A Dissertation

Multisensory Integration and Crossmodal Interactions:
Exploring the Interaction Between
Visual Attention and Language Comprehension

Eric M. Chiu

Cognitive and Information Sciences

University of California, Merced

5200 North Lake Rd., Merced, CA 95343

ABSTRACT

Convergence of multisensory information can improve the likelihood of detecting and responding to an event as well as more accurately identifying and localizing it. The ubiquitous nature of crossmodal processing is observed in everything from basic signal detection to speech recognition. Although we observe multisensory integration throughout various modalities, this dissertation reviews and discusses research focusing primarily on the relationship between the auditory-linguistic system and visual systems. A series of experiments and simulations presented here show graded effects of crossmodal processing that are reflected in reaction time data and motor output, measured through streaming x-y coordinates from eye-movements. A model simulates and makes predictions about real-time crossmodal processing that argue against the traditional serial and parallel approach to visual attention and supports a perspective with a single underlining mechanism. A purely parallel process is introduced as a means for reconciling both traditional and continuous accounts of visual attention. A broad philosophical discussion follows, in which an integrative and continuous approach to crossmodal processing is proposed and discussed.

BIOGRAHICAL SKETCH

Eric M. Chiu was born in San Francisco, CA in 1983. After receiving an associate of arts degree in liberal arts with an emphasis on psychology from De Anza College he moved to the University of California, Santa Cruz and obtained a bachelors degree in psychology. Following his bachelor's degree he worked in the field of mental health as a caseload mental health worker before entering the Cognitive and Information Sciences department at the University of California, Merced where he plans, with a lot of luck, to complete his graduate studies.

ACKNOWLEDGEMENTS

I am eternally grateful to Michael Spivey for providing me never-ending encouragement while expertly guiding my development as a cognitive scientist. Thank you also to the community of professors, graduate students, and undergraduate researchers in both UC Merced Cognitive and Information Science and UC Santa Cruz Psychology for providing a nurturing environment in which I have had the pleasure to learn and grow. I would like to especially acknowledge Markie Johnson, Norma Cardona, Mauricio Cifuentes, Courtney Griffin-Oliver, Lydia Goes, Maria Vega, and Lilly Rigoli for assisting with the experiment design process and data collection and Andreas Kolling for assisting with programming. Special thanks to Bruce Bridgeman for encouraging me to continue my academic career and providing me the resources and opportunities to do so. I would also like to thank my parents and brother for their loving support throughout my studies.

This dissertation, “Multisensory Integration and Crossmodal Interactions: Exploring the Interaction Between Visual Attention and Language Comprehension,” by Eric M. Chiu is hereby approved for Degree of Doctorate of Philosophy in Cognitive and Information Sciences:

_____

Michael J. Spivey, Ph.D. (Advisor)

_____

David C. Noelle, Ph.D. (Committee Member)

_____

Jack L. Vevea, Ph.D. (Committee Member)

Cognitive and Information Sciences

University of California, Merced

2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

The Contiguity of Mind

Introduction

Most everyday objects and events generate multisensory inputs that appear concurrently or with some amount of overlap. Taking advantage of the shared information in these signals, rather than processing them individually, can be advantageous for task performance and learning (de Sa & Ballard, 1998; Calvert, Hansen, Iversen, & Brammer, 2001; Soto-Faraco, Foxe, & Wallace, 2005). In basic signal processing, such as air traffic control, it is imperative that a sensor (e.g., radar) detects the presence of an external event. The current technology involved in this sort of event detection generally depends on a single sensor, which can be very effective when the event to be detected produces a strong signal with a unique signature and few other competing signals that may activate the sensor and confound operators. Unfortunately this is not the case. To be certain that an event is not missed due to a weak or ambiguous signal or a noisy environment, the threshold for activating the sensor must be set very low making this approach to event detection very limited because it generates far too many signals, which effectively makes the information from such a sensor extremely difficult to manage. This predicament can be remedied by functionally coupling two or more sensors, each tuned to a different form of environmental energy (e.g. visible light and infrared light, radar and sonar, or light and sound, etc.). By specifying the criteria for activation and temporally synchronizing each sensor before activating

their common central processor, the thresholds of these sensors can be set very low while still minimizing false-positives and accurately disambiguating events.

Human perception has evolved in a similar manner. We automatically integrate information delivered by a variety of sensory systems (e.g., visual, auditory, somatosensory, etc.) each tuned to detect different forms of environmental energy to create a single percept of the world. Correspondingly, we observe countless examples of automatic perceptual interactions in human cognition across various systems including but not limited to vision, audition, touch, and assorted linguistic systems. One of the most famous examples of a perceptual interaction is the McGurk effect where visual input alters linguistic percept. The McGurk effect is experienced when you see a televised face repeatedly saying "ga-ga," but synchronized with the mouth movements the audio stream actually delivers "ba-ba," which when observed together constructs a convincing percept of hearing "da-da" (McGurk & MacDonald, 1976). This finding illustrates a dynamic and immediate integration of visual and linguistic processing, which is a great example of multisensory integration and the topic of this examination.

Multisensory Integration

Although each of our senses is tuned to different forms of environmental energy, it is now widely accepted that much of our sensory cortexes are fundamentally multisensory in nature (Ghazanfar & Schroeder, 2006). This is contrary to the early widespread acceptance of Fodorian modularity, a concept that argues for separate structures in the mind with specific functional purposes and information encapsulation (Fodor, 1983). In human cognition the field of crossmodal interaction, the study of perception that involves two or more sensory modalities, and multisensory integration is still young; even now little is known about the structure of the mind

and the specifics of information processing that allows us to integration unimodal cues into a cohesive perception of the world.

In the past two decades of psychological experiments, we have learned that when signals from different sensory modalities integrate it is done optimally for maximal activation. This is done in a way as when two or more cues from different modalities appear in close spatial and temporal proximity they are combined in the brain in a way so that the more statistically reliable cue is weighted more strongly (Alais & Burr, 2004). For instance, we largely rely on the visual system for spatial acuity (DeValois & DeValois, 1993), while relying on the auditory system for perceiving temporal events (Tyler & Hamer, 1990; Viemeister & Plack, 1993).

This weighting strategy results in a variety of perceptual anomalies or crossmodal illusions such as the *ventriloquism effect*, which occurs when minute discrepancies are introduced to a perceptual event changing the temporal and/or spatial relationship among multisensory stimuli that are usually derived from the same event (Howard & Templeton, 1966). The illusion exists because the brain weighs the superior spatial resolution of the visual system more strongly than that of the auditory system when determining the location of an auditory-visual event, thus causing an observer to perceive the speech originating from the doll rather than the skilled performer (Driver & Spence, 2000). On the other hand, when a single flash of light is accompanied by two auditory beeps, the brain weighs the superior temporal resolution of the auditory system more heavily causing observers, more often than not, to perceive the single flash of light as two flashes (Shams, Kamitani, & Shimojo, 2000). Moreover, it has been demonstrated that it can be advantageous for task performance and for learning to take advantage of both of the shared and unique information in these signals, rather of processing them with

segregated modules (Zellner & Kautz, 1990, de Sa & Ballard, 1998; Calvert et al., 2001; Soto-Faraco et al., 2005).

The term *modality,* which is most commonly used to refer to one aspect of perceived stimuli such as light, sound, taste, or another sensory event but the term modality is not limited to only sensory experiences. Language is another modality, albeit non-sensory, that has been found to integrate with sensory processes. In addition to the aforementioned McGurk effect, studies have also found that hearing the name of a letter prior to a detection task (e.g., hearing "emm" when detecting the letter "M") improves perceptual sensitivity and detection; visual pre-cues of to-be-detected stimulus and unmatched auditory cues were not found to improve detection in this study (Lupyan & Spivey, 2008; 2010). These results demonstrate an immediate top-down conceptual influence on visual recognition, which implies that visual perception depends on more than simply what something looks like but also what it represents. Another intriguing example of the immediate integration of vision and audition comes from a study by Calvert and colleagues (1997) that found auditory cortex activation of a skilled lip-reader during silent lip-reading. This is surprising because there is no actual auditory information input to activate the auditory cortex. The finding suggests that the skill of lip reading appears to recruit linguistic representations in order to understand what was said, which is closely intertwined with auditory processing (Calvert et al., 2001). Findings like this further demonstrate the multimodal integrated nature of sensory systems once thought to be modular.

The ability for humans and other species to integrate divergent sensory information is extremely fascinating and complex. Take for instance the fact that although vision and audition detect distinct environmental energies (light waves vs. sound waves, respectively) that vary in arrangement (retinotopy vs. tonotopy, respectively), one would assume integration would be

incompatible or at the very least cognitively taxing yet we are able to immediately and automatically integrate this information with little to no effort (Spence & Driver, 1996). Synesthesia is a great example where divergent information, whether sensory or conceptual, automatically integrates.

Synesthesia, from the ancient Greek syn, "together", and aesthēsis, "sensation," is when stimulation of one sensory modality (e.g., audition) leads to the automatic and involuntary experience in a second sensory modality (e.g., vision). For example, sonogenic synesthesia is where hearing music automatically provokes intense visual experiences or somatosensory paraesthesias. Neurophysiologically, synesthesia reflect a fusion of sensory experiences via association phenomena, in which independent groups of neurons are activated in close temporal proximity to one another via long chains of synaptic connections. Their concurrent activity produces a perceptual synthesis after repeated pairings much like any other conditioned experience.

The occurrence of synesthetes appears to cut across a variety of social milieus and personalities, and exhibit underlining similarities between synesthetes. Some theorize that these underlining similarities are a result of common couplings regularly experienced in an individual's environment especially during early cognitive development. If the statistical prominence of underlining similarities is indeed learned (e.g., in a kindergarten classroom where the letter "A" is almost always coupled with a "red apple," binding the letter "A" with the color "red") then it suggests a cultural linguistic difference in grapheme correspondences. Unfortunately, research in this field has historically been stunted by false claims of synesthesia, thus without the capacity to evaluate their physical basis there was no way to differentiate true

synesthetes from false. Only recently has research in this field been able to progress. I am excited to learn what they will uncover.

Shifts in sensory attention often precede motor action, thus the study of attention is an important tool in multimodal literature. Spatial attention paradigms, such as Posner's (1978) cuing paradigm, are commonly used to study multisensory integration. Spatial attention can be separated into two major areas, overt exogenous attention and covert endogenous attention. For example, in vision, changes in spatial attention can occur overtly with eye movements or covertly with the eyes stationary. Within the eye, only a relatively small section known as the fovea is capable of high visual acuity. The fovea is necessary during actions such as recognizing facial features or reading. As a result, the eyes must continually make saccades, small jerky ballistic movements, to direct the fovea to the necessary locations to perform desired action but before the eyes move to a target location overtly, attention shifts to that location covertly. As a result, exogenous and endogenous spatial attention in vision can be studied separately by controlling eye movements.

Traditionally, studies of multisensory interactions have been limited to exogenous spatial attention but with the advent of relatively low cost eye tracking technology that allows us to record overt eye movements, research in the past four decades has shifted focus from overt attention to endogenous spatial attention (Spence & Driver, 1994). The field of endogenous spatial attentional subsystems has been populated by three main possible architectures: entirely supramodal, entirely modality-specific, and a blend of the two theories (Farah, Wong, Monheit, & Morrow, 1989). The first, entirely *supramodal*, is a perspective where perception is modulated as a function of location across all modalities. This 'supramodal' attentional subsystem allocates salience to locations in space regardless of the modality of the target being

attending. The traditional view of this hypothesis presumes that the size of spatial-attention effects should always be similar across all modalities, regardless of which sensory modality currently played the primary task-relevant role. Though there has been evidence of larger spatial attention effects within modalities that are primary to a task than for a secondary modality, these findings are reconciled by suggesting that the interaction between supramodal spatial selection and task-relevant modalities are perhaps nonlinear (Spence & Driver, 2004).

The second, *entirely modality-specific*, hypothesis claims separate encapsulated modality-specific spatial-attentional systems that operate independently in their individual representations of space whether it is auditory, visual, or otherwise. In this hypothesis when spatial salience increases for a specified location it is done so independently by each modality. According to this idea, synergies would not exist between the separate modalities. Thus greater activation of vision in the right spatial field would have no effect on auditory activation in either the right or the left spatial field. As a result, any observed relationship between modalities in this case would be entirely coincidental.

The third blend of the two previous possibilities is an intermediate of the two extremes takes two possible manifestations. The first manifestation proposed by Posner (1990) is a *unimodal-plus-supramodal hybrid* perspective, which postulates a system that has unimodal subsystem that map on to a higher-level supramodal component within a hierarchical attentional network. There is neurophysical evidence that such a supramodal map may exist in the posterior parietal cortex and the superior colliculus (Farah et al., 1989; Driver & Spence, 1998). For better understanding let us envision a supramodal map of attentional salience that integrates an unimodal map of vision from the extrastriate visual cortex (Itti & Koch, 2001; Parkhurst, Law, & Niebur, 2002) and an unimodal map of audition from the auditory cortex and the inferior

colliculus (King, 1999). This hybrid perspective of spatial attention can be captured by a three-dimensional topographically arranged layers of neurons (fig. 1.1) where the height dimension (z-axis) represent the salience of that specific space such that areas where spatial activation correspond in the unimodal maps would produce a subsequently larger activation in the supramodal map (Spivey, 2007). It is important to note that the integration of these different sensory inputs is not solely a feed-forward process but rather a bidirectional interaction.

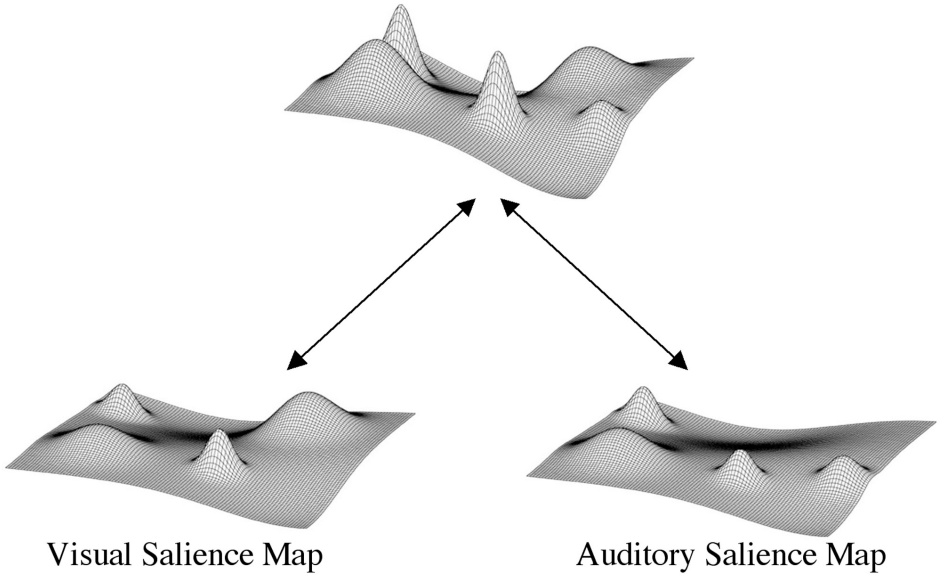

Figure 1.1: Depiction of a supramodal hybrid theory of attentional saliency. A supramodal salience map receives input from and sending feedback to unimodal salience maps. Note that areas with overlapping activation in the unimodal maps would produce a subsequently larger activation in the supramodal map (Figure adapted from Spivey (2007) with permission.)

The second manifestation proposed by Spence and Driver (1996) is described as a "*separable-but-linked*" perspective that postulates that there are indeed separable modality-specific attention systems, but with links such that auditory orienting tends to result in visual orienting to the corresponding location in visual space and vice versa (Spence & Driver, 1994).

The separate-but-linked hypothesis predicts that attention can be simultaneously directed to difference positions in two modalities under at least some circumstances with or without performance cognitive costs, a phenomenon completely rules out by a purely supramodal perspective. In this perspective the synergies between modalities such as visual-auditory spatial attention appear to change depending on the task. Whatever the mechanism may be there is clearly a need for cognitive science to study crossmodal interactions and multisensory integration as part of the endeavor to understand human cognition and the mind.

Vision, Audition, and Attention

Now we shift our discussion to attention in vision. Historically the visual system has been thought of as a functionally independent cognitive process (Fodor, 1983), but recent research demonstrates a more dynamic and immediate integration of visual information with information from other modalities (McGurk & MacDonald, 1976; Shams et al., 2000). For instance, it has been demonstrated that while observing a leftward and a rightward moving circle animated on a computer display, the type of sound delivered when they pass through each other will influence how this event is perceived. If the observer hears a *"*whoosh*"* sound just as the circles pass through each other, they will appear to travel past one another on slightly different depth planes. However, with identical visual input, if the observer hears a *"*boing" sound the two circles will appear to bounce off of each other and reverse their respective directions (Sekuler, Sekuler, & Lau, 1997).

Additional evidence of the distributed functioning of the visual system is seen with a series of experiments by Spence and Driver (1996) that investigated endogenous covert spatial orienting in hearing and vision using a modified Posner cuing paradigm (Posner, 1980) where

observers judged the elevation (up vs. down) of auditory or visual targets that appeared either on their left of right visual field. They demonstrated that when observers were informed that targets were more likely on one side in one or both modalities elevation judgments were faster on that side regardless of modality or laterality and even if the modality of the target was uncertain. This is consistent with the previously presented supramodal theory of attention because the results suggest that observers directed endogenous attention wholly, meaning all modalities (e.g., vision and audition), to the side that they were informed is most likely for the target to appear. However, they also demonstrated that it was possible to "split" auditory and visual attention when targets in the two modalities were consistently expected on opposite sides throughout a block but at a cost. Covert orienting effects were larger when targets were expected on the same side in both modalities, suggesting that endogenous covert attention may not operate within an exclusively supramodal system but exhibits strong spatial synergies between visual and auditory attention (Spence & Driver, 1996).

While such examples of interactions between vision and audition are extremely interesting and informative, recent advances in methodological techniques have greatly aided in progressing the field of cognitive science by providing us with novel insight into human cognition and perception. Dense-sampling techniques such as eye-tracking and reach-tracking measures, like mouse-tracking, allow us to develop a more detailed illustration of the temporal dynamics of cognitive processes such as with the mechanisms involved in how visual information immediately impacts lexical and sentence processing among a myriad of other mechanisms. A great example of the use of this innovation comes Tanenhaus and colleagues' (1995) study that investigated the rapid mental processes that accompany spoken language comprehension by recording eye movements while observers followed instructions and

manipulated real objects. They found that visual context influences spoken word recognition and mediated syntactic processing even during the earliest moments of language processing (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Beginning with this pioneering study, eye-tracking has been extensively used to investigate the real-time interactions of visual information and language comprehension.

Early word recognition findings are another example where dense-sampling methods have reconciled a cognitive process that was once deemed equivocal. These initial reaction-time studies suggested that as a word is heard it is initially ambiguous with other words that share similar sounding onsets, implying that even during the earliest moment of processing visual context might influence word recognition and syntactic processing (Zwitserlood, 1989). This perspective proposes that for a brief period after the onset of a word all words beginning with the same phonemic input compete but as more phonemic input is received the target becomes less and less ambiguous, thus competing distractors drop out as a word unfolds over time (Marslen-Wilson, 1987). To test this hypothesis Allopenna, Magnuson, & Tanenhaus (1998) (see also, Spivey-Knowlton, 1996) recorded eye movements as observers heard and responded to instructions like, “Pick up the beaker” while viewing a visual display (Allopenna et al., 1998). The visual displays contained four items: the target item (e.g., a beaker), an onset-competitor (e.g., a beetle), a rhyme competitor (e.g., a speaker), and an unrelated referent (e.g., a carriage). The results revealed that during the first half of the spoken target word, the probability of fixating the target or competitor equally and gradually increased but around the offset of the spoken target word, the proportion of fixations to the target began to rise sharply and subsequently decreasing the proportion of fixations to the competitor.

