(32)=1.19, p=0.24	4: t(32)=0.24, p=0.82 6: t(32)=0.14, p=0.89 8: t(32)=1.65, p=0.11	

Table 13 B:

	V-CC Distance 0	V-IC Distance 1	V-IC Distance 2
V-CC Distance 0			

V-IC Distance 1	4: t(32)=0.27, p=0.79 6: t(32)=1.61, p=0.12 8: t(32)=1.16, p=0.25		
V-IC Distance 2	4: t(32)= 0.58, p=0.57 6: t(32)=0.59, p=0.56 8: t(32)=0.11, p=0.91	4: t(32)=1.36, p=0.18 6: t(32)=2.46, p=0.02 8: t(32)=1.46, p=0.16	

Table 14: Reaction Time

Table 14A:

	A-CC Distance 0	A-IC Distance 1	A-IC Distance 2
A-CC Distance 0			
A-IC Distance 1	4: t(32)=0.68, p=0.50 6: t(32)=0.94, p=0.35 8: t(32)=0.61, p=0.55		
A-IC Distance 2	4: t(32)=1.06, p=0.30 6: t(32)=2.18, p=0.04 8: t(32)=1.19, p=0.24	4: t(32)=0.24, p=0.81 6: t(32)=0.85, p=0.40 8: t(32)=1.76, p=0.09	

Table 14B:

	V-CC Distance 0	V-IC Distance 1	V-IC Distance 2
V-CC Distance 0			
V-IC Distance 1	4: t(32)=0.36, p=0.72 6: t(32)=1.61, p=0.12 8: t(32)=1.28, p=0.21		
V-IC Distance 2	4: t(32)=0.49, p=0.63 6: t(32)=0.60, p=0.56 8: t(32)=0.23, p=0.82	4: t(32)=1.36, p=0.18 6: t(32)=2.46, p=0.02 8: t(32)=1.46, p=0.16	

Table 15: Capacity

Table 15 A:

	A-CC Distance 0	A-IC Distance 1	A-IC Distance 2
A-CC Distance 0			
A-IC Distance 1	4: t(32)=0.52, p=0.61 6: t(32)=1.63, p=0.11 8: t(32)=0.87, p=0.39		
A-IC Distance 2	4: t(32)=0.26, p=0.79 6: t(32)=0.80, p=0.43 8: t(32)=0.29, p=0.77	4: t(32)=0.35, p=0.73 6: t(32)=0.98, p=0.34 8: t(32)=0.76, p=0.45	

Table 15 B:

	V-CC Distance 0	V-IC Distance 1	V-IC Distance 2
V-CC Distance 0			
V-IC Distance 1	4: t(32)=1.53, p=0.14 6: t(32)=0.19, p=0.85 8: t(32)=0.23, p=0.82		

V-IC Distance 2	4: t(32)=0.28, p=0.78 6: t(32)=1.07, p=0.29 8: t(32)=0.03, p=0.98	4: t(32)=1.85, p=0.07 6: t(32)=0.82, p=0.42 8: t(32)=0.29, p=0.77	

Appendix D. Statistical results of performance comparisons across experimental sessions in Chapter 4

Repeated measures ANOVA with session as a between-subject factor, with load, and condition as the within-subject factors.

Simple Stimuli

Table 1: Accuracy

	F	p	η_p^2
Condition	13.21	<0.001	0.26
Condition x Session	0.58	0.63	0.02
Load	155.42	<0.001	0.81
Load x Session	0.17	0.84	0.01
Condition x Load	7.22	<0.001	0.16
Condition x Load x Session	0.49	0.81	0.01
Session	0.07	0.79	<0.001

Table 2: Reaction Time

	F	p	η_p^2
Condition	2.94	0.04	0.07
Condition x Session	1.00	0.40	0.03
Load	74.99	<0.001	0.67
Load x Session	0.15	0.86	<0.001
Condition x Load	0.82	0.55	0.02
Condition x Load x Session	0.75	0.61	0.02
Session	0.59	0.45	0.02