Thus, early in the presentation of an auditory stimulus, before the target has been uniquely identified, competition between the partially active representations manifests itself in eye movement patterns. Moreover, the data revealed a greater probability of fixations to the rhyme competitor than to the neutral distractor object, which is congruent to the rhyme competitor effects predicted by McClelland and Elman's (1986) interactive neural network simulation of speech perception. This "TRACE" model is named so because of the network of units that form a dynamic processing structure termed "the Trace," which serves as both the model's perceptual processing mechanism and as the system's working memory.

Similar to words, some sentences are just as momentarily ambiguous across time. Many early investigations in the processing of temporarily ambiguous sentences looked at sentences in isolation with results supporting a modular process. For instance, in the sentence "Since Jay always jogs a mile doesn't seem far," inflated reading times were observed when readers encountered the disambiguating word, "…doesn't…" (Frazier & Rayner, 1982). These early researchers postulated that the increased reading time was the manifestation of an encapsulated syntactic processing module separate from other perceptual and cognitive systems, accordingly arguing for its autonomy from other information sources, such as semantics and visual information. Just like with words, dense-sampling techniques tell a different story about the mechanisms involved in processing these sentences illustrating a drastically different process for syntactically ambiguous sentences in conjunction with a visual scene than previously postulated from reaction time data (Tanenhaus et al., 1995).

Tanenhaus and colleagues (1995) observed that when listeners experienced the *garden-path effect* when they heard a temporarily ambiguous sentence such as "Put the apple on the towel in the box" while viewing a scene containing four objects: an apple (target object), a towel

(incorrect goal location), a box (correct goal location), and a flower (neutral unrelated referent), they experience the *garden-path effect*. The garden-path effect occurs when observers temporarily interpreting a sentence incorrectly before migrating and ultimately settling on the correct meaning. In the case of "the apple on the towel," they temporarily interpret "…on the towel…" as the destination of the putting event, only to later realize that this parse is incorrect. In these experiments, the garden-path effect manifests itself as inflated reading times at the point of disambiguation, exhibiting itself as an increased probability of saccades to the incorrect destination (the towel). Trials containing unambiguous sentences like "Put the apple *that's* on the towel in the box" did not exhibit an increased probability of looks (Tanenhaus et al., 1995). These findings are one of the many examples of a connection between linguistic processing, whether it is visual or auditory, and visual motor output.

Links between audition and vision are plentiful but insofar the precise relationship has eluded us. For instance, an eye-centered frame of reference is most common because vision is the sensory modality with the highest spatial acuity. As a result, many aspects of auditory spatial perception seem to depend on eye position or the location of a visual stimulus (Lewald & Ehrenstein, 1996; Lewald, 1997, 1998; Shams et al., 2000; Shimojo & Shams, 2001). For example, Lewald and Ehrenstein (1996) demonstrated that observers shifted judgments of sound bursts toward eye position in both focused gaze, while fixated on target, and unfocused gaze, in darkness. Supporting the idea that, in humans and other animals, vision plays a key role in calibrating the auditory system's capacity to localize a sound source, explaining auditory neurons that exhibit spatially selective receptive fields that shift with eye position. This specific synergy leads to such perceptual illusions as the previously mentioned ventriloquism effect.

I have reviewed a number of ways in which the auditory system, including linguistic processing, is impacted by vision as well as how auditory processing influences the vision. I have also shown the need to look beyond reaction time data, as it is crucial when investigating how cognitive processes unfold in real time. Dense-sampling approaches such as eye-tracking and reach-tracking provide a window into how ambiguous stimuli such as partially active lexical and syntactic representations are activated and compete over time. These methods also allow researchers to investigate the way objects in a visual scene can immediately guide language processing and vice versa.

Visual Search

Two basic phenomena define the topic of visual attention. The first basic phenomenon is limited capacity for processing information. Humans are inherently limited capacity creatures and as a result the aforementioned crossmodal interactions bestow considerable behavioral advantages. At any given time, only a small amount of the existing information on the retina can be processed and mapped onto motor output. Correspondingly, giving attention to any one stimulus leaves less processing for any others. The second basic phenomenon is selectivity, the ability to filter out unwanted information. Correspondingly, one is aware of attended stimuli and largely unaware of unattended ones. Thus, accuracy in identifying an attended stimulus may be independent of the number of non-targets in a display (Duncan, 1980).

Traditionally two divergent perspectives, originating from Treisman & Gelade's (1980) *Feature Integration Theory*, have populated the field of attention in visual search. According to the Feature Integration Theory, when executing a search of objects the first stage that describes the beginning of the perceptual process is called the *preattentive stage*, where the stimulus is

analyzed for details such as shape, color, orientation and movement, with each aspect being processed in different specialized areas of the brain. Each of these brain areas creates feature spatial maps of each perceived feature. During this stage perception occurs automatically, unconsciously, and effortlessly; meaning observers are not aware of this process since it occurs early in perceptual processing before the to be detected stimulus becomes conscious (Treisman & Gelade, 1980).

Following the first preattentive stage of the Feature Integration Theory is the second *focused attention stage*, where individual features are combined to create a percept of the stimulus as a whole. In this stage, attention is used to combine the individual features maps, which is succeeded by selection of that object within a spatial "master map." This master map of locations contains all the locations in which features have been detected and is generated by integrating the multiple feature maps. When attention is focused at a particular location on the map, the features currently in that position are attended to and stored in "object files." Identification of an object occurs when the attended object is familiar, or in other words an association is made between prior knowledge and the attended object.

Researchers often refer to patients suffering from a form of Balint's syndrome (oculomotor apraxia as suppose to the other two forms of Balint's syndrome, which are simultanagnosia and optic ataxia) as evidence of this stage of perceptual processing (Posner, Walker, Friedrich, & Rafal, 1984; Humphreys & Riddoch, 1993; Desimone & Duncan, 1995). Due to extensive bilateral damage to the parietal lobe, these patients suffer from a severe neuropsychological impairment often referred to as the psychic paralysis of gaze are unable to voluntarily guide eye movements and focus attention on individual objects. Patients suffering from Balint's syndrome have the inability to focus attention long enough to combine the features

of a stimulus that requires combining multiple features, which provides support for the focused attention stage of this theory.

Treisman and Gelade's (1980) Feature Integration Theory distinguishes between two perspectives for processing two kinds of visual search arrays. First perspective is the initial *parallel processing stage* (competitive), which institutes the aforementioned independent spatial maps that identify the location of features in a visual field. The first type of search array is termed a *single-feature search* or simply "feature search" and accounts for the majority of observations of the parallel processing perspective where responses are based on a single map of partially active representations of objects simultaneously contending for probabilistic mapping onto motor output. These feature search arrays often inducing what is called a perceptual "pop-out" effect, where the unique target object that differ from distractor objects by the only feature (e.g., color, orientation, intensity, etc.) in the array appears to pop-out from the group (Treisman & Gelade, 1980; Treisman & Gormican, 1988).

A *conjunction search*, the second sort of search array, uses multiple features thus multiple maps would be needed to identify the presence and subsequently map the location of each feature in a visual field. Accordingly, decisions in conjunction search require combining information from multiple feature maps and as previously mentioned this integration requires a process of "focal attention," which is only accurate and reliable when dealing with one array element at a time. In this model, the perspective responsible for processing a traditional conjunction-search is referred to as a *serial search process* (attentive), which claims that observers allocate complete attentional resources discretely and wholly to individual objects one at a time (Treisman & Gelade, 1980; Treisman, 1988).

As a reaction to Treisman's Feature Integration Theory, Wolfe (1994) proposed perspective in a model called the *Guided Search Model 2.0*. Similar to the Feature Integration model this model distinguishes between two different stages. According to this model the first initial largely parallel *preattentive stage*, processes information about basic visual features (e.g., color, motion, various depth cues, etc.) across large portions of the visual field. Subsequently, a *limited-capacity stage* performs additional operations that are more complex (e.g., face recognition, reading, object identification, etc.) over a smaller more limited portion of the visual field. Deployment of the limited attentional resources is guided by the output of the earlier parallel process effectively making it a bottom-up and top-down process, Wolfe points out that this is the heart and primary discerning factor of the Guided Search Model 2.0. The information acquired through this bottom-up and top-down processing is then ranked according to priority. The priority ranking is what guides visual search and effectively makes the search more efficient (Wolfe, 1994).

More recently, findings have demonstrated that instead of two apparently dichotomous perspectives, parallel and serial processing, attention in visual search may be better described as a single process of graded enhancement of feature salience, which is supported by observations of gradual improvements of efficiency in visual search tasks (Olds, Cowan, & Jolicoeur, 2000a; 2000b; 2000c). In a series of experiments Olds and colleagues (2000a; 2000b; 2000c) observed facilitatory effects as a result of very brief presentations (less than 100 ms in some conditions) of displays with only single-feature distractors before transitioning to conjunction-search displays. Although observers' responses were not as fast as with pure "pop-out" displays, they observed a graded improvement of search efficiency.

Furthermore, an examination of 2,500 visual search studies each by Wolfe (1998) with a few hundred trials (totaling approximately 1 million trials) failed to find a bimodal distribution of search efficiency, despite including a wide variety of search tasks. Eckstein (1998) found no evidence supporting the existence of an initial serial mechanism where information binds across feature dimensions when low-level effects such as physical similarity of target and distractor, element eccentricity, and eye movements where carefully controlled. Instead, the study showed that the previously observed conjunction search dichotomy is likely the result of the noisy neural processing of features in the human visual system, which is well supported by physiological recordings of cells in the visual cortex (Maunsell & Newsome, 1987; Tolhurst, Movshon, & Dean, 1983).

A study by Maioli, Benaglio, Siri, Sosta, and Cappa (2001) found no differences in a visual search task where observers had to locate a "Q" among "O" distractors or vice versa. Furthermore, they found that the only accurate predictor of reaction time was the number of saccades made during a search, which were discovered to be independent of the number of stimulus items. Correspondingly, Watson, Brennan, Kingstone, and Enns (2010) found that a passive cognitive strategy, that is allowing the target to "pop" into mind rather than trying to actively guide attention, increased search efficiency by decreasing the number of necessary saccades and improving the use of information from each fixation despite delaying onset of the initial eye movement (Watson, Brennan, Kingstone, & Enns, 2010).

To account for these findings, Maioli and colleagues (2001) argue for a time-limited competitive model for attention in visual search, in which both parallel and serial processing mechanisms are integrated, which provides a unified conceptual framework for all types of visual search. This perspective is supported by identification of neural mechanisms that are

mediated by biased competition in the extrastriate visual cortex, forming a compelling argument against the serial processing perspective and for a completely parallel processing perspective, which we have learned claims attention is better characterized as a function of partially active representations of objects simultaneously contending for probabilistic mappings onto motor output (Desimone & Duncan, 1995; Desimone, 1998; Reynolds & Desimone, 2001).

Findings like Old and colleagues' (2000a; 2000b; 2000c) "search assistance" along with the various other studies that have been presented (Eckstein, 1998; Wolfe, 1998; Maioli et al., 2001; Watson et al., 2010) has largely shifted the serial-parallel dichotomy terminology of visual search efficiency with a dialogue that is graded and continuous (e.g., Nakayama & Joseph, 1998). Further support for this trend comes from work by Spivey, Tyler, Eberhard, and Tanenhaus (2001) that discovered another type of "search assistance" phenomenon. Observers in an *Audio/Visual Concurrent* (A/V-concurrent) condition, where the conjunction-search display is presented concurrently with target identity via auditory linguistic queries (e.g. "Is there a red vertical?"), displayed dramatically improved search efficiency when compared to an *Auditory-First* control condition, where the same spoken query of target identity was provided prior to visual display onset. The findings suggest that in A/V-concurrent trials upon hearing the first-mentioned adjective in the spoken query visual attention is able to begin the search with only that feature, thus initiating the process more efficiency in a single-feature like search. Then after hearing the second adjective, several hundred milliseconds later, observers can subsequently quickly identify the target among the now smaller more salient subset of objects.

This finding has been repeatedly reproduced and extended by Reali, Spivey, Tyler, and Terranova (2006) as well as Chiu and Spivey (2012) in a variety of follow up experiments. For instance, despite altering the order of the adjective delivery from color-first to orientation-first, a

significant improvement in visual search efficiency continued to be observed when the identity of the conjunction target was delivered incrementally via a spoken target query while the stimulus display was visible but not when delivered prior to stimulus onset. Interestingly, Gibson, Eberhard, and Bryant (2005) found that with faster speech (4.8 syllables/second vs. 3.0 syllables/second) the A/V-concurrent condition no longer provided an enhanced efficiency effect on conjunction-search tasks, indicating that linguistic mediation of visual search is sensitive to speech rate.

More recently, experiments by Jones, Kaschak, and Boot (2011) used eye-tracking to examine an alternative perspective to one that proposes search efficiency is increased due to language enhancing perceptual processing. Jones and colleagues (2011) observed eye movement patterns that suggests previously observed improvements in search efficiency with concurrent speech was not likely due to linguistic enhancement of perceptual processes but rather from delaying the onset of target-seeking eye movements. They explicate the findings by Gibson et al. (2005) are better explained by this "preview" of search display because slower speech provides observers with additional search display viewing time, which affords additional information about potential target locations independently of the information conveyed by auditory linguistic speech stream.

It is clear that the field of visual search continues to be a hotly debated topic that remains to have many mysteries yet to be solved. Whatever the true multimodal relationship between linguistic processing and visual attention may be or how to best describe attention in visual search with either the Guided Search Model 2.0 (Wolfe, 1998), the Feature Integration Theory (Treisman & Gelade, 1980), or a third purely parallel perspective (Maioli et al., 2001). The following series of studies is part of a research program that accompanies findings like the one

mentioned above by Spivey et al. (2001) that explores the degree to which the incremental processing of spoken words in a full sentence can interact with concurrent visual search processes.

CHAPTER TWO

Replication and Semi-concurrent Experiment with a Localist Attractor Model

Experiment 1: Replication Experiment

In this experiment, I replicated the design of Spivey et al. (2001: Experiment 1) and Reali et al. (2006: Experiment 1), with the exception that I utilized a blocked between-subjects design to rule out any concerns about observers noticing different types of trials and developing search strategies based on that knowledge. This new design also allows us to compare auditory-first control trials to novel types of trials in later experiments (see Experiment 2). Despite the more controlled and less statistically powerful (power = 1 - β) blocked experimental design, I expect to reproduce the core effect from Spivey et al. (2001) and Reali et al. (2006) where the concurrent onset of the visual search display and the target-identifying auditory query elicits a more efficient search strategy compared to trials when target-identifying auditory queries are presented prior to visual search displays. Replicating these results with this new design will greatly support the previously proposed notion that observers utilize concurrent delivery of auditory and visual information to improve search strategies (Spivey et al, 2001; Reali et al, 2006).

Method

Participants

One hundred and sixty-seven University of California, Merced undergraduate students received course credit for participating in this experiment. Participants were randomly assigned to one of two slightly different conditions: 90 participated in the auditory-first control condition and 77 participated in the A/V-concurrent condition. Fourteen participants in the auditory-first condition and 17 participants in the A/V-concurrent condition were unable to perform the task at above 80% accuracy and were omitted from the analysis. For this experiment and all remaining experiments in this dissertation all incorrect responses and trials with reaction times 2.5 interquartile ranges from the median were also omitted from the analysis. Utilization of IQR for data culling over standard deviation (SD) was chosen for its superior resistance to the influence of outliers (McCluskey & Lalkhen, 2007). The participants in this and all subsequent experiments were naïve as to the purpose of the experiments and all reported normal hearing as well as normal or corrected-to-normal vision.

Stimuli and Procedure

The experiment was composed of two slightly different types of trials, auditory-first trials and A/V-concurrent trials. Observers were randomly assigned to one of the conditions and participated in a 32 trial "practice block" that was not part of the final analysis before participating in a 96 trial "experiment block" that was used in the final analysis.

Participants in the auditory-first condition were presented with the entire target query via spoken query (e.g., "Is there a red vertical?") prior to visual display onset. Participants in the A/V-concurrent condition were viewing the visual search display when hearing the adjectives in the target query. The same female speaker recorded all speech files for this experiment and all following experiments in this dissertation that utilize an auditory linguistic query. Each speech

file had an identical 1 second preamble recording, “Is there a...” spliced onto the beginning of each of the four target queries. The two descriptive adjectives (color: “red” or “green” and orientation: “vertical” or “horizontal”) averaged 1.5 s. Each stimulus bar, in this experiment and subsequent experiments, subtended 2.8˚ X 0.4˚ of visual angle and neighboring bars were separated from one another by an average of 2˚ of visual angle (see fig. 2.1). Participants sat comfortably with their backs against a stationary chair such that their eyes were a measured distance of 57 cm from the display (at a 57 cm viewing distance from a display 1 cm on the display is equivalent to 1 degree of visual angle) in order to control for size of objects on their retina (Hubel, 1988). A more natural sitting position was opted over the use of any sort of restraint such as a chin rest because small difference in viewing distance from trial to trial translates to a relatively small change in visual angle, thus no other apparatus was used to control for viewing distance.

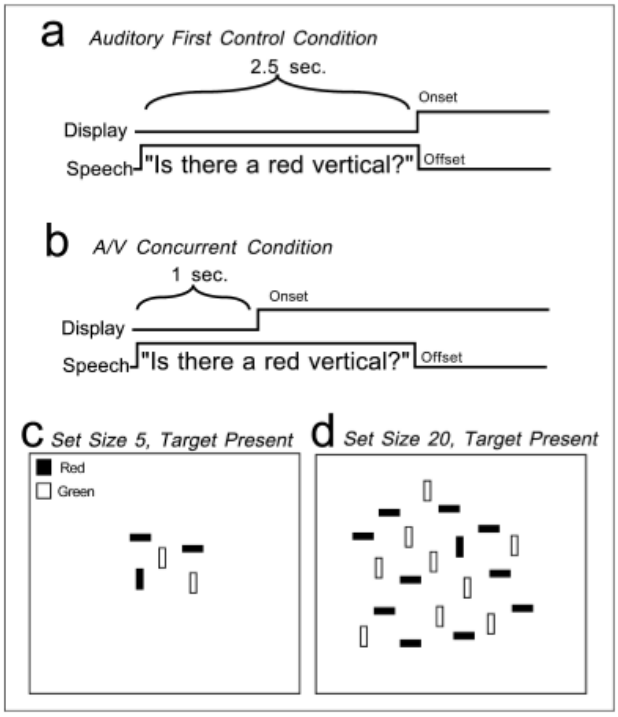

Figure 2.1: Examples of the auditory and visual stimuli. In the auditory-first control condition (a) the onset of the visual display coincided with the offset of the spoken target query. In the audiovisual-concurrent (A/V-concurrent) condition (b), the onset of the visual display coincided with the onset of the first target-feature word in the spoken query. The example displays show target-present trials with a set size of 5 (c) and 10 (d). In these displays, the target is a red vertical bar, which is accompanied by vertical green distractor bars and horizontal red distractor bars. (Figure adapted from Spivey et al. (2001))

Participants were instructed to respond to each display as quickly and accurately as possible by pressing the labeled "Yes" button if the queried object was present in the display and the labeled "No" button if it was absent. Participants initiated each trial manually by pressing the space bar. The keyboard keys "1" and "0" were used for absent and present responses, respectively, allowing participants to comfortably rest a finger from each hand on the response keys, as instructed, while keeping a thumb on the space bar. The distance from response keys to the space bar was 6 cm. A fixation-cross preceded the onset of the visual display in order to direct participants' gaze to the central region of the display. Half of the trials were target-present and half with target-absent. Set sizes of 5, 10, 15, and 20 were used. Based on the two target features (color: red or green, and orientation: vertical or horizontal) four unique targets appeared equally and randomly throughout the trials. This design is utilized throughout the experiments in this dissertation. The duration of the entire experiment was approximately fifteen minutes. Two 20" Apple iMac's, in conjunction with noise minimizing headphones, were used to run this and all following experiments. Programing and execution of this and all following experiments was completed with Mathworks Matlab software. No additional software packages were used.

Results and Discussion

A hierarchal linear model (HLM), which accounts for the unbalanced number of subjects by condition and the repeated measures design, was used for this analysis along with the analysis for Experiment 2. The naturally positively skewed raw reaction time data was log-transformed in order to fulfill the assumption of a normal distribution for all inferential statistics. However, we report descriptive statistics (slopes and intercepts of reaction times in milliseconds) from an untransformed HLM.

In this experiment, we replicated previous findings demonstrated by Spivey et al. (2001) and Reali et al. (2006) with a between-subjects design. Figure 2.2 shows the RT-by-set-size functions for target-present (filled symbols) and target-absent (open symbols) trials in the A/V-concurrent (triangles) and auditory-first (circles) conditions. Next to each regression line for all results figures is the best-fit linear equation and the proportion of variance accounted for (r^2). The error bars indicate standard error of the mean. In the auditory-first condition, the RT-by-set-size function was highly linear in both target-present, $r^2 = .971$, and target-absent trials, $r^2 = .996$, as typically observed in standard conjunction search tasks. Similarly, the RT-by-set-size function for the A/V-concurrent condition was highly linear for target-present trials, $r^2 = .915$, and target-absent trials, $r^2 = .947$. Likely due to the delay of complete delivery of target identity by approximately 1.5 s relative to the auditory-first condition, overall mean reaction time (as well as y-intercepts) were significantly slower in A/V-concurrent conditions for both target-present, $t(59) = 3.28$, $p = .001$, and target-absent, $t(59) = 3.03$, $p = .003$, trials. Mean accuracy across target-present and -absent trials after culling for outliers was 93.0% for the auditory-first

condition and 94.5% for the A/V-concurrent condition, which is similar to previous observations (Spivey et al., 2001; Reali et al., 2006).

For this and all following experiments the most important analysis is the comparison of the slopes of functions relating reaction time to set size. This slope value is an indicator of how efficient the search process is; that is, how much it resembles a serial process where each new distractor object increases reaction time by a sizeable fixed duration, or how much it resembles a parallel process where each new distractor object increases reaction time by little or no amount. The slopes of the RT-by-set-size functions reveal that A/V-concurrent conditions produce more efficient visual search compared with the auditory-first conditions (see fig. 2.2). An HLM analysis revealed significantly shallower slopes for the A/V-concurrent condition compared to the auditory-first condition in target-present trials (12.7 ms per item vs. 17.7 ms per item), $t(75) = 5.5$, $p < .001$, and target-absent trials (19 ms/item vs. 36.6 ms/item), $t(59) = 9.9$, $p < .001$, replicating the key results of Spivey et al. (2001) and Reali et al. (2006)[1].

[1] The HLM analysis of the untransformed data set also revealed significantly shallower slopes for A/V-concurrent condition than auditory-first condition in target-present trials, $t(75) = 2.02$, $p = 0.04$, and target-absent trials, $t(59) = 4.64$, $p < .001$.

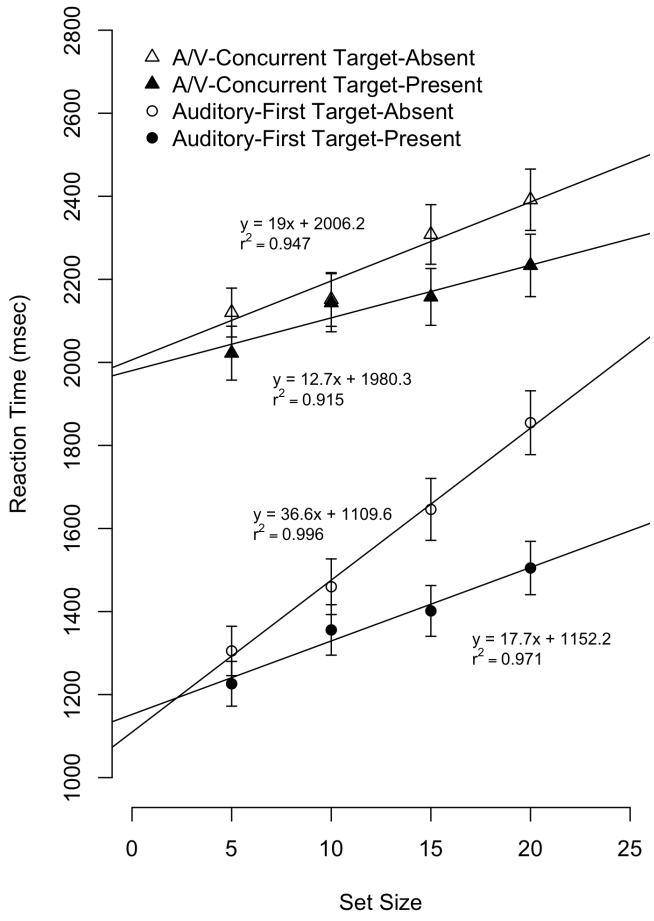

Figure 2.2: Results from Experiment 1. Shown separately for target-present (filled symbols) and target-absent (open symbols) trials for both the auditory-first control (circles) and the A/V-concurrent conditions (triangles). Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

Similar to Spivey et al. (2001) and Reali et al. (2006), we found a near 2:1 ratio between target-absent and target-present trials in the auditory-first condition (36.6 ms/item vs. 17.7 ms/item) but found a slightly lower than 2:1 ratio for the A/V-concurrent condition (19.0 ms/item vs. 12.7 ms/item). This 2:1 ratio between target-absent and -present trials has been regarded as consistent with a standard serial search account (Treisman & Gelade, 1980),

suggesting that, as a result of subtle timing changes made to target identity delivery and display onset, observers may utilize a different strategy when performing an exhaustive search before terminating search and responding "absent" in target-absent trials.

The results of this experiment indicate that by simply adjusting the timing of a spoken query so that the two target feature words were presented at the same time the visual display was visible allowed participants to find the target object in a way that was substantially less affected by the number of distractors, replicating Spivey et al. (2001) and Reali et al. (2006). We observed this finding for the first time with a between-subjects design that effectively rules out strategic accounts that might suggest participants notice the difference between the two types of trials (A/V-concurrent and auditory-first) and then approach the tasks differentially. The results observed in the auditory-first condition are of the type that are traditionally interpreted as consistent with the construction of a conjunction template of the target object followed by a serial process of sequentially comparing each display object with the target template (Treisman & Gelade, 1980). However, by simply shifting the relative timing of visual onset and speech onset, the A/V-concurrent results become more consistent with a parallel or "partial parallel" (Maioli et al., 2008) search process. It appears that the incremental nature of speech processing allows the visual search process to begin when only a single feature of the target identity has been heard. When the initial feature is identified the search proceeds in an efficient nearly parallel fashion so when the second adjective is heard, a substantial amount of the target identification process has been completed – and thus the presence of multiple distractors is less disruptive.

One possible account for the improvement in search efficiency might be that hearing the first adjective triggers a genuinely parallel search mechanism that extracts the objects that exhibit

the target color (typically half of the objects in the display), and then hearing the second adjective triggers a serial search mechanisms that searches for the target among that extracted subset. However, this should produce search slopes (in terms of milliseconds of reaction time per distractor in the display) that are half that of the auditory-first condition, because half of the objects would have been ruled out via an instantaneous parallel mechanism, and many such experiments have instead found slopes far below half (Spivey et al., 2001; Reali, Spivey, Tyler, & Terranova, 2006). Conversely, it may be tempting to conceive of this process as a sequentially nested pair of parallel single-feature searches (first selecting the color, and then selecting the orientation among that selected subset), the results are not quite consistent with that account either. While the slope of the reaction-time-by-set-size function is reliably shallower in the Audio/Visual Concurrent condition, it is not flat (as would be predicted by two nested parallel processes). A biased competition approach (Desimone & Duncan, 1995) to accounting for this linguistic modulation of visual search suggests that, rather than having to choose between parallel (flat slopes) and serial (steep slopes), some continuously graded improvement in efficiency is possible.

Localist Attractor Model

To further investigate the influence of incremental information processing on visual search, Spivey and Dale (2004) and later Reali et al. (2006) implemented a localist attractor network model that easily simulated a potential mechanism by which the visual search process may be influenced by incremental linguistic input. A number of implementations of Desimone and Duncan's (1995) biased competition framework have focused on fitting data from the firing rates of individual neurons in monkey cortex (Reynolds & Desimone, 2001; Spratling &

Johnson, 2004). To complement that approach and to more easily fit reaction time data from humans, this framework was abstracted to a level of functionally-unitized population codes, by creating vectors of nodes that varied in value between 0 and 1, that represent objects competing against one another. Inspired by the biased competition framework Spivey and Dale (2004) developed biased competition framework inspired simulations of visual search reaction times. In the present implementation of this model, one feature vector of nodes represented the target property redness (positive activation) and non-redness (zero activation). Another feature vector represented the target property verticalness (positive activation) and non-verticalness (zero activation). Finally, an integration vector received input from those feature vectors and represented each objects' likelihood of being the target (see fig. 2.3). The lengths of these vectors vary depending on set size between 5 and 25 by intervals of 5; e.g., for a set size of 15 the length of both feature vectors and the integration vector would be 15.

At the beginning of the simulation, initial activation of each node in either feature vector is 1/N, where N is the number of nodes in the vector. Hearing "red" and "vertical" provides input to these feature vectors by multiplying each node by 1 if the object exhibits the appropriate property and by 0 if the object does not exhibit the appropriate property. Similar to a probability distribution, during the network's settling process each timestep begins with the normalization of each feature vector to sum to 1. Those feature vectors are then averaged to produce the activation pattern at the integration layer.

The integration layer sends point-wise multiplicative cumulative feedback to each of the feature vectors, such that each feature node adds to its current activation the product of itself and its corresponding integration node. Since the integration layer's activation pattern is an average of the feature vectors, this feedback functions as a form of crosstalk between the feature vectors,

allowing each feature node to add to its current activation and corresponding integration node such that matching activation peaks strengthen one another over time. For each timestep (treated as 30 ms), this cycle of normalization to integration to feedback repeats until a node in the integration layer exceeds some criterion activation, 0.95 in this case, at which point the target has been found and a settling time (i.e., reaction time) is recorded. Treating each timestep as 30 ms intervals stems from extensive reaction time studies (e.g., Pöppel, 1994) that point to a processing window of approximately 30 ms for perceptual processing. This has been supported by 40 Hz brain wave recordings that have implicated 30 ms as the smallest interval in which features can be bound together within system states, thus events and features occurring within the interval area are perceived as one (Ballard, Hayhoe, Pook, & Rao, 1997).

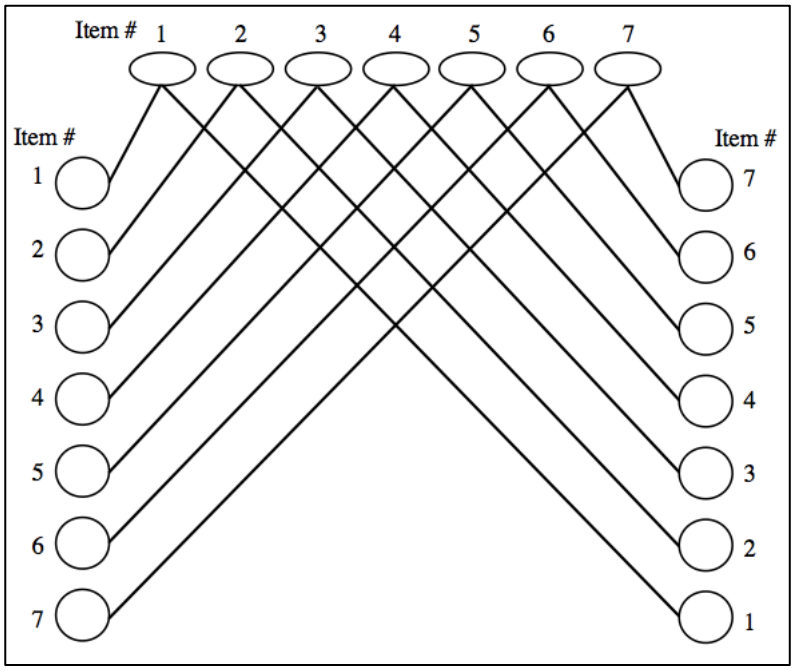

Figure 2.3: Integration-competition model of visual search. A localist attractor network model that simulates a potential mechanism by which the visual search process may be influenced by incremental linguistic input. One feature vector of nodes represented the target property redness (positive activation) and non-redness (zero activation). Another feature vector represented the

target property verticalness (positive activation) and non-verticalness (zero activation). Finally, an integration vector (top) receiving input from those feature vectors represented each object's likelihood of being the target. The lengths of these vectors vary depending on set size, 7 in this example.

This normalized recurrence competition algorithm allows the integration layer to be updated and evaluated in parallel at each timestep, rather than imposing a serial search of one object at a time. This reflects the human data and produces a strikingly linear increase in settling time as set size increases. It should be noted that this competition algorithm does not simulate target-absent trials, since termination of search is not likely the result of a representation winning a competition process (Chun & Wolfe, 1996).

Simulation

When simulating the auditory-first condition the redness and verticalness vectors received input at the same time (Appendix A). The result was an RT-by-set-size slope of 16 ms/item (see fig. 2.4). In simulating the A/V-concurrent condition, the verticalness feature received its input 30 timesteps after the redness feature vector received its input allowing the network to pursue its settling first (the equivalent of 900 ms, corresponding to the point at which the second adjective becomes recognizable). Under these circumstances, the RT-by-set-size slope was reduced to 9.1 ms/item (fig. 2.4). A constant of 900 ms for auditory-first and A/V-concurrent conditions is then added to the RT for perceptual registration and motor execution. The only difference between the parameters of this simulation and that of Reali et al., (2006), is that the constant in our model is slightly longer than that used in Reali et al. (2006) (900 vs. 700

ms). This may be due to the larger proportion of English as a second language-speaking subjects in our participant pool at the University of California, Merced compared to that of Cornell University.

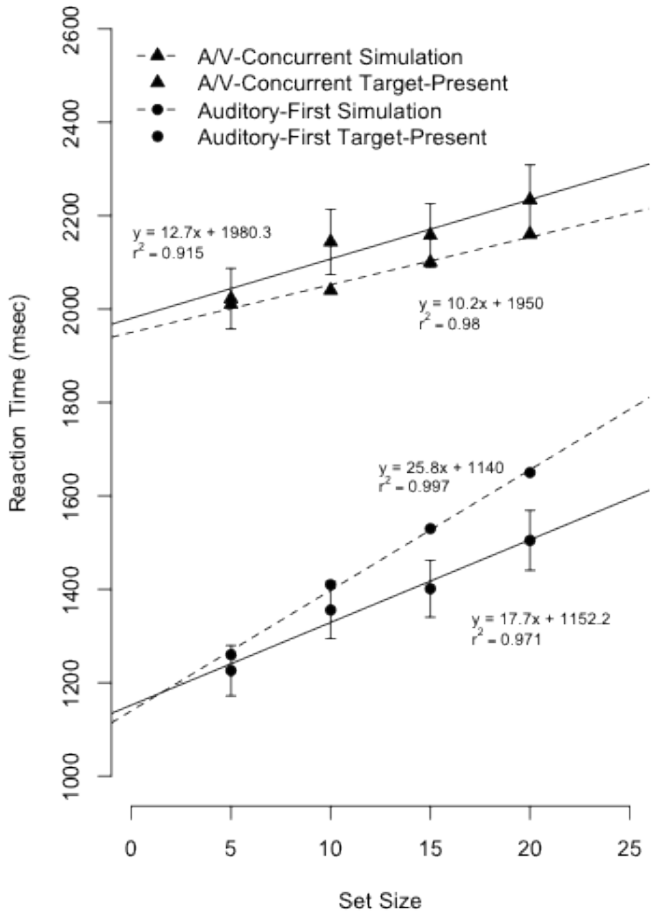

Figure 2.4: Results from the localist attractor network simulation. Dashed lines show the simulation with human data (solid lines) from Experiment 1 for target-present trials. Each line is accompanied by the best-fit linear equation. The results of Experiment 1 are accompanied by the accounted proportion of variance (r^2). The error bars indicate standard error of the mean.

Figure 2.4 shows the target-present trials from Experiment 1 overlapped with the results of the localist attractor simulation. With the only adjustable parameters being the milliseconds

per timestep (30 ms) and the fixed duration for sensory registration and motor output (900 ms), the model is well correlated with the human data collected from Experiment 1 as evident with a RMSE = 87.55 ms and a highly significant Pearson r-squared value, $r^2 = .989$, $p < .001$. Root mean square error (RMSE) can be interpreted as the standard deviation of the unexplained variance or error between the model and the human data, and thus has the valuable property of being in the same metric as the response variable (i.e., milliseconds). In this example a RMSE value of 137.14 ms is a relatively nominal difference, given the 2200+ ms range of this dataset, and reflects a good model fit with only two adjustable parameters.

Predictions

The goal of this model is to do more than merely fit existing data, but to also make predictions for new experiments. To test the model's scalability, and to further investigate the mechanisms of linguistically mediated visual search, we use the same localist attractor network with a minor adjustment to make some predictions for a semi-concurrent condition, where the search display appears immediately after the first target-feature (color) is mentioned but before the second target-feature is presented (orientation). This manipulation of the model predicts the effects of graded difference in feature identification delivery rates, which is novel to the visual search literature.