Table 3: Capacity Estimates

	F	p	η_p^2
Condition	8.70	<0.001	0.19
Condition x Session	0.31	0.82	0.01
Load	3.60	0.03	0.09
Load x Session	0.15	0.86	<0.001
Condition x Load	4.71	<0.001	0.11
Condition x Load x Session	0.15	0.99	<0.001
Session	0.07	0.80	<0.001

Complex Stimuli

Table 4: Accuracy

	F	p	η_p^2
Condition	9.30	<0.001	0.20
Condition x Session	0.25	0.86	0.01
Load	196.89	<0.001	0.84
Load x Session	1.11	0.33	0.03
Condition x Load	3.71	<0.001	0.09
Condition x Load x Session	0.23	0.97	0.01
Session	0.31	0.58	0.01

Table 5: Reaction Time

	F	p	η_p^2
Condition	1.14	0.34	0.03
Condition x Session	1.89	0.14	0.05
Load	55.92	<0.001	0.60
Load x Session	0.25	0.78	0.01
Condition x Load	5.40	<0.001	0.13
Condition x Load x Session	0.02	1.00	0.00
Session	32.19	<0.001	0.47

Table 6: Capacity Estimates

	F	p	η_p^2
Condition	6.74	<0.001	0.15
Condition x Session	0.26	0.85	0.01
Load	5.09	0.01	0.12
Load x Session	2.31	0.13	0.06
Condition x Load	2.94	0.01	0.07
Condition x Load x Session	0.23	0.97	0.01
Session	0.78	0.38	0.02

Appendix E. Pairwise comparisons of working memory performance (accuracy, reaction time, capacity) across conditions for tests with simple and complex stimuli in Chapter 4

Simple Stimuli

Effects of Modality

Table 1: Accuracy

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(38)=1.00, p=0.35 4: t(38)=3.00, p<0.01 8: t(38)=3.00, p<0.01			
AV-Test V	2: t(38)=4.69, p<0.001 4: t(38)=5.00, p<0.001 8: t(38)=1.00, p=0.23	2: t(38)=4.21, p<0.001 4: t(38)=8.17, p<0.001 8: t(38)=1.49, p=0.15		
AV-Test A	2: t(38)=1.66, p=0.11 4: t(38)=1.71, p=0.10 8: t(38)=3.30, p<0.01	2: t(38)=0.98, p=0.33 4: t(38)=0.75, p=0.46 8: t(38)=0.50, p=0.62	2: t(38)=2.00, p=0.07 4: t(38)=5.00, p<0.001 8: t(38)=2.00, p=0.08	

Table 2: Reaction Time

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(38)=1.00, p=0.40 4: t(38)=2.86, p<0.01 8: t(38)=1.00, p=0.18			
AV-Test V	2: t(38)=0.02, p=0.99 4: t(38)<0.001, p=0.73 8: t(38)=1.00, p=0.49	2: t(38)=0.85, p=0.40 4: t(38)=2.98, p<0.01 8: t(38)=1.87, p=0.07		
AV-Test A	2: t(38)=0.16, p=0.88 4: t(38)=1.35, p=0.18 8: t(38)=0.73, p=0.47	2: t(38)=0.85, p=0.40 4: t(38)=0.65, p=0.52 8: t(38)=0.19, p=0.85	2: t(38)<0.001, p=0.86 4: t(38)=2.00, p=0.11 8: t(38)=2.00, p=0.08	

Table 3: Capacity

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(38)=1.00, p=0.32 4: t(38)=3.15, p<0.01 8: t(38)=3.31, p<0.01			
AV-Test V	2: t(38)=4.08, p<0.001 4: t(38)=4.45, p<0.001 8: t(38)=0.83, p=0.41	2: t(38)=3.79, p=0.001 4: t(38)=7.14, p<0.001 8: t(38)=1.71, p=0.10		
AV-Test A	2: t(38)=2.07, p=0.05 4: t(38)=1.71, p=0.10 8: t(38)=3.43, p=0.001	2: t(38)=1.00, p=0.18 4: t(38)=1.00, p=0.57 8: t(38)=1.00, p=0.56	2: t(38)=1.28, p=0.21 4: t(38)=4.00, p<0.001 8: t(38)=2.00, p=0.05	