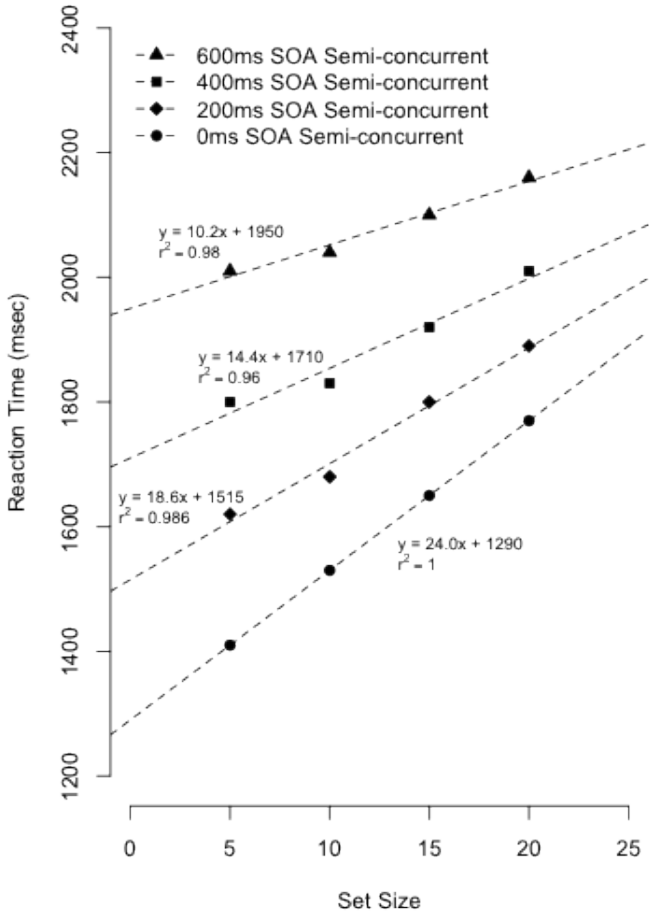

Figure 2.5: Localist attractor network predictions for semi-concurrent conditions. Each line is accompanied by the best-fit linear equation.

Similar to the A/V-concurrent simulations, when simulating the semi-concurrent conditions, the redness feature vector received input slightly before the verticalness feature received input. This allows the network to begin settling the redness vector slightly before the verticalness vector is activated. Initial semi-concurrent simulations activated the redness vector first and then, to allow time to process the first adjective (color), the verticalness vector was activated 10 timesteps (300 ms) later. Under these circumstances, the RT-by-set-size slope differed only slightly from the auditory-first condition with a slope of 24.0 ms/item compared to

25.8 ms/item (fig. 2.5). Additional simulations added the equivalent of 200 ms, 400 ms, and 600 ms to that delay (with 17, 23, and 30 timesteps), resulting in a graded shallowing of RT-by-set-size slope, 18.6 ms/item, 14.4 ms/item, and 10.2 ms/item, respectively. As with the auditory-first and A/V-concurrent simulation a constant of 900 ms was also added to the semi-concurrent predictions to account for perceptual registration and motor execution.

The same localist attractor model (with no adjustments of any parameters) which approximated data from Experiment 1, predicted a systematic shallowing of RT-by-set-size slopes as the timesteps increased from 10 to 30 between the activation of the first feature vector and the second. Essentially, the model produces an RT-by-set-size slope that is comparable to those equated with a serial search process -- even though the model processes its activation patterns in parallel. Interestingly, the conditions with 0 ms and 200-ms SOA behave in a range that is much like that seen in serial or inefficient search processing, and the conditions with 400 ms and a 600-ms SOA behave in a range that is rather close to parallel or efficient search processing. Does this progression of partial improvement in search efficiency (with more and more time between use of the first adjective and use of the second adjective) mirror visual search processing in people? In Experiment 2, these display manipulations first tested on the model are now tested with human participants.

Experiment 2: Semi-concurrent Experiment

This experiment explores the predictions made by the localist attractor network model on what we call the semi-concurrent condition illustrated in Figure 2.5. We utilize four different stimulus onset asynchronies (SOAs) between the visual onset of the display and the auditory

onset of the second adjective, the first adjective was fixed and is always presented immediately prior to the onset of the search display, to imitate the four different timestep durations used in our localist attractor predictions.

Method

The method and design of this experiment followed that of Experiment 1 with the only difference being the onsets of the visual displays and auditory queries. In Experiment 2 observers were now randomly placed in one of four semi-concurrent conditions, where they were presented with one adjective describing the target identity before onset of the visual search display and the other concurrently with, or subsequent to, onset of the visual search display.

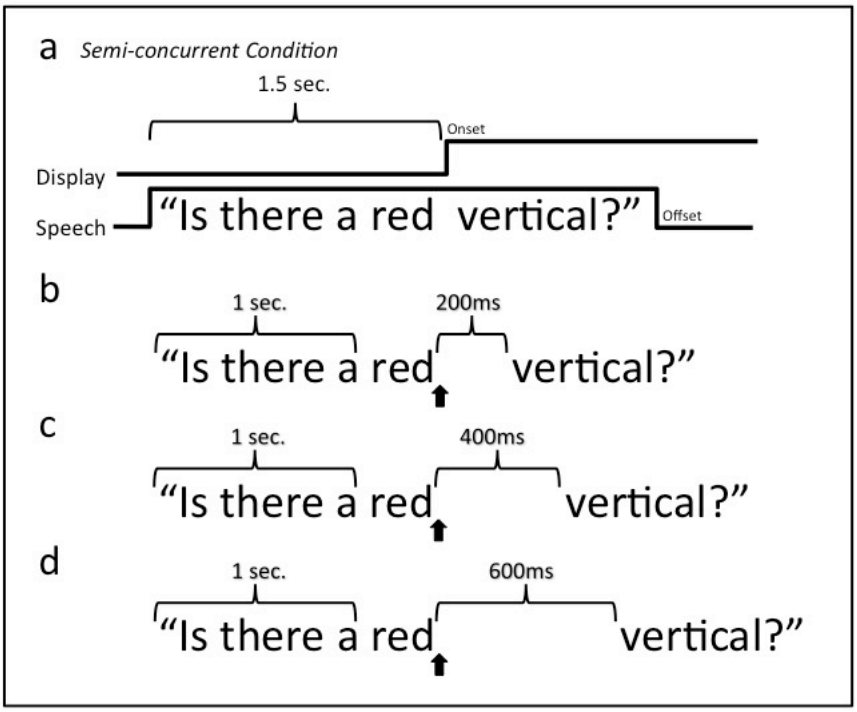

Figure 2.6: Examples of auditory stimuli for semi-concurrent conditions. In the 0-ms SOA semi-concurrent (a) condition, which is similar to the A/V-concurrent condition of Experiment 1, the onset of the visual display coincided with the end of the first descriptive adjective (color).

The arrows indicate display onset for the 200-ms SOA semi-concurrent (b), 400-ms SOA semi-concurrent (c), and the 600-ms SOA semi-concurrent (d) conditions.

The same audio files were utilized from Experiment 1 but with four SOAs introduced between the end of the first descriptive adjective, color, and the second adjective, orientation (e.g. "Is there a green -*SOA*- horizontal"). These SOAs were inserted into the speech files resulting in a total of 16 different speech files, given four SOAs and four types of target objects. In these semi-concurrent conditions, the search display was presented immediately following the first target descriptor, after which subsequent SOAs would elapse before the second target descriptor, was mentioned (see fig 2.6). We used SOAs of 0 ms, 200 ms, 400 ms, and 600-ms.

Participants

A new sample of 275 University of California, Merced undergraduates participated in this experiment for course credit. Participants were randomly assigned to one of four SOA semi-concurrent conditions and only participated in that one SOA condition and no other. Forty-two, 107, 66, and 60 participants were assigned to the 0 ms, 200 ms, 400 ms, and 600-ms SOA condition respectively. Five, 11, 7, and 9 participants did not meet a minimum accuracy of 80% for the 0, 200, 400, and 600-ms conditions respectively and were omitted from the analysis.

Results and Discussion

Figure 2.7 shows the RT-by-set-size functions for target-present (filled symbols) trials in the 0-ms SOA semi-concurrent (circles), 200-ms SOA semi-concurrent (squares), 400-ms SOA semi-concurrent (diamonds), and 600-ms SOA semi-concurrent (triangles) conditions.

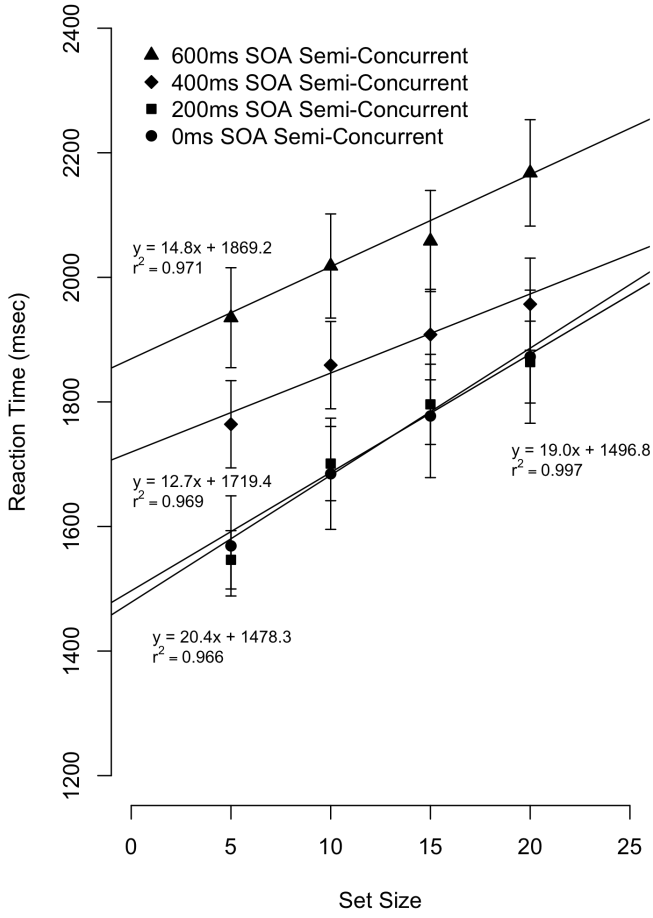

Figure 2.7: Results from Experiment 2. Shown are target-present trials for the 0-ms SOA semi-concurrent conditions (circle), 200-ms SOA semi-concurrent conditions (square), 400-ms SOA semi-concurrent conditions (diamond), 600-ms SOA semi-concurrent conditions (triangle). Each line is accompanied by the best-fit linear equation. The results from Experiment 2 are accompanied by the accounted for proportion of variance (r^2). Error bars indicate standard error of the mean.

Overall mean reaction time and subsequent y-intercepts were slower as SOAs increased because complete notification of target identity was delayed by the duration of the SOA. Mean accuracy was 94.6% for the 0-ms SOA condition, 93.5% for the 200-ms SOA condition, 95.4%

for the 400-ms SOA condition, and 94.0% for the 600-ms SOA condition, which are consistent to previous observations of accuracy on this task (Spivey et al., 2001; Reali et al., 2006).

Since all conditions were between-subjects, an HLM analysis compared the SOA conditions in this experiment to the auditory-first condition from Experiment 1. This analysis revealed significantly shallower slopes, compared to the auditory-first condition, for the 400-ms SOA semi-concurrent condition for target-present trials, $t(58) = 4.48$, $p < .001$, and target-absent trials, $t(58) = 8.81$, $p < .001$, as well as for the 600-ms SOA semi concurrent condition for target-present trials, $t(50) = 3.88$, $p < .001$, and target-absent trials, $t(50) = 8.32$, $p < .001$. Reaction-time-by-set-size slopes where not significantly shallower for the 0-ms SOA for target-present trials, $t(36) = 1.78$, $p = .08$, and target-absent trials, $t(98) = 6.54$, $p < .001$. RT-by-set-size slopes for the 200-ms SOA semi-concurrent condition were not significant for target-present trials, $t(36) = 1.96$, $p = .05$, but were significant for target-absent trials, $t(98) = 3.87$, $p < .001$.

We continue to observe an approximate 2:1 ratio between target-absent and target-present trials in all four of the semi-concurrent conditions (35.7 ms/item vs. 19.0 ms/item for 0-ms SOA, 28.6 ms/item vs. 20.4 ms/item for 200-ms SOA, 22.6 ms/item vs. 12.7 ms/item for 400ms SOA, 25.7 ms/item vs. 14.8 ms/item for 600-ms SOA) with a possible nonlinear shift from 3:2 ratio to a 2:1 as SOA's increase from 0 ms to 600 ms. Future research should investigate the progression of target-absent and target-present ratio trends. This observed nonlinear progression in ratio may provide insight into target-absent search strategies.

The findings are generally consistent with the localist attractor network predictions (fig. 2.7), where we saw a progression of partial improvement in search efficiency as the SOA increases. As predicted by the model, the slopes of the RT-by-set-size functions for target-present trials revealed a gradual shallowing from the 0-ms SOA condition to the 600-ms SOA

condition. With SOAs of 400 ms and 600 ms, the slopes of the RT-by-set-size functions for target-present trials were very near the range of parallel or efficient search – just as observed with the model simulation. One apparent deviation between the model and the human data is in comparing the 0 ms and 200-ms SOA conditions. In the model, there is a moderate difference in slope and a moderate difference in mean reaction times as well. Curiously, in the human data, there is almost no difference between these two conditions in slope or even in mean reaction times. For these two conditions, the human data produce slopes and mean reaction times that are nearly perfectly in between the respective slopes and mean reaction times that are produced by the model for these two conditions.

It appears that the model's predictions for Experiment 2 were generally well fitted. Further investigation reveals that the model is a good fit to the data collected from Experiment 2 with a RMSE = 77.21 ms and a highly significant Pearson r-squared value, $r^2 = .994$, $p < .001$. In this comparison, a RMSE value of 77.21 ms is a relatively nominal difference and reflects a good model fit given the nearly 2200 ms range of the data.

CHAPTER THREE

Non-linguistic Preview Experiments

Introduction

Experiments by Jones, Kaschak, and Boot (2011) used eye-tracking to examine an alternative view to one that proposes search efficiency is increased due to language enhancing perceptual processing. Jones and colleagues (2011) observed patterns of eye movements suggesting increased efficiency with concurrent speech was not likely due to linguistic enhancement of perceptual processes but instead delaying the onset of target-seeking eye movements. They contend the findings by Gibson et al. (2005) are better explained by this "preview" of search display (when observers are presented with the search display prior to being notified of the target object's identity) because slower speech provides observers with more search display viewing time, which provides additional information about potential target locations independently of the information conveyed by auditory linguistic speech stream.

Experiment 3A

The purpose of the present study is to examine the role of preview of search display on visual processing. This study is part of an ongoing effort to understand exactly how language comprehension and visual search interact in real-time. In this experiment, we utilized visual cues to deliver simultaneously a two-feature target identity in a conjunction-search task to test the role of preview on visual search.

Method

In this experiment we employed three SOA conditions (0 ms, 350 ms, and 750 ms) when identifying the target object. Participants were either presented with the target identifying visual cue simultaneously with the search display (0-ms SOA) or with either a 350 ms or 750 ms delay after onset of search display (see fig. 3.1). All three SOAs appeared equally and randomly.

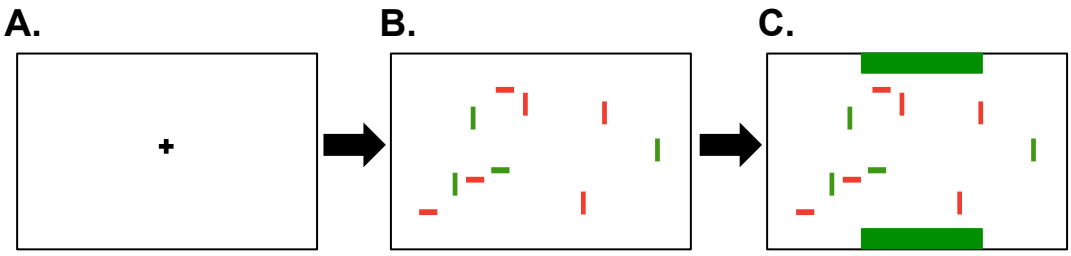

Figure 3.1: Example of nonlinguistic visual cues trial presentation for Experiment 3. Duration of search display (B) varied between 0, 350, & 750 ms in Experiment 3A and 0 & 1500 ms in Experiment 3B before the target identifying visual cues appeared (C).

Participants

One hundred and fifty-seven University of California, Merced undergraduate students received course credit for participating in this experiment. Twenty-four participants were unable to perform the task with an accuracy of 80% or better and were removed from the analysis.

Stimuli and Procedure

Target identifying visual cues were either red or green horizontal bars that appeared at the top and bottom of the search display or were red or green vertical bars that appeared on the left

and right of the search display. Dimensions of the visual cues were designed to resemble the dimensions of the stimulus objects but four times larger. Stimulus bars were identical to the ones in first two experiments. The first block was referred to as the "practice" block, consisting of 32 trials, and was followed by an experimental block with 96 trials. The design of the experiment was consistent with previously mentioned experiments (Experiments 1 & 2). The duration of the entire experiment took approximately fifteen minutes to complete.

Results and Discussion

In this experiment we demonstrate with various conditions that search efficiency does not increase in a conjunction-search task when target features are delivered simultaneously, despite having time to preview the search display. The RT-by-set-size functions for target-present trials (filled symbols) are shown in Figure 3.2 and 3.3 for target-absent trials (open symbols) in the three SOA conditions, 0 ms (circles), 350 ms (diamonds), and 750 ms (triangles). We should note at this time that RT's were recorded from display onset, irrespective of condition, until a response was made. In the 0-ms SOA control condition, the RT-by-set-size function was highly linear in both target-present, $r^2 = .994$, and target-absent trials, $r^2 = .984$, as typically observed in standard conjunction-search tasks. Similarly, the RT-by-set-size functions for the 350 ms and 750 ms SOA conditions were highly linear in target-present trials, $r^2 = .925$ and $r^2 = .992$, and target-absent trials, $r^2 = .977$ and $r^2 = .961$, respectively.

Since our primary interest is to assess the effects of preview on visual search efficiency, analysis in this experiment compared the 350 ms and 750-ms SOA conditions to the 0-ms SOA condition. Overall mean RTs, as well as y-intercepts, were significantly slower in the 350 ms and 750-ms SOA conditions because delivery of target identity was delayed by 350 ms and 750

ms, respectively, relative to the 0-ms SOA condition for both target-present, $t(132) = 2.38$, $p = .017$, and $t(132) = 8.21$, $p < .001$, and for target-absent, $t(132) = 4.05$, $p < .001$, $t(132) = 9.31$, $p < .001$, trials. Similar to previous observations mean accuracy was 94.7% for all three conditions (Spivey et al., 2001; Reali et al., 2006).

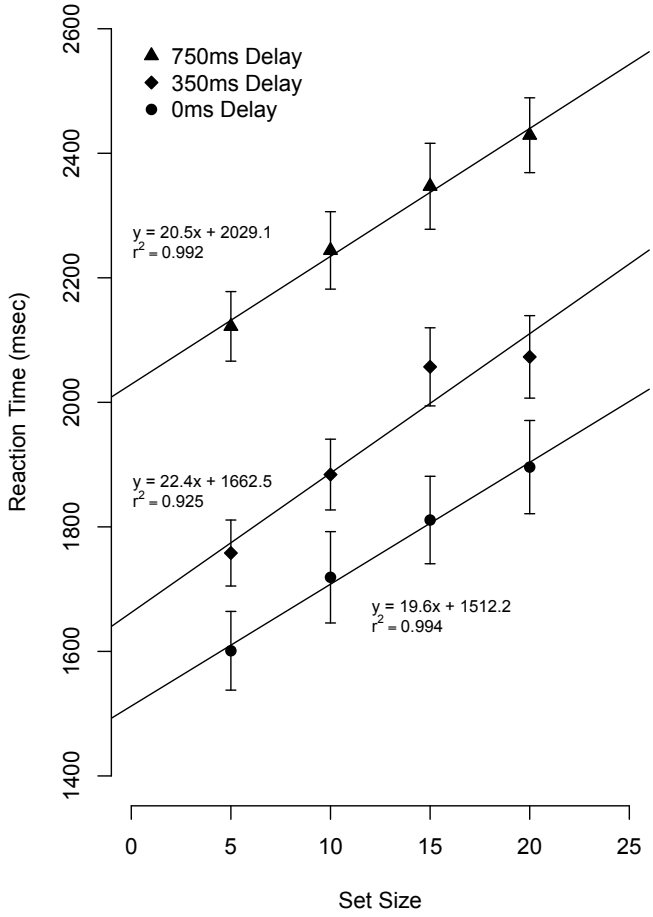

Figure 3.2: Results from Experiment 3A for target-present trials (filled symbols). Shown for the 0 ms delay condition (circle), 350 ms delay condition (diamond), and 750 ms condition (triangle). Each line is accompanied by the best-fit leaner equation and the accounted proportion of variance (r^2). Error bars indicate standard error of the mean.

The slopes of the RT-by-set-size functions reveal that 350 ms and 750-ms SOA conditions did not produce more efficient visual search compared with the 0-ms SOA conditions (see fig. 3.2 & 3.3). Contrary to findings by Jones et al. (2011) an analysis revealed slopes for the 350 ms and 750-ms SOA conditions compared to the 0-ms SOA condition were not significantly different for target-present trials (22.4 ms/item & 20.5 ms/item vs. 19.6 ms/item), $t(132) = 0.61$, $p = .543$, and $t(132) = 0.21$, $p = .835$, and target-absent trials (37.0-ms/item & 35.7 ms/item vs. 41.9 ms/item), $t(132) = -0.99$, $p = .323$, and $t(132) = -1.26$, $p = .207$.

Although observers' eye-movements in Jones et al. (2011: Experiment 2) were constrained, unlike for the aforementioned experiments, for either a "short" 350 ms or a "long" 750 ms while they viewed the search display, observers were presented with the target's identity prior to the onset of the search display. This is the primary difference between the study by Jones et al. (2011: Experiment 2) and the one we conducted and likely explains the contrary results. Thus it appears that when observers are presented with the target identity and allowed to view the search display, albeit with eye-movements constrained to a central fixation cross, before responding, observers' search efficiency improves (Jones et al., 2011) but only when given sufficient time for processing and not immediately following target identification as with the auditory-first conditions used in previous experiments (Spivey et al., 2001; Reali et al., 2006; Chiu & Spivey, 2012).

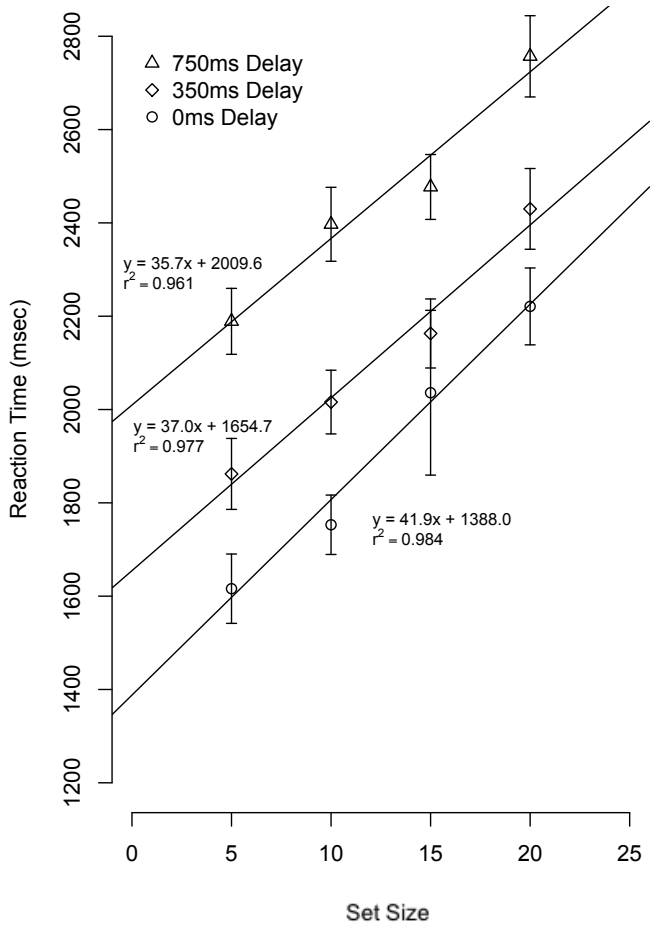

Figure 3.3: Results from Experiment 3A for target-absent trials (open symbols). Shown for the 0 ms delay condition (circle), 350 m delay condition (diamond), and 750 ms condition (triangle). Each line is accompanied by the best-fit leaner equation and the accounted proportion of variance (r^2). Error bars indicate standard error of the mean.

Similar to Spivey et al. (2001) and Reali et al. (2006), we found a near 2:1 ratio between target-absent and -present trials in all three conditions (37.0-ms/item vs. 22.4 ms/item for 0-ms SOA, 35.7 ms/item vs. 20.5 ms/item for 350-ms SOA, and 41.9 ms/item vs. 19.6 ms/item for 750-ms SOA).

The results of this experiment indicate that simply delivering target identity simultaneously in a conjunction-search task with a variety of SOAs so that observers are allowed preview time does not substantially affect search efficiency. However when target identity is known prior to display onset and given sufficient preview before requiring a response, search efficiency does improve (Jones et al., 2011). Is it possible that with the search display available but target identity unknown, as in this experiment, a maximum SOA of 750 ms does not provide sufficient preview to initiate any meaningful visual processing?

Experiment 3B

In this experiment we extended the methods in Experiment 3A to first, mimic the entire duration (1500 ms) of the auditory linguistic query, which identified the target object, in previous work by Spivey et al. (2001) and to, secondly, explore the effects of a relatively long preview duration of search display on visual search processing.

Method

The methods of this experiment follow that of Experiment 3A with the exception that only two SOAs (0 ms and 1500 ms) were used for the target identifying visual cue.

Participants

Fifty-nine University of California, Merced undergraduate students received course credit for participating in this experiment. Five participants were unable to perform the task with an accuracy of 80% or better and were subsequently removed from the analysis.

Stimuli and Procedure

The same stimuli and target identifying visual cues from Experiment 3A were used in this experiment. Participants were presented with both SOAs equally and randomly in a within-subjects experimental design. The same testing apparatuses and software were used in this experiment as the previous one.

Results and Discussion

As with Experiment 3A, we continue to demonstrate with a slightly different condition that search efficiency does not increase with simultaneous delivery of target feature in a conjunction-search task, despite having time to preview the search display. Figure 3.4 shows the RT-by-set-size functions for target-present trials (filled symbols) and target-absent trials (open symbols) in the 0-ms SOA (triangles) and 1500-ms SOA (circles). In the 0-ms SOA condition, the RT-by-set-size function was highly linear in both target-present, r^2 = .995, and target-absent trials, r^2 = .979, as typically observed in standard conjunction-search tasks. Similarly, the RT-by-set-size functions for the 1500-ms SOA condition was highly linear in target-present trials, r^2 = .975, and target-absent trials, r^2 = .958.

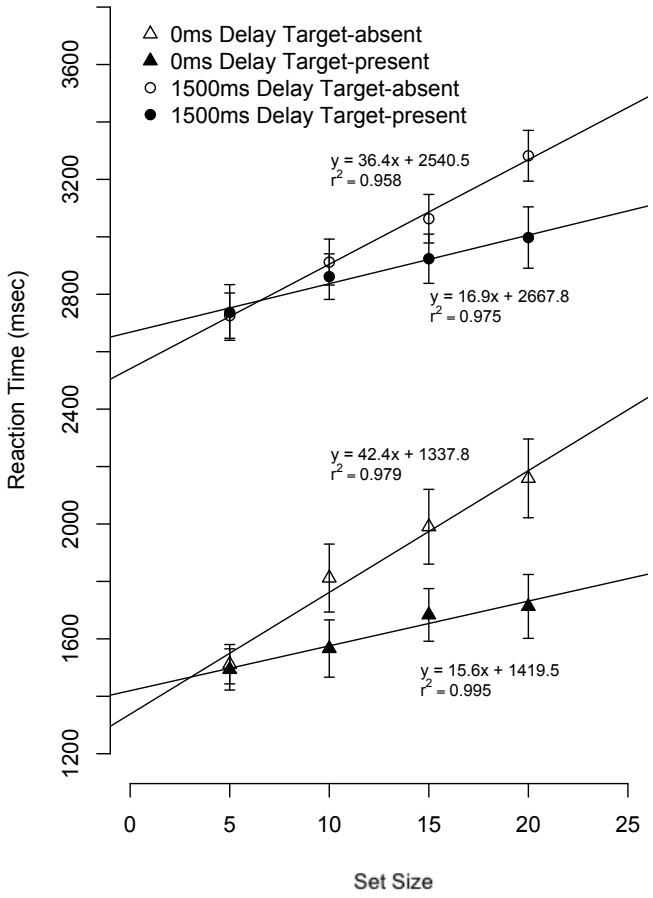

Figure 3.4: Results from Experiment 3B. Shown for target-present trials (filled symbols) and target-absent (open symbols) for the 0 ms delay conditions (triangle) and 1500 ms delay conditions (circle). Each line is accompanied by the best-fit linear equation, the accounted proportion of variance (r^2). Error bars indicate standard error of the mean.

Overall mean RTs, as well as y-intercepts, were significantly slower in the 1500-ms SOA condition because delivery of target identity was delayed by 1500 ms relative to the 0-ms SOA condition for both target-present, $t(53) = -3.05$, $p = .002$, and target-absent, $t(53) = -3.06$, $p < .002$, trials. Similar to previous observations mean accuracy was 94.5% for both conditions.

The slopes of the RT-by-set-size functions reveal that the 1500-ms SOA condition did not produce more efficient visual search compared with the 0-ms SOA conditions (fig. 11). An

analysis revealed slopes for the 1500-ms SOA condition compared to the 0-ms SOA condition were not significantly different for target-present trials (16.9 ms/item vs. 15.6 ms/item), $t(53) = 0.22$, $p = .825$, and target-absent trials (36.4 ms/item vs. 42.4 ms/item), $t(53) = -0.85$, $p = .398$.

Consistent with the previous experiments (Experiments 1-3A), we found a near 2:1 ratio between target-absent and -present trials in both conditions (36.4 ms/item vs. 16.9 ms/item and 42.4 ms/item vs. 15.6 ms/item). The results of this experiment continue to indicate simply delivering target identity simultaneously in a conjunction-search task, even with a relatively long SOA (1500 ms), does not significantly affect search efficiency. It appears that concurrent and incremental target-identity delivery may be vital to elicit a modified search strategy as with even a 1500-ms SOA, that mimics the overall duration of the linguistic query used in Experiment 1 & 2, search strategies do not significant change.

CHAPTER FOUR

Non-linguistic Incremental Experiments

Introduction

Although Experiment 3B (in Chapter 3) utilized an SOA with an overall duration equivalent to the entire length of the linguistic query used in Experiment 1 (Chapter 2), it was unsuccessful in eliciting the same significant improvement in search strategies previously observed (see Chapter 2; also Spivey et al., 2001; Reali et al., 2006; Chiu & Spivey, 2012). This raises the question, if not preview than what is it about a concurrent linguistic and visual delivery that improves search efficiency? Auditory spoken language unfolds information overtime and is, subsequently, processed incrementally as evident by studies of real-time linguistic processing. These studies assert that language comprehension immediately takes into account information as it is presented, directing attention to relevant objects as a sentence progresses (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). Our findings thus far indicate that the incrementality of language comprehension is crucial in the interaction that produces a more efficient use of visual attentional resources. This experiment is designed to test the role of incremental information processing, characteristic of language comprehension, on visual attention. This is done by visually imitating the temporal characteristics of the auditory linguistic query used in previous work to identify a target object during a conjunction-search task (e.g., Spivey et al., 2001; Gibson et al., 2005). I expect that with the addition of incremental information delivery, albeit visual therefore

unimodal, search efficiency will improve just as it does with a concurrent auditory linguistic query.

Experiment 4A

This experiment explores the effect of incremental non-linguistic information delivery on visual search processing by visually replicating the temporal characteristics of the auditory linguistic query that was used to identify the target object in previous work by Spivey et al. (2001) (also see Gibson et al., 2005; Reali et al., 2006; Chiu & Spivey. 2012).

Methods

In this experiment two slightly different conditions were used to simulate the auditory-first and A/V-concurrent condition first used by Spivey et al. (2001). A *cue-first* condition similar to the auditory-first condition, delivers target identity incrementally via a visual cue prior to display onset, and a *cue-concurrent* condition similar to the A/V-concurrent condition, delivers target identity incrementally via an identical visual cue but concurrently with display onset. For Reali et al. (2006: Experiment 1) several participants spontaneously reported being unaware of any difference in display onset timing; no participant reported experiencing a difference between trials when auditory-first and A/V-concurrent trials appeared in a random order within one block of 192 trials. However, due to the unimodal nature of this task the difference between timing of display onset for cue-first and cue-concurrent trials is much more apparent and may elicit conscious adjustments to search strategies, which would not allow us to test the natural interaction between information processing and visual attention. In order to

reduce this possibility, I opted for a blocked trial design. Participants participated in both types of trials, cue-first control and cue-concurrent, in random order.

Participants

Forty-six University of California, Merced undergraduate students received partial course credit for participating in this experiment. Eight participants were unable to perform the task with a minimum accuracy of 80% and were subsequently removed from the analysis. As with Experiments 1-3, all incorrect responses and trials with reaction times greater than 2.5 IQRs from the median were also omitted from the analysis.

Stimuli and Procedure

Stimulus objects were identical to the ones used in the previously mentioned experiments (Experiments 1-3). In order to visually simulate the incremental information delivery of the spoken query (e.g., "Is there a red vertical?" 500 ms to utter the first feature color, "red" or "green," and 1000 ms to utter the second feature orientation, "vertical" or "horizontal") used in Spivey et al. (2001), target identifying visual cues that identified the color of the target began as either all red or all green horizontal and vertical bars that appeared on all sides (top, bottom, left, and right) of the search display for 500 ms. To identify the orientation of the target, the visual cues then transitioned from the colored bars to grey horizontal or vertical bars that appeared only at the top and bottom or the left and right of the search display, respectively, for 1000 ms before disappearing (see fig. 4.1). Dimensions of the visual cues were identical to Experiments 3A & 3B.

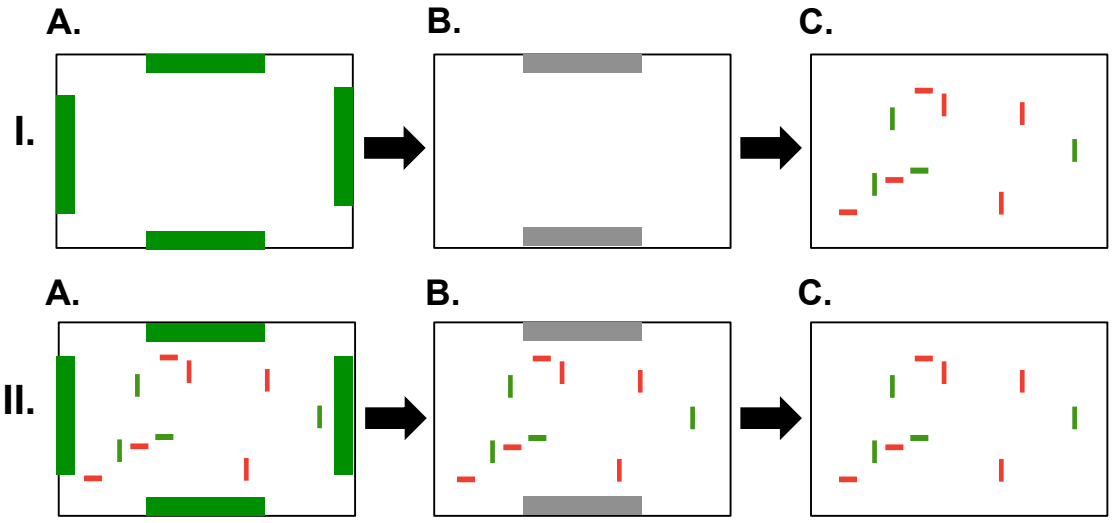

Figure 4.1: Example of nonlinguistic visual cue trial presentation for Experiment 4. Shown separately for cue-first (I) and cue-concurrent (II) conditions. Trial presentation for Experiment 4B are identical to Experiment 4A (500 ms for color & 1000 ms for orientation) with the exception that the duration of the color cue (A) lasted for 300 ms and the duration of the orientation cue (B) lasted for 600 ms. Experiment 4C uses the same cue timing as Experiment 4A but present orientation first (B to A to C)

Prior to participating in the two experiment blocks observers participated in two practice blocks one of each condition (cue-first and cue-concurrent). Each practice block consisted of 16 trials each for a total of 32 practice trials. Two experiment blocks of 64 trials, for a total of 128 trials, followed the practice blocks. One experimental block contained cue-first trials only and the other cue-concurrent trials only. The order of the experiment blocks (cue-first first or cue-concurrent first) were randomly assigned to participants, each order was used equally. Participants were instructed to respond to each display as quickly and accurately as possible by

pressing the labeled “YES” button on the keyboard if the target is present in the display and the labeled “NO” button if it is absent. The duration of the entire experiment lasted approximately 20 minutes.

Results and Discussion

The findings demonstrate an improvement in search strategies when visual non-linguistic cues deliver target features incrementally and concurrent with the visual display onset. There was no effect observed when target features were delivered prior to display onset. Figure 4.2 shows the RT-by-set-size functions separately for target-present trials (filled symbols) and target-absent trials (open symbols) in the cue-first (triangle) and cue-concurrent (circle) conditions. The RT-by-set-size function remained highly linear in the cue-first condition for target-present, $r^2 = 0.768$, and target-absent, $r^2 = 0.962$, trials as well as in the cue-concurrent condition for target-present, $r^2 = 0.314$, and target-absent, $r^2 = 0.698$. This linearity is consistent with the visual search literature including the linguistically mediated visual search literature (Spivey et al., 2001 & Reali et al., 2005). Overall mean reaction time, as well as y-intercepts, were significantly slower in the cue-concurrent condition because complete delivery of target identity was delayed by 1500 ms relative to the cue-first control condition for both target-present, $t(37) = 4.49$, $p < .001$, and target-absent, $t(37) = -4.32$, $p < .001$, trials. Mean accuracy was 94.0% for both conditions.