Effects of Congruency

Table 4: Accuracy

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(38)=1.00, p=0.51 4: t(38)=1.00, p=0.53 8: t(38)=2.00, p=0.07			
V-CC	2: t(38)=1.42, p=0.16 4: t(38)=2.45, p=0.02 8: t(38)=2.81, p<0.01	2: t(38)=0.74, p=0.47 4: t(38)=3.30, p<0.01 8: t(38)=1.04, p=0.31		
V-IC	2: t(38)=1.96, p=0.06 4: t(38)=4.20, p<0.001 8: t(38)=1.16, p=0.25	2: t(38)=0.83, p=0.41 4: t(38)=6.00, p<0.001 8: t(38)=1.00, p=0.27	2: t(38)<0.001, p=0.80 4: t(38)=2.00, p=0.03 8: t(38)=2.00, p=0.08	

Table 5: Reaction Time

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(38)<0.001, p=0.84 4: t(38)<0.001, p=0.81 8: t(38)=1.00, p=0.38			
V-CC	2: t(38)=0.22, p=0.83 4: t(38)=0.10, p=0.33 8: t(38)=0.76, p=0.45	2: t(38)<0.01, p=0.10 4: t(38)=0.77, p=0.45 8: t(38)=1.74, p=0.09		
V-IC	2: t(38)=0.22, p=0.83 4: t(38)=1.70, p=0.10 8: t(38)=0.45, p=0.66	2: t(38)=0.06, p=0.95 4: t(38)=2.00, p=0.12 8: t(38)=2.00, p=0.09	2: t(38)<0.001, p=0.95 4: t(38)=1.00, p=0.20 8: t(38)<0.001, p=0.75	

Table 6: Capacity

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(38)=1.00, p=0.29 4: t(38)<0.001, p=1.00 8: t(38)=1.00, p=0.44			
V-CC	2: t(38)=2.00, p=0.10 4: t(38)=3.00, p<0.01 8: t(38)=3.00, p<0.01	2: t(38)=0.60, p=0.56 4: t(38)=2.97, p<0.01 8: t(38)=1.94, p=0.06		
V-IC	2: t(38)=1.16, p=0.26 4: t(38)=3.73, p=0.001 8: t(38)=0.50, p=0.62	2: t(38)=0.15, p=0.88 4: t(38)=4.29, p<0.001 8: t(38)=0.50, p=0.62	2: t(38)=0.68, p=0.50 4: t(38)=2.00, p=0.14 8: t(38)=2.48, p=0.02	

Complex Stimuli

Effects of Modality

Table 7: Accuracy

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(38)=1.00, p=0.32 4: t(38)=4.34, p<0.001			

	8: t(38)=3.88, p<0.001			
AV-Test V	2: t(38)=1.64, p=0.11 4: t(38)=0.82, p=0.42 8: t(38)=2.00, p=0.11	2: t(38)=0.41, p=0.69 4: t(38)=4.54, p<0.001 8: t(38)=1.37, p=0.18		
AV-Test A	2: t(38)=1.20, p=0.24 4: t(38)=3.09, p<0.01 8: t(38)=1.66, p=0.11	2: t(38)=0.04, p=0.97 4: t(38)=0.36, p=0.72 8: t(38)=1.71, p=0.10	2: t(38)<0.001, p=0.63 4: t(38)=4.00, p<0.001 8: t(38)<0.001, p=0.99	

Table 8: Reaction Time

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(38)<0.001, p=0.83 4: t(38)=3.15, p<0.01 8: t(38)<0.001, p=0.94			
AV-Test V	2: t(38)=1.66, p=0.11 4: t(38)=0.13, p=0.90 8: t(38)=0.44, p=0.66	2: t(38)=1.53, p=0.14 4: t(38)=2.57, p=0.01 8: t(38)=0.45, p=0.66		
AV-Test A	2: t(38)=0.90, p=0.38 4: t(38)=4.13, p<0.001 8: t(38)=0.02, p=0.98	2: t(38)=1.00, p=0.31 4: t(38)=2.00, p=0.07 8: t(38)<0.001, p=0.93	2: t(38)=2.42, p=0.02 4: t(38)=4.33, p<0.001 8: t(38)<0.001, p=0.69	

Table 9: Capacity

	Visual alone	Auditory alone	AV-Test V	AV-Test A
Visual alone				
Auditory alone	2: t(38)=1.01, p=0.32 4: t(38)=4.32, p<0.001 8: t(38)=3.86, p<0.001			
AV-Test V	2: t(38)=1.83, p=0.08 4: t(38)=1.11, p=0.27 8: t(38)=1.68, p=0.10	2: t(38)=0.58, p=0.56 4: t(38)=4.82, p<0.001 8: t(38)=1.12, p=0.27		
AV-Test A	2: t(38)=0.67, p=0.51 4: t(38)=2.99, p<0.01 8: t(38)=1.54, p=0.13	2: t(38)=0.43, p=0.67 4: t(38)=0.09, p=0.93 8: t(38)=1.73, p=0.09	2: t(38)=0.97, p=0.34 4: t(38)=3.66, p<0.01 8: t(38)=0.22, p=0.83	

Effects of Congruency

Table 10: Accuracy

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(38)=1.00, p=0.30 4: t(38)=1.00, p=0.15 8: t(38)=1.00, p=0.54			
V-CC	2: t(38)=0.50, p=0.62 4: t(38)=4.36, p<0.001 8: t(38)=0.51, p=0.62	2: t(38)=0.90, p=0.38 4: t(38)=2.42, p=0.02 8: t(38)=0.14, p=0.89		
V-IC	2: t(38)=1.29, p=0.21 4: t(38)=2.81, p<0.01 8: t(38)=0.12, p=0.91	2: t(38)=0.21, p=0.84 4: t(38)=1.23, p=0.23 8: t(38)=1.00, p=0.58	2: t(38)=1.00, p=0.33 4: t(38)=1.00, p=0.15 8: t(38)<0.001, p=0.72	

Table 11: Reaction Time

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(38)=1.00, p=0.23 4: t(38)=2.00, p=0.13 8: t(38)=3.00, p<0.01			
V-CC	2: t(38)=2.02, p=0.05 4: t(38)=4.13, p<0.001 8: t(38)=1.64, p=0.11	2: t(38)=2.75, p<0.01 4: t(38)=2.41, p=0.02 8: t(38)=1.09, p=0.28		
V-IC	2: t(38)=0.74, p=0.47 4: t(38)=3.70, p=0.001 8: t(38)=1.96, p=0.06	2: t(38)=1.70, p=0.10 4: t(38)=2.22, p=0.03 8: t(38)=1.00, p=0.35	2: t(38)=1.00, p=0.25 4: t(38)<0.001, p=0.99 8: t(38)<0.001, p=0.77	

Table 12: Capacity

	A-CC	A-IC	V-CC	V-IC
A-CC				
A-IC	2: t(38)=1.00, p=0.36 4: t(38)=2.00, p=0.11 8: t(38)<0.001, p=0.72			
V-CC	2: t(38)=1.00, p=0.52 4: t(38)=4.00, p<0.001 8: t(38)<0.001, p=0.85	2: t(38)=0.47, p=0.64 4: t(38)=2.07, p=0.05 8: t(38)=0.55, p=0.59		
V-IC	2: t(38)=1.61, p=0.12 4: t(38)=3.27, p<0.01 8: t(38)=0.30, p=0.77	2: t(38)=0.63, p=0.53 4: t(38)=1.43, p=0.16 8: t(38)=0.10, p=0.92	2: t(38)=1.18, p=0.24 4: t(38)=1.00, p=0.36 8: t(38)<0.001, p=0.66	