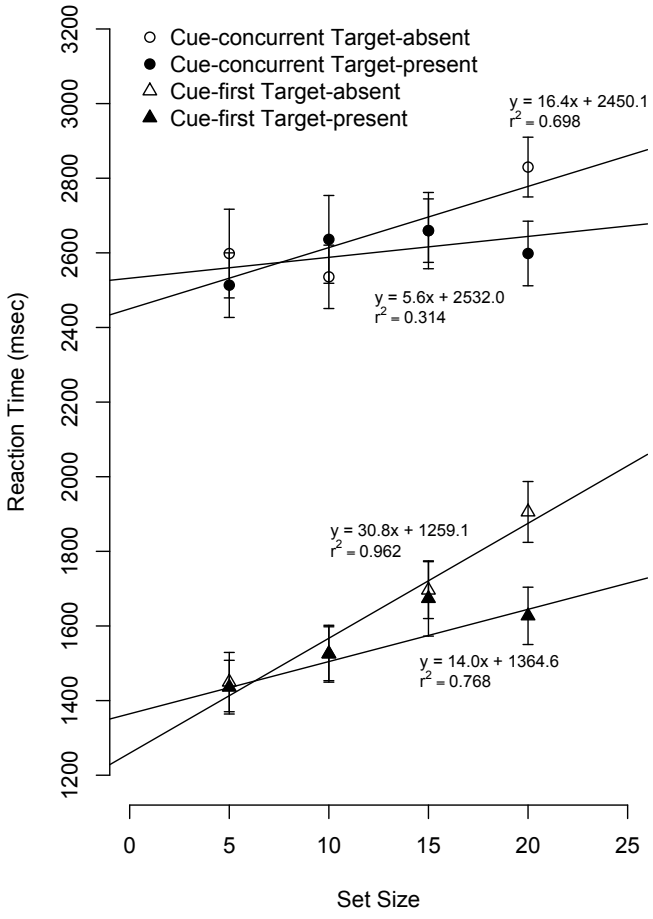

Figure 4.2: Results for Experiment 4A. Shown separately for target-present (filled symbols) and -absent trials (open symbols) for both cue-first (triangles) and cue-concurrent (circles) conditions. Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

The slopes of the RT-by-set-size functions reveal that the cue-concurrent conditions produced more efficient visual search compared with the cue-first conditions (see fig. 4.2). An analysis revealed slopes for the cue-concurrent condition compared to the cue-first condition were significantly different for target-present trials (5.6 ms/item vs. 14.0 ms/item), $t(37) = -2.77$, $p = .010$, and target-absent trials (16.4 ms/item vs. 30.8 ms/item), $t(37) = -2.75$, $p = .006$. The

results here, as with previous studies (Spivey et al., 2001; Reali et al., 2006), exhibit a near 2:1 ratio between target-absent and -present trials for both cue-concurrent conditions (16.4 ms/item vs. 5.6 ms/item) and cue-first conditions (30.8 ms/item vs. 14.0 ms/item).

The results of this experiment indicate that visual non-linguistic delivery of target features presented incrementally and concurrently with the visual display onset has a facilitatory effect on visual search efficiency, but not when the target features are delivered prior to display onset. It could in principle be seen that the findings observed in the cue-first condition appear to be consistent with a traditionally serial approach where observers wholly and discretely compare each display object with the target template (Treisman & Gelade, 1980) and the findings in the cue-concurrent condition, which simply involves shifting the timing of display onset relative to target identity cues, are more consistent with a parallel or "partial parallel" search process where search begins before complete biased competition converges to a solution (Maioli, Benaglio, Siri, Sosta, & Cappa, 2001). However, rather than a hybrid model that posits two separate cognitive mechanisms for search processing, I proposed and implemented a simulation (Chapter 2: Localist Attractor Network) of a single mechanism that is purely parallel in processing arrays of input.

This localist attractor network is able to produce parallel-like and serial-like behavior as well as graduations between the two by using a process that is purely parallel in nature. A hybrid (half parallel and half serial) partial parallel search process would in theory elicit a RT-by-set-size slope half that of a purely series search process (detailed in Chapter 2) but this is not the case for the findings in this literature (Experiment 4A; also see Spivey et al, 2001; Reali et al., 2006; Chiu & Spivey, 2012). Instead we see a range of ratios that vary from a low of approximately 1:5 (Reali et al., 2006: Experiment 2) to high of approximately 2:3 (Chapter 2:

Experiment 1). Findings like this, in addition to an absence of a bimodal distribution when examining 2,500 visual search experiments (Wolfe, 1998), provides evidence against a dichotomous view of visual search and supports a continuum of search strategy propelled by a single mechanism.

Experiment 4B

The phenomenon of linguistically mediated visual search is based on subtle, a few hundred millisecond, changes to display onset in relation to target identifying linguistic queries. As mentioned before Gibson et al. (2005) found with a faster speech rate (4.8 vs. 3.0 syllables/second) the A/V-concurrent condition, first used in Spivey et al. (2001), no longer provides an enhanced efficiency effect on conjunction-search tasks when compared to auditory-first conditions. This indicates that linguistic mediation of visual search is affected by speech rate, which is not surprising given the subtly timing nature of this effect. In the last experiment (Experiment 4A) we saw that the facilitory effect of an A/V-concurrent onset could be replicated using nonlinguistic and incremental visual cues. To test the difference between the effect of incremental visual cue and linguistic query on visual attention this experiment explores the role of a faster rate of incremental information delivery on visual search by visually replicating the slightly faster temporal characteristics of the auditory linguistic query that was used to identify the target object in previous work by Gibson et al. (2005) using the same incremental non-linguistic information delivery used in Experiment 4A. Past findings, where increasing speech rate eliminates improvements in search efficiency previously observed with A/V-concurrent trials by Spivey et al. (2001) (Gibson et al., 2005), would predict that with a faster rate of information delivery improvements in search strategy will no long be present.

Method

Methods in this experiment largely follow the previous experiment (Experiment 4A). In order to simulate the faster auditory-first and A/V-concurrent conditions used by Gibson et al. (2005) in this experiment, the timing of the target identifying visual cues from Experiment 4A were slightly modified and detailed below.

Participants

Thirty-eight University of California, Merced undergraduate students received partial course credit for participating in this experiment. Fourteen participants were unable to perform the task with a minimum accuracy of 80% and were subsequently removed from the analysis. We continued to omit all incorrect responses and trials with reaction times greater than 2.5 IQRs from the final analysis.

Stimuli and Procedures

Stimulus bars were identical to Experiments 1-4A. In order to visually simulate the incremental information delivery of the faster spoken query in Gibson et al. (2005) compared to the one used by Spivey et al. (2001), the same target identifying visual cues from Experiment 4A that began as all red or all green horizontal and vertical bars that appeared on all sides (top, bottom, left, and right) of the search display was presented for a shorter 300 ms (vs. 500 ms in Experiment 4A) to identify the color of the target. In order to identify the orientation of the target, the same visual cue from earlier then transitioned to grey horizontal or vertical bars that appeared either at the top and bottom or the left and right of the search display for a shorter 600

ms (vs. 1000 ms in Experiment 4A) before disappearing (see fig. 4.1). Dimensions of the visual cues were identical to the previous experiment along with all other design, procedures, and instructions.

Results and Discussion

Figure 4.3 shows the RT-by-set-size functions for target-present trials (filled symbols) and target-absent trials (open symbols) in the cue-first (triangle) and cue-concurrent (circle) conditions. As predicted by findings by Gibson et al. (2005) the results were unable to demonstrate a facilitatory effect when visual non-linguistic delivery of target features were presented either prior or concurrently with the visual display onset when target identifying visual cue rate is increased. The RT-by-set-size functions continue to produce highly linear regressions for both target-present and -absent trials in both the cue-concurrent condition, $r^2 = .910$, $r^2 = .977$, respectively, and the cue-first condition, $r^2 = .670$, $r^2 = .949$, respectively. Overall mean reaction time was significantly slower in the cue-concurrent condition because complete delivery of target identity was delayed by 900 ms relative to the cue-first condition for both target-present, $t(24) = -2.60$, $p < .01$, and target-absent, $t(24) = -1.72$, $p < .01$, trials. Mean accuracy was 93.3% for both conditions.

The slope coefficients of the RT-by-set-size functions reveal that the cue-concurrent conditions did not produced more efficient visual search compared with the cue-first conditions (see fig. 4.3). An analysis revealed slopes for the cue-concurrent conditions were not significantly different than the cue-first conditions for target-present trials (25.1 ms/item vs. 18.4 ms/item), $t(24) = 0.67$, $p = .504$, and target-absent trials (48.2 ms/item vs. 45.2 ms/item), $t(24) = 0.26$, $p = .796$. We continue to observe a near 2:1 ratio between target-present and -absent trials

for both cue-concurrent conditions (25.1 ms/item vs. 48.2 ms/item) and cue-first conditions (18.4 ms/item vs. 45.2 ms/item).

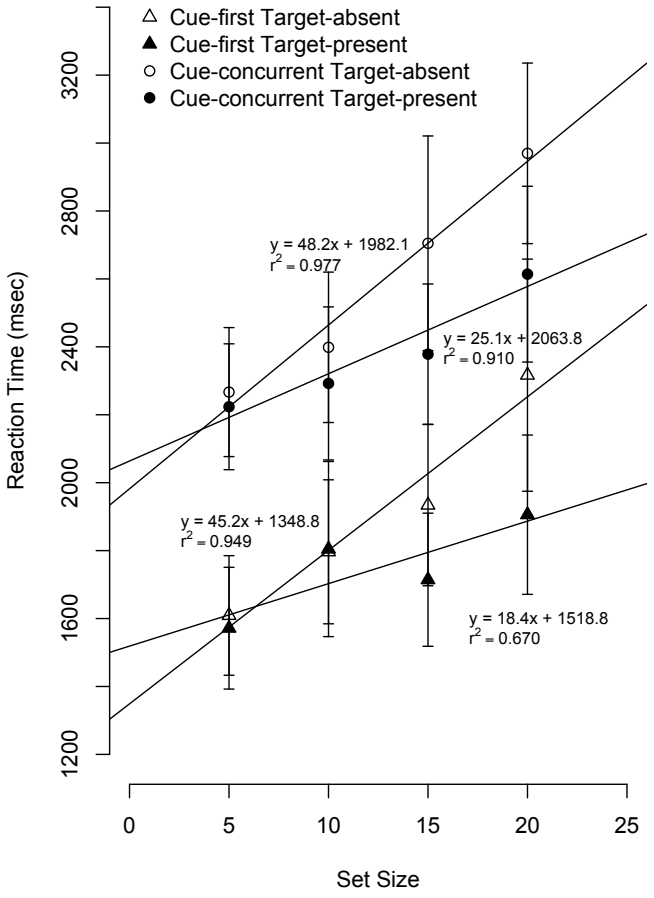

Figure 4.3: Results for Experiment 4B. Shown separately for target-present (filled symbols) and -absent trials (open symbols) for both cue-first (triangles) and cue-concurrent (circles) conditions. Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

The results of this experiment show that when the rate of visual non-linguistic delivery of target features is increased (900 vs. 1500 ms overall) the facilitatory effect previously observed with visual search when target identity is presented incrementally and concurrently with the

visual display onset is no longer present. This finding suggest that although visual search can be made more efficient with concurrent incremental information delivery, this interaction is sensitive to temporal constraints such that simply increasing the rate of information delivery elicits a pattern of search that is more consistent with a conventional serial processing account (Treisman & Gelade, 1980).

This contradicts the "preview" interpretation by Jones et al. (2011), which postulates the difference in results among Spivey et al. (2001) and Gibson et al. (2005) is solely due to the longer preview period afforded by a slower speech rate rather than the consequence of linguistic enhancement as originally proposed by Spivey et al. (2001). Instead it appears, as we see now in this study (Experiment 4B), that the immediate interaction between visual attention and incremental information processing, such as with understanding spoken language, is one that requires adequate time to activate and integrate internal attentional saliency maps before being able to map onto motor output. Thus, not only is it necessary for the target identity to be both delivered incrementally and presented concurrently with display onset in order to produce an interaction that improves search efficiency it is also necessary for the information delivery to be at a rate that permits the real-time interaction between information processing (visual/unimodal, or linguistic/multimodal) and visual attention.

Experiment 4C

When it comes to language mediated visual search the order of adjective delivery (color adjective first vs. orientation adjective first) has been found to effect search strategies (Spivey et al, 2001; Reali et al., 2006). In a four-query design, where all four queries were orientation-first (e.g., "Is there a vertical red?"), Spivey et al. (2001: Experiment 2) continued to observe a

facilitory effect for target-present trials but improvement in target-absent trials was only marginal. In a more complex eight-query design, where half of the queries were color-first and the other half orientation-first, Reali et al. (2006: Experiment 2) observed a facilitory effect in the color-first condition for both target-present and –absent trials as well as for target-absent trials in the orientation-first condition however there was no effect observed for target-present trials. Although in both instances when the effects were not found to be significant RT-by-set-size slopes for the A/V-concurrent trials were still numerically smaller or shallower than for auditory-first control trials. This suggests that the order of adjectives may be an important factor in eliciting the observed interaction between language processing and visual search. This is consistent with the preferred order of feature type delivery observed in other studies (Olds & Fockler, 2004) that show a color preview and a color preview followed by an orientation preview facilitated search but not vice versa or with orientation alone, the last of which was even hurtful when presented alone before the search display.

We also see growing evidence that words acquired earlier in development are processed more accurately and quickly than words acquired later (Boulenger, Décoppet, Roy, Paulignan, & Nazir, 2007). It is safe to say, for English, that color adjectives (e.g. red, green, etc.) are acquired well before orientation adjectives (e.g. vertical, horizontal, etc.), even now participants occasionally need to be reminded of the definition of orientation adjectives. The purpose of this experiment is to, first, investigate whether the weaker effect of orientation-first target identity delivery is primarily due to linguistic factors or simply the result of a stronger effect with color descriptions over orientation and to, secondly, continue expanding the generalizability of the paradigm. In this study we continue with our unimodal visual paradigm but reverse the order of the feature cues with the incremental target identifying visual cues.

Method

The design of this experiment is identical to that of Experiment 4A with the exception that target identifying visual cues appear in reverse order (orientation first then color). The duration of feature presentation remains the same, 1000 ms for orientation and 500 ms for color.

Participants

Thirty-three University of California, Merced undergraduate students received partial course credit for participating in this experiment. Eight participants were unable to perform the task with a minimum accuracy of 80% and were removed from the analysis. As with the previous experiments we omitted all incorrect responses and trials with reaction times greater than 2.5 IQRs from the median.

Stimuli and Procedures

This experiment used the same search displays and timing of stimuli as in Experiment 4A. The practice block lasted 16 trials and the experiment block consisted of 128 trials. Practice block trails were not included in the final analysis. Half of the trials were target-present and half were target-absent; set sizes of 5, 10, 15, and 20 were used. Four unique targets, given two features (orientation and color), appeared equally and randomly.

Results and Discussion

As expected, the cue-concurrent condition elicited reaction times that were overall significantly slower than those in the cue-first control condition for both target-present, $t(24)$ =

15.75, $p < .001$, and target-absent, $t(24) = 12.76$, $p < .001$, trials because complete target identity was delivery 1500 ms later for cue-concurrent trials than for cue-first trials. It should be noted that observers in Experiment 2a had a shorter delay for cue-concurrent trails when compared to cue-first trails since the first presented target feature for Experiment 2a, color, lasted for 500 ms verses the longer 1000 ms for orientation for this experiment, this is reflected when comparing the average reaction time for cue-concurrent trials between this experiment and Experiment 4A (average 1500 ms vs. 1000 ms). Mean accuracy across conditions was 92.1%.

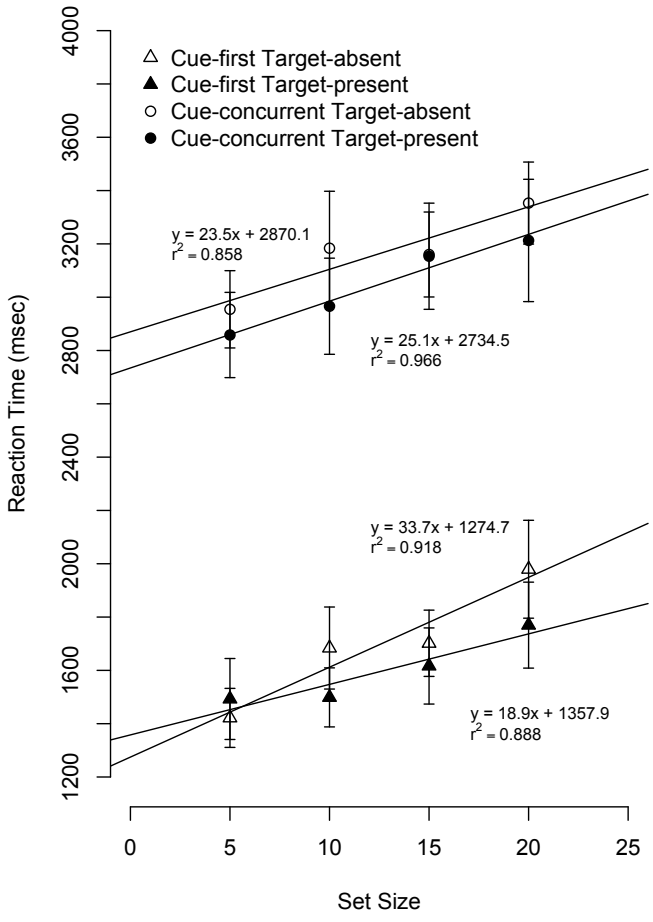

Figure 4.4: Results from Experiment 4C. Shown separately for target-present (filled symbols) and –absent trials (open symbols) for cue-first (triangles) and cue-concurrent (circles) conditions.

Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

The RT-by-set-size slopes reveal that the difference between color-first and orientation-first cue presentation was not significantly different both target-present, $t(24) = 0.77$, $p = .441$, and target-absent, $t(24) = -1.39$, $p = .165$, trials. It is interesting to note that although neither comparison was significantly different, the slope coefficients indicate that cue-concurrent trials did elicit a shallower, more efficient, search than cue-first trails for target-absent trials (33.7 ms/item vs. 23.5 ms/item) but not for target-present trials (18.9 ms/item vs. 25.1 ms/item). This is consistent with Spivey et al. (2001: Experiment 2) and Reali et al. (2006: Experiment 2) both of which found inconsistencies between target-present and –absent trials when target feature delivery was altered to have orientation presented first.

The finding here along with those of Spivey et al., (2001: Experiment 2) and Reali et al. (2006: Experiment 2) suggest that during the parallel stage of perceptual grouping, that precedes competition between inputs, color input produce stronger weights than orientation to the extent that observers are able to make faster and more accurate matches due to a more robust internal description of the information needed to guide action and awareness (Duncan & Humphreys, 1989). This indicates that the color feature encourages more intense perceptual grouping that in the end permits more spreading activation of appropriate targets and suppression distractors when compared to an orientation feature. Balota and Abrams (1995) show a similar pattern of results with word frequency. They discovered that motor movements exhibit more force in response to high frequency words (e.g., color words such as “red”) than low frequency words

(e.g., orientation words such as “horizontal”), suggesting that word frequency not only influences the time required to recognize a word, but also influences the subsequent response dynamics.

Additional studies have also observed asymmetrical effects between color and orientation words such as a study by Boucart & Humphreys (1997) where in a matching task that assessed the effect of semantic information on visual attention by manipulating the semantic relations among pictures surrounding a reference target (a line segment) that appeared first followed by two objects each containing a target and a distractor (a line segment). When observers were asked to match the referent line segment by color or orientation, semantic information was found to affect performance in the orientation-matching task, but not in the color-matching task (Boucart & Humphreys, 1997). This may be due to a more strongly weighted awareness of color, which made the color-matching task more resistant to outside influence. This supports the claim that the observed asymmetry with color- and orientation-first trials in incremental information (visual or linguistic) mediated visual search may be due to differences in the overall saliency of color descriptors over orientation.

CHAPTER FIVE

Eye-tracking Experiment

Introduction

Although the studies in Chapter 4 (Experiments 4A – C) provide us with further understanding of the effect of incremental information processing on visual search processing, it is still a bit of a stretch to generalize the process of integrating target identifying visual cues, albeit incremental, as the same cognitive process as integrating auditory linguistic information. In this chapter I observe, using eye-tracking methods, the mechanisms of visual search during a conjunction search task mediated by language. Dense-sampling techniques, such as eye-tracking and mouse-tracking, allow us to develop a more detailed mechanistic illustration of the temporal dynamics of phenomenon, such as with how visual information immediately impacts lexical and sentence processing (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

The Dense-sampling Approach

Ever since Tanenhaus and colleagues (1995) assessed the immediate mental processes that accompany spoken language comprehension by inspecting eye-movement recordings as observers manipulated real objects, dense-sampling methods such as eye-tracking have been used extensively to investigate the real-time interactions between visual information and language comprehension. Dense-sampling across the time course of a phenomenon is extremely informative with events that have longer time scales of cognition, such as with language comprehension. For example, when studying behavior over the course of hours, a time series of

thousands of reaction times has been shown to be much more informative than thousands of independent word recognition events. Rather than analyzing a thousand separate word recognition events as though they were independent of one another, the time series of those reaction times can be analyzed as one temporally-extended process of cognitive performance that reveals statistical patterns of fractal structure in the variance, which is one that is naturally predicted only by an interactive dynamical account of cognition (Kello, Beltz, Holden, & Van Orden, 2007; Van Orden, Holden, & Turvey, 2003).

It is crucial that our field resolve how and when multiple sources of information from different modalities interact, such as with language-mediated vision and vision-mediated language, to produce a response behavior, which provides additional understanding of the eventual response as well as the offline latency of the behavior. Thus dense-sampling methods (such as eye-tracking and reach-tracking) are, in addition to studying cognitive performance over an extended time period, also suited to investigate real-time cognitive processes in individual behaviors. It must be kept in mind that when measuring a change of state, a coarse time scale may present a change as more or less instantaneous, but with a finer time scale that same state change will appear gradual. Therefore in order to discover the processes or mechanisms that actually elicit a change of state, it is crucial that our science operate at a time scale that reveals the underlying gradualness of that change (Spivey, 2007). With these methods the immediately available gradations of partially active representations revealed by the oculomotor and skeletomotor system through their respective movements as well as the evolution of responses from the competition of multimodal interactions, over several hundred milliseconds, are observable and recordable.

Multimodal interactions, once only theorized, from studies that observed improvements in response speed from congruent auditory-visual information have been resolved using dense-sampling methods (e.g., Todd, 1912). Dense-sampling the response movement itself has produced evidence that supports a multimodal interaction between auditory-visual signals and motor output, which was only hypothesized by early reaction time data in experiments that utilized multimodal redundant-signals. In a study by Giray and Ulrich (1993) this redundant-signals effect was further examined by measuring reaction time in addition to the force of a response, which was measured with an apparatus that resembled an old-fashioned telegraph key, for both unimodal and bimodal trials. The findings revealed a decrease in reaction times as well as an increase in force for trials where multimodal information presented together. The authors used this to support the continuous integration of sensory information onto motor output, as opposed to the traditional assumption that once a motoric response is initiated it is impervious to sensory manipulation.

A wave of literature has emerged with the availability of dense-sampling methods that measure kinematic features of the motor movement during a response, which provide insight into the temporal dynamics of activation accumulation. In an experiment by Abrams and Balota (1991) participants make rapid limb movements in opposite directions in order to indicate whether a string of letters was a word or not. In addition to high lexical frequency speeding response and increasing force, they also found effects in movement duration, peak acceleration, final velocity, and initial velocity. These effects found early and continuously in behavior are extremely important for distinguishing between models of perception and cognition that make predictions regarding the intermediate stages of processing, where an early effect of velocity can

mean the difference between an encapsulated modular stage or a partially active distributed representation (Anderson, Chiu, Huette, & Spivey, 2010).

The temporal dynamics of motor output can be especially informative when the stimulus delivery itself is inherently extended in time as with language. One of the many concerns and core characteristics of investigating spoken language is the temporal nature of acoustic events (i.e., sounds arrive in a linear order to form words, sentences, and discourse). Dense-sampling methods such as eye-tracking can reveal probabilistic activations for visual referents available in the environment otherwise missed by coarser timescales. This close time-locking of saccades to speech allows for direct time-sensitive measurements of processing that can address fine-grained aspects of language comprehension (Tanenhaus et al., 1995).

Early reaction time data suggests that as a word unfolds over time it is initially ambiguous with other words that share similar sounding onsets, suggesting that even during the earliest moment of processing visual context influences word recognition and syntactic processing (Zwitserlood, 1989). This theory proposes that for a brief period after the onset of a word, all words beginning with the same phonemic input compete, but as more phonemic input is received the target becomes less ambiguous, causing words to drop out of the competition (Marslen-Wilson, 1987). To test this hypothesis Allopenna and colleagues (1998) (see also, Spivey-Knowlton, 1996) presented observers with visual displays containing four items: the target (e.g., a beaker), an onset-competitor (a beetle), a rhyme competitor (a speaker), and an unrelated referent (a carriage). Eye movements were recorded as observers heard and responded to instructions like, "Pick up the beaker." The results revealed that during the first half of the spoken target word, the probability of fixating the target or competitor both gradually increased equally but around the offset of the spoken target word, the proportion of looks to the target

began to rise sharply and subsequently decreasing the proportion of looks to the competitor. Thus, early in the auditory stimulus, before the target has been uniquely identified, competition between the partially active representations manifests itself in the eye movement patterns. Furthermore, the data revealed a greater probability of fixations to the rhyme competitor than to the neutral distractor object, which is consistent to the rhyme competitor effects predicted by McClelland and Elman's (1986) interactive-activation neural network simulation of speech perception, detailed in Chapter 6.

Some sentences, like words, are just as temporarily ambiguous across time. Dense-sampling methods have also helped to elucidate the processing of these ambiguous sentences. Many early investigations in the processing of temporarily ambiguous sentence looked at sentences in isolation, with results supporting a modular process. For instance, in this sentence, "Since Jay always jogs a mile doesn't seem far," inflated reading times were observed when readers encountered the disambiguating word, "doesn't" (Frazier & Rayner, 1982).

These early researchers postulate the increased reading time was the manifestation of an encapsulated syntactic processing module separate from other perceptual and cognitive systems, thus arguing for its autonomy from other information sources, such as semantics and visual information. However, dense sample methods illustrate a drastically different process for syntactically ambiguous sentences in conjunction with a visual scene (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). When observers hear a temporarily ambiguous sentence such as "Put the apple on the towel in the box" while viewing a scene containing an apple (target object), a towel (incorrect goal location), a box (correct goal location), and a flower (neutral unrelated referent), they experience the garden-path effect, temporarily interpreting "…on the towel…" as the destination of the putting event, only to later realize that this parse is ultimately

incorrect. In these experiments, the garden-path effect manifests itself as inflated reading times at the point of disambiguation, manifesting itself as an increased probability of saccades to the incorrect destination (the towel). Trials containing unambiguous sentences like "Put the apple *that's* on the towel in the box" did not exhibit an increased probability of looks. Furthermore it has been found that during spoken word recognition eye movements are made not only to referred target as well as to competitor objects with phonologically similar names (Allopenna et al., 1998; Spivey-Knowlton, Tanenhaus, Eberhard, & Sedivy, 1995) but also to semantically related properties (Huettig & Altmann, 2005; Yee & Sedivy, 2006) and visually similar shapes (Huettig & Altmann, 2007). These findings reveal that in a "visual world" paradigm, saccades appear to be driven by partially active representations.

Even more samples per second can be collected when one records the temporal dynamics of a reaching movement, again revealing competition between multiple potential movement destinations (Tipper, Howard, & Jackson, 1997). One should note that reach movements are often initiated after a first eye movement, and therefore this compensatory strength and weakness (denser sampling but later measurement in reach-tracking) should encourage one to treat these two methods as complementary, not adversarial as continuous mouse-tracking like eye-tracking, provides support for the concept of continuous parallel processing in cognition (Magnuson, 2005).

Another popular and extremely informative dense-sampling approach is found with reach-tracking, particularly now with the development of a computer-mouse-tracking paradigm by Spivey, Grosjean, and Knoblich (2005), which has simplified the method and made it readily available. This method of sampling full mouse-movement trajectories at 60 Hz, and looking at their curvatures, velocity and acceleration profiles, distribution of maximum deviations, as well

as measures of entropy or disorder, has aided in distinguishing between alternative computational simulations of the temporal dynamics of ambiguity in spoken word recognition (Spivey, Dale, Knoblich, & Grosjean, 2010; van der Wel, Eder, Mitchel, Walsh & Rosenbaum, 2009). Computer-mouse tracking effects have also been found for a variety of studies including attention (Song & Nakayama, 2006), sentence processing (Farmer, Anderson, & Spivey, 2007), semantic categorization (Dale, Kehoe, & Spivey, 2007), color categorization (Huette & McMurray, 2010), as well as the time course of high-level cognitive processes, such as fuzzy-truth decision making (McKinstry et al., 2008), social preferences such as racial biases (Wojnowicz, Ferguson, Dale, & Spivey, 2009), and gender recognition (Freeman, Ambady, Rule, & Johnson, 2008) are reflected in the trajectory of mouse movements.

The reason that dense-sampling of motor output during a response is so informative is that it allows one to observe the cognitive process of the motor system before it reaches completion, sampling the system as it generates a movement associated with the results of that cognitive process. Providing a simple way to record the evolution of multifarious neural activity patterns associated with a given cognitive process over the course of several hundred milliseconds, which ultimately influences the initial generation of movement in the oculomotor cortex (Gold & Shadlen, 2000) and primary motor cortex (Cisek & Kalaska, 2005). The presence of reciprocal neural projections between these motor areas and frontal cortex suggests an undivided process whereby cognition and action are not quite separable (e.g., Barsalou, 2008; Chemero, 2009; Hommel, 2004; Pulvermüller, 2005; Spivey, 2007).

With new advances in dense-sampling techniques, as with eye-tracking researchers can now construct robust illustrations of real-time cognitive processes such as identify fixation rich regions over a time period. These illustrations, called "heat maps" because of its use of color in

representing quantity and duration of eye fixations on specific area of a search display, resemble the graphical representation of data in physical sciences, where the individual values contained in a matrix are represented as colors. Generally, areas where users look the most are colored red, yellow areas indicate fewer fixations, and least-viewed areas are colored blue; gray areas generally denote areas without fixations. I use this method to investigate differences in eye-movement and –fixations during a linguistically mediated conjunction search task.

Method

All methods, stimuli, and procedures are identical with Experiment 1 from Reali et al. (2006), which utilized a mixed within subjects design, with the exception that the search displays used are the same for all participants but presented in random order. Half of the trials are presented in the A/V-concurrent condition and the other half presented in the auditory-first condition. Participants were randomly assigned to one of two groups. Participants in the first group, *Group A*, all received the same search display in one of the two conditions (A/V-concurrent or auditory-first) and the other half in the remaining condition. Participants in the second group, *Group B*, received the same search displays but presented in the opposite condition as the participants in the Group A, such that any given display was presented as both auditory-first and A/V-concurrent across both groups. This allows for the between subject comparison of search strategies among SOAs for any given search display. Target-present and –absent trials along with the four set sizes (5, 10, 15, & 20) appeared randomly and equally. Novel to this experiment is that while performing in the conjunction search task observers' eye-movements were recording for all trials using an Eye-Link II head mounted eye-tracker.

Participants

Sixty-eight undergraduate students from the University of California, Merced received partial course credit for participation in this experiment. All of the participants had normal, non-corrected, vision as well as normal color perception. As with the aforementioned experiments (Experiment 1 – 4), those participants who failed to complete the experiment with at least 80% accuracy were omitted from the analysis. Three participants, two of which scored close to 50% accuracy and clearly were not invested in the task, did not perform to these standards and were subsequently removed form the analysis. It should be noted that although the amount of participants that were removed from the previous analyses for the other experiments (Experiment 1-4), as a result of our accuracy requirements, were not a large portion of the entire data set (range: 8.47% - 36.8%, M = 17.1%, SD = 8.16%), the amount here (three participants) is a much smaller portion (4.41%) of the data set than previously observed. This improvement in performance may be the result of using the eye-tracking system or the presence of an experimenter monitoring the equipment during participation in the experiment.

Stimuli and Apparatus

Identical pre-generated search displays were used for each observer. The same stimulus bars were used that subtended 2.8° X 0.4° of visual angle and neighboring bars were separated from one another by an average of 2.0° of visual angle. The green and red bars had the same luminance of 13.4 cd/m². Appearance of the target object in the four quadrants (top-left, top-right, bottom-left, and bottom-right) as well as the type of target (e.g. green horizontal), and set sizes of objects (5, 10, 15, & 20) was controlled for to insure they appeared equally. Observers were randomly assigned to participate in one of two groups (A or B). The two groups were

indistinguishable but differed in that identical search displays were presented in an auditory-first trial for one group and an A/V-concurrent trial for the other group. This was achieved by utilizing two experiment files that provided a slight variation of the experiment in which each display from was switched either from being a control condition to A/V-concurrent or vice versa.

In half of the trials, a spoken query (e.g., "Is there a red vertical?") informed participants of the targets' identity before they were presented with the visual display (auditory-first condition), and in the other half of the trials, the first adjective of the spoken query coincided with the appearance of the visual display (A/V-concurrent condition). The identical 1000 ms prelude recording ("Is there a…") was used with two target-identifying adjectives (color and orientation), which together averaged 1500 ms. As a result of this design our usual RT-by-set-size analysis would continue to be a mixed within subjects analysis but the comparison of search strategies between groups, via eye-tracking data, would then be a mixed between group analysis.

An Eyelink II head mounted video-based eye-tracker (SR Research Ltd., Mississauga, Ontario, Canada) with a temporal resolution of 250 Hz and a spatial resolution of 0.025º recorded eye movements by tracking pupil and the corneal reflection. The video-based eye-tracker used two infrared LEDs mounted on the headband to illuminate each eye. Tracking was monocular although viewing was binocular. It classified an eye movement as a saccade when its distance exceeded 0.2 degrees and its velocity reached 30 degrees per second or when its distance exceeded 0.2 degrees and its acceleration reached 9500 degrees per second squared. The displays were generated using Mathworks MATLAB software and the experiment was designed using Experiment Builder by SR Research Ltd. No additional software packages were used. Stimuli were presented on a 22" ThinkVision LCD monitor with 1280 x 1024 resolution. The

prerecorded speech queries, recorded from the same female speaker, are identical to Experiment 1 and 2 and were presented through Harmon Kardon HK206 desktop computer speakers.

Procedure

The Eyelink eye-tracker was calibrated using the standard nine-point calibration method for each participant. Calibration was followed by 16 practice trials to allow participants to familiarize themselves with the task and wearing the head mounted eye-tracker. The experiment consisted of 128 trials containing an equal amount of auditory-first control and A/V-concurrent trials mixed together in a randomized order for each participant. Observers were instructed to keep their fingers resting on the marked response keys and to respond as quickly and accurately as possible by pressing "Yes" and "No" if the target was present or absent, respectively. Before each trial, participants were required to fixate their gaze on a fixation cross in the center of the screen so that the experiment would continue on to the next trial; this was also used to as a "drift correct," which verified that the initial calibration remained valid. Participants initiated each trail by pressing the space bar while fixating on the fixation cross. If the drift correct was invalid, the trial would not begin and the experimenter was prompted, very rarely, to recalibrate the participant. Calibration varied between five to ten minutes and the experiment itself lasted approximately 15 minutes; the entire experiment lasted approximately 30 minutes.

Results and Discussion

As with Experiment 1, a hierarchal linear model (HLM) was used for this analysis because it accounts for the unbalanced N and repeated measures design, as the result of data culling. To fulfill the assumption of distribution normality the inferential statistics were

performed on log-transformed reaction times, as reaction time response data is bound on the left but not the right thus naturally positively skewed (Luce, 1986). However, descriptive statistics (slopes and intercepts of reaction times in milliseconds) continue to be reported from an untransformed HLM.

In this experiment, we replicated previous findings demonstrated by Spivey et al. (2001) and Reali et al. (2006) with a within subjects design. Figure 5.1 shows the RT-by-set-size functions for target-present (filled symbols) and target-absent (open symbols) trials in the A/V-concurrent (triangles) and auditory-first control (circles) conditions. Next to each graph line is the best-fit linear equation and the proportion of variance accounted for (r^2); the error bars indicate standard error of the mean. The RT-by-set-size function are highly linear for the auditory-first condition in both target-present, $r^2 = .561$, and target-absent trials, $r^2 = .951$, as well as for the A/V-concurrent condition for target-present trials, $r^2 = .773$, and target-absent trials, $r^2 = .903$, which is typically observed in standard conjunctions search tasks. Mean accuracy across all trials is 95.0%, which is consistent with previous observations (Spivey et al., 2001; Reali et al., 2006, Chiu & Spivey, 2012).

As expected the slopes of the RT-by-set-size functions reveal that A/V-concurrent conditions produce more efficient visual search when compared with the auditory-first conditions (see fig. 5.1). The HLM analysis revealed significantly shallower slopes for the A/V-concurrent condition compared to the auditory-first condition in target-present trials (5.5 ms per item vs. 8.7 ms per item), $t(64) = -3.23$, $p < .001$, and target-absent trials (29.4 ms/item vs. 45.2 ms/item), $t(64) = -10.24$, $p < .001$, as previously observed by Spivey et al. (2001), Reali et al. (2006), and Chiu and Spivey (2012).

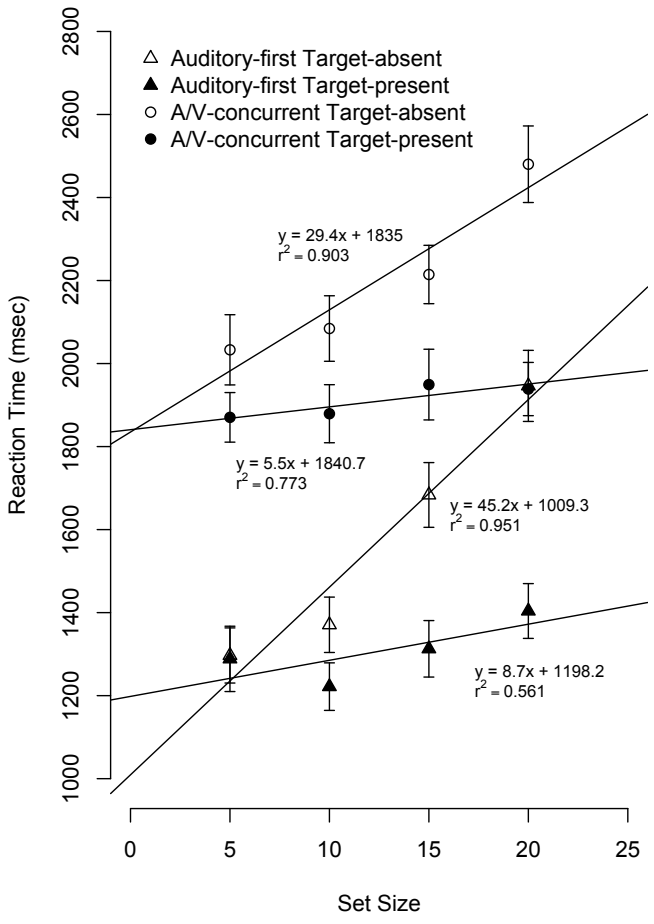

Figure 5.1: Results from Experiment 5. Shown separately for target-present (filled symbols) and –absent trials (open symbols) for cue-first (triangles) and cue-concurrent (circles) conditions. Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

Interestingly, the ratios of target-present and -absent trials in this experiment for both conditions are smaller than with previous experiments (Experiment 1-4; Spivey et al., 2001; Reali et al., 2006; Chiu & Spivey, 2012). The ratios are closer to 1:5 than the previously observed 1:2 for the auditory-first condition (8.7 ms/item vs. 45.2 ms/item) and the A/V-concurrent condition (5.5 ms/item vs. 29.4 ms/item). The smaller ratio may be the result of a

more stringent regulation of object distribution across the search display, which can in principle allow for faster, more accurate, and possibly stronger activation of target objects and/or suppression of distractors. It should be noted that the target-absent trials RT-by-set-size slopes in this experiment remain similar to those of past experiments but target-present slopes are smaller thus what ever the effect that produced the smaller ratios appear to primarily affect target-present trials. Additional testing is necessary to investigate this claim. Overall mean reaction time, as well as y-intercepts, were significantly slower in A/V-concurrent conditions because complete delivery of target identity was delayed by approximately 1500 ms relative to the auditory-first condition for both target-present trials, $t(64) = 184.79$, $p < .001$, and target-absent trials, $t(64) = 250.27$, $p < .001$.

We see the results of this experiment continue to show observers were able to find the target object in a way that was substantially less affected by the number of distractors simply by adjusting the timing of spoken query, so that the two target-feature words are heard while the visual display was visible. It appears that the incremental nature of speech processing allows the visual search process to begin when only a single feature of the target identity has been heard. When the initial feature is identified the search proceeds in an efficient nearly-parallel fashion so when the second adjective is heard, a substantial amount of the target identification process has been completed – and thus the presence of multiple distractors is less disruptive.

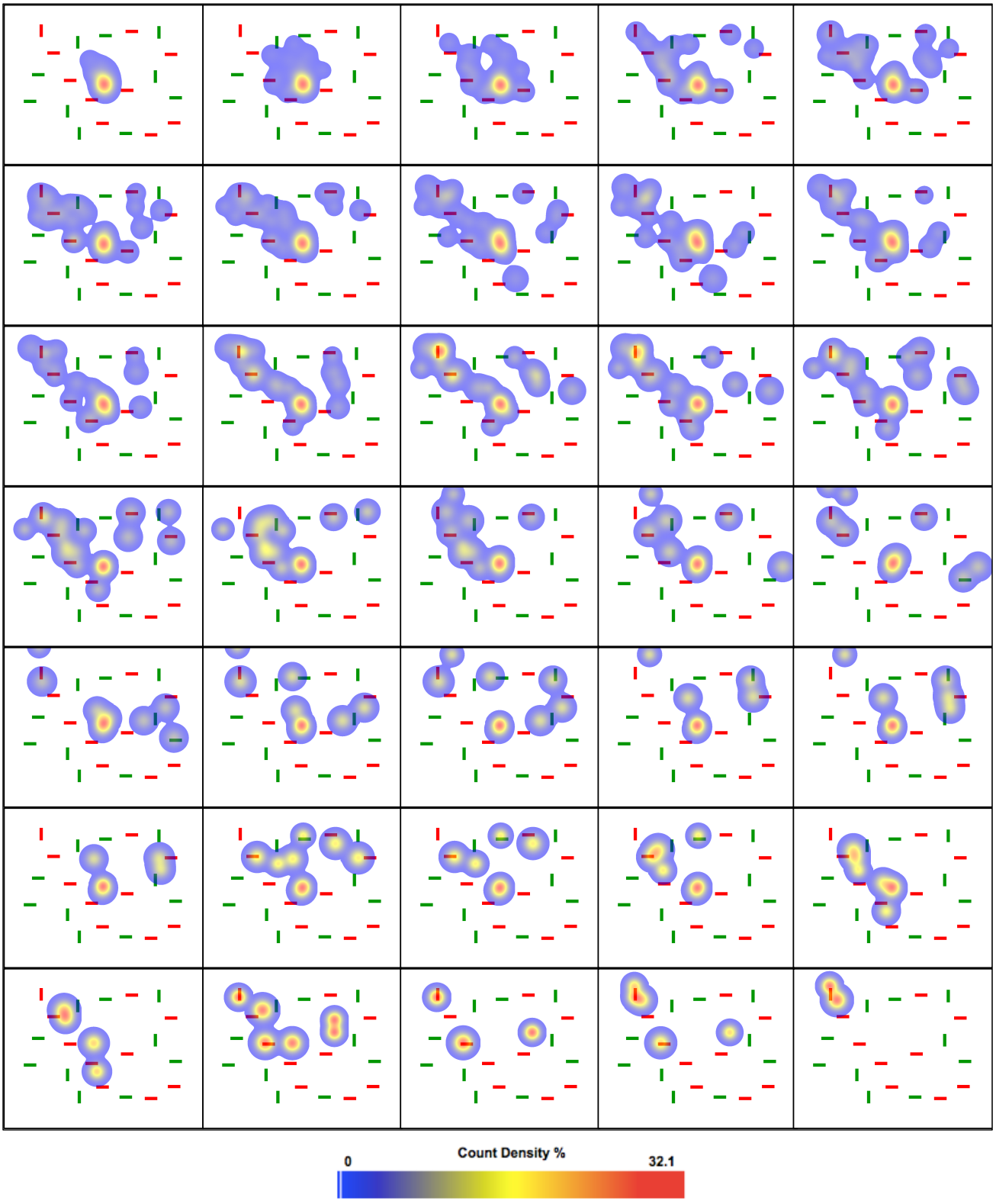

Figure 5.2: Eye-tracking results for Experiment 5. Search displays are overlapped with a heat map representing fixation activity (blue = low, yellow = medium, and red = high) for a target-

present trial with a set size of 20. The target in this trial is a red vertical bar located in the top-left of the search display. Fixations for this figure are comprised of participants in Group B who received this search display in the auditory-first condition. Each frame represents 100 ms timesteps.

Of primary and novel interest in this experiment is the analysis of eye-movement patterns during the linguistically mediated conjunction search task that has been replicated here and in other studies (Spivey et al., 2001; Reali et al., 2006; Chiu & Spivey, 2011). Figure 5.2 and 5.3 show eye fixation patterns over a target-present (target is a red vertical bar) trial with a set size of 20 for both the A/V-concurrent (see fig. 5.3) and auditory-first control (see fig. 5.2) condition. Participants in Group A were presented with this trial in the A/V-concurrent condition; subsequently participants in Group B were presented with this trial in the auditory-first condition. Thus the eye-fixations portrayed in figure 5.2 are comprised solely of participants from Group B and Group A for figure 5.3. Each frame depicts eye-fixations at 100 ms time intervals. At first glance it appears that eye-fixations, over the duration of this trial, are more efficient for the A/V-concurrent condition (see fig. 5.3) than in the auditory-first condition (see fig. 5.2). There appears to be fewer fixations in figure 5.3, for the A/V-concurrent condition, than in figure 5.2, for the auditory-first condition, and the fixations are seemingly directed mostly to red objects throughout the trial. This is not surprising since the first mentioned target-identifying adjective, delivered concurrently with display onset, is the color red as the target for this trial is a red vertical bar located in the top-right of the search display. Thus it appears that observers in Group A use the concurrent linguistic query to their advantage and perform a smaller more efficient search of a subset of objects for the unique target than observers in the

auditory-first condition, which leads to a more efficient search strategy that is less affected by quantity of distractors.

This improved efficiency is supported by the significantly fewer number of fixations observed across each trial when presented in the A/V-concurrent condition (M = 13.08, SD = 6.89) than in the auditory-first condition (M = 17.23, SD = 23.76), $t(64) = 17.18$, $p < .001$. From beginning to end of the trial for figure 5.3 in the A/V-concurrent condition there are few fixations, if at all, to distractors that do not match the color of the inquired target (see fig. 5.4), which supports the idea that upon hearing the first target-identifying adjective a rapid parallel-like search process weeds out the conflicting colored distractors and subsequently increases the saliency of fitting objects. This phenomenon is not observed with the auditory-first condition. Further analysis finds that the number of fixations are significantly smaller for A/V-concurrent than auditory-first condition across all four set sizes, $f(64) = 116.7$, $p < .001$ (see Table 5.1 for values). It should be noted that the following descriptive statistics reported here solely involve target-present trials because, as mentioned before, search strategies in target-absent trials have been found to differ from that of target-present and are notoriously difficult to simulate, which is reflected by research in this area that has not uncovered much.

Number of Fixations for Target-present Trials

Set Size	A/V-concurrent		Auditory-first	
5	M = 12.1	SD = 5.1	M = 13.4	SD = 7.3
10	M = 12.7	SD = 4.9	M = 14.0	SD = 5.7
15	M = 14.2	SD = 10.0	M = 14.5	SD = 6.2
20	M = 13.2	SD = 5.9	M = 14.5	SD = 5.4

Table 5.1: Number of fixation mean and standard deviation values in target-present trials for the A/V-concurrent and auditory-first condition across all four set sizes.

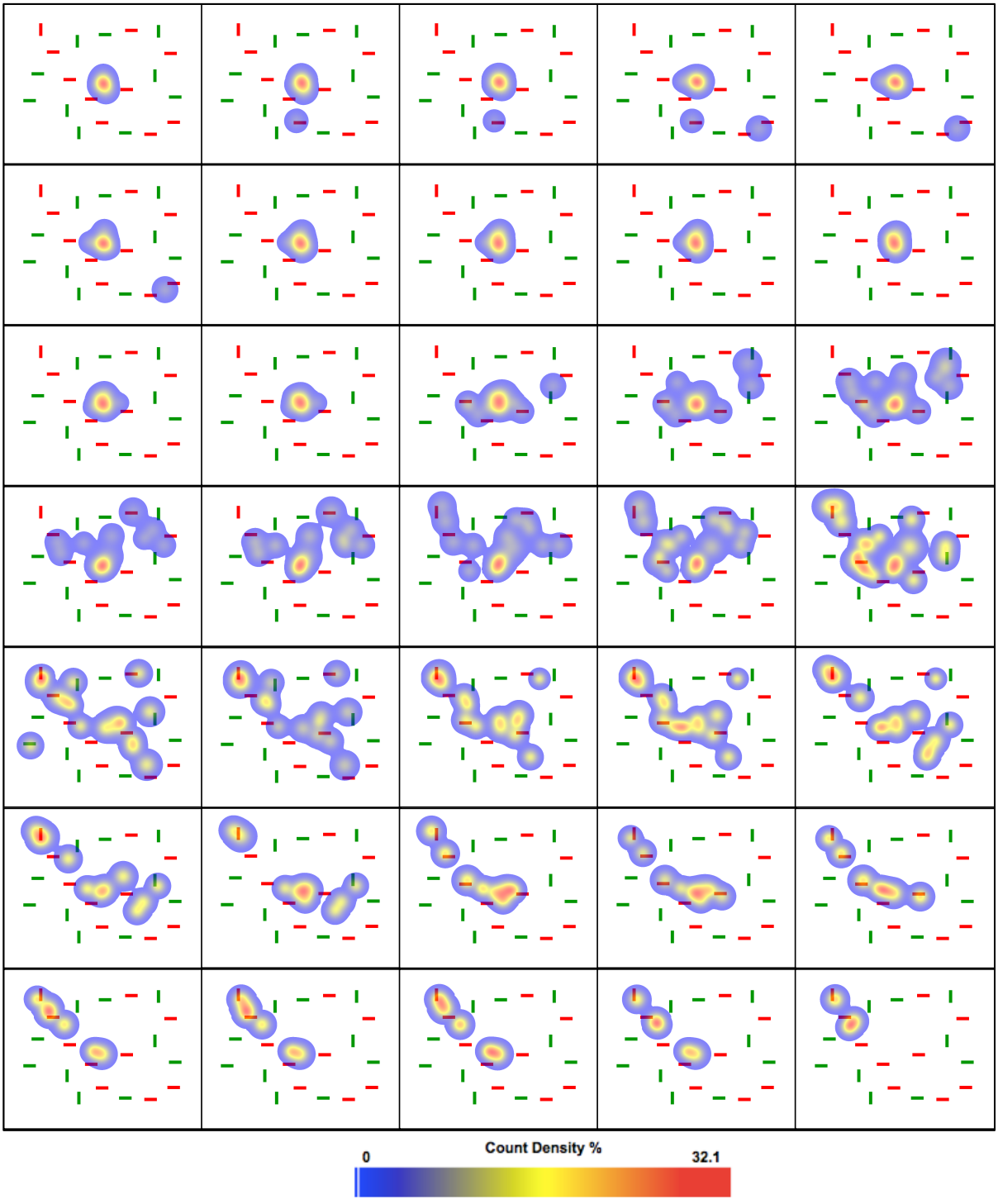

Figure 5.3: Eye-tracking results for Experiment 5. Search displays are overlapped with a heat map representing fixation activity (blue = low, yellow = medium, and red = high) for the same target-present (red vertical bar) trial with a set size of 20 depicted in Figure 5.2. Fixations for this figure are comprised of participants in Group A, who received this search display in the A/V-concurrent condition. Each frame represents 100 ms timesteps.

In addition to the difference in number of fixations the average duration of each fixation is also significantly shorter for the A/V-concurrent condition, 335.66 ms (*SD* = 395.00), than for the auditory-first condition, 382.18 ms (*SD* = 503.44), $t(64) = 7.33$, $p < .001$. Since observers in the auditory-first condition receive both target-identifying adjectives before the onset of the search display, it is possible that they are judging each object fixated on to both features at once in search of the unique target, which would explain the longer fixation durations when compared to A/V-concurrent trials. Since observers in the A/V-concurrent condition have already isolated attention to objects that match the identified color feature they would only have to judge each fixated object on one remaining feature, the second uttered target-identified adjective (orientation), which would be speed the process. Further analysis of this effect has found that fixation durations are significantly shorter for A/V-concurrent trials when compared to auditory-first trials across all four set sizes (5, 10, 15, & 20), $f(64) = 39.1$, $p < .001$ (see Table 5.2 for values).

Fixation Duration for Target-present Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 361.6	*SD* = 414.4	*M* = 421.6	*SD* = 547.5
10	*M* = 334.6	*SD* = 376.2	*M* = 375.5	*SD* = 488.0
15	*M* = 327.2	*SD* = 397.6	*M* = 368.3	*SD* = 484.9
20	*M* = 320.8	*SD* = 391.2	*M* = 367.9	*SD* = 494.5

Table 5.2: Mean and standard deviation values of fixation durations, measured in milliseconds, for A/V-concurrent and auditory-first trials across all four set sizes.

Unexpectedly, eye-fixation duration across both conditions (A/V-concurrent and auditory-first) is found to decrease (392.6 ms, 356.2 ms, 349.0 ms, and 345.8 ms, respectively) as set size increases from 5, 10, 15, and 20; an analysis of this effect finds it to be significant, $f(64) = 24.83$, $p < .001$. This phenomenon may be the result of a desire to respond quickly as instructed at the beginning of the experiment. Thus when observers see there are more objects, they speed their search strategy, surprisingly with no significant affect on accuracy as set size increases from 5 to 20 (5.1%, 4.7%, 5.0%, & 6.2%, respectively, $f(64) = 0.46$, $p < .497$), and improve their response time in relation to the increased set size. This effect requires further investigation to understand the mechanisms that cause it.

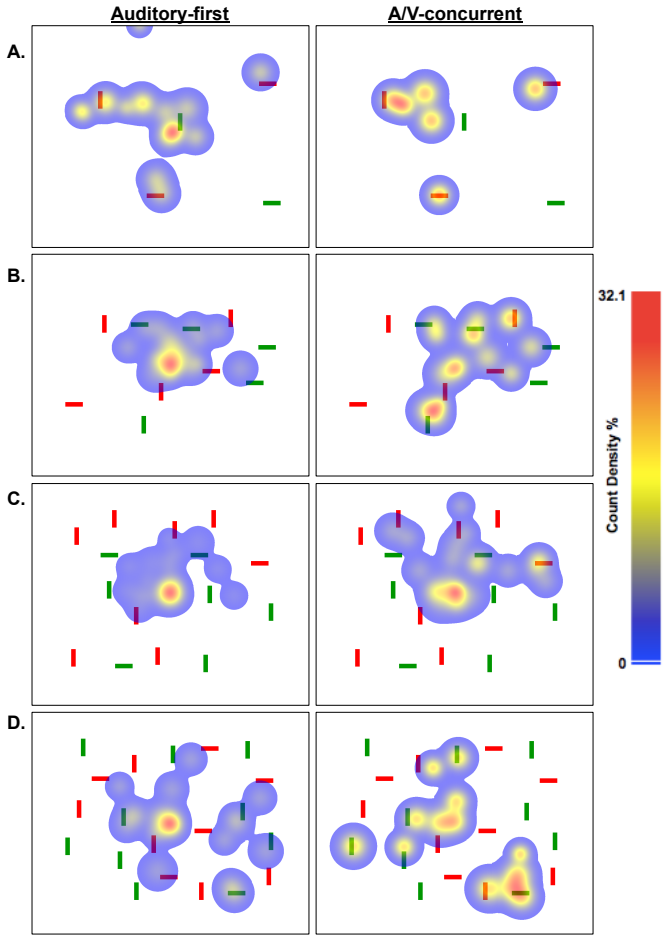

Figure 5.4: A closer look at the eye-tracking results from Experiment 5. Target-present trials are shown separately for auditory-first control and the A/V-concurrent trials. Search displays are overlapped with a heat map representing fixation activity (blue = low, yellow = medium, and red = high). A single 100 ms time period is depicted for each set size: 3100-3200 ms for 5 (A), 2400-2500 ms for 10 (B), 1700-1800 ms for 15 (C), and 2600-2700 ms for 20 (D). Targets for each trial are as follows: 5 = red vertical, 10 = green vertical, 15 = red horizontal, and 20 = green horizontal.

Interestingly the length of saccades, rapid ballistic movements of the eye between fixation points, measured in amplitude (degrees of visual angle) are significantly longer for A/V-

concurrent trials, 5.24 (*SD* = 6.88), than for auditory-first trials, 4.73 (*SD* = 7.03), $t(64) = -4.95$, $p < .001$. If it is the case that observers in an auditory-first trial are performing a traditional serial search process, where they are attending to each object wholly and discretely to judge it to be the inquired target object, then we can presume their attention would jump from one object to the next closest object in some sort of ordered fashion to optimize their search strategy until the target object was found. This scenario would describe why saccade amplitudes are shorter for auditory-first trials than for A/V-concurrent trials. Because half of the objects are effectively ruled out in A/V-concurrent trials, moving gaze from one object to another plausible object, that matches the mentioned color, would probabilistically be longer than simply shifting to the next closest object. Further analysis of this effect has found that saccade amplitudes are significantly shorter for auditory-first trials when compared to A/V-concurrent trials across all four set sizes (5, 10, 15, & 20), $f(64) = 14.21$, $p < .001$ (see Table 5.3 for values).

Saccade Amplitude for Target-present Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 5.33	*SD* = 7.00	*M* = 5.22	*SD* = 8.06
10	*M* = 4.45	*SD* = 6.14	*M* = 4.38	*SD* = 6.55
15	*M* = 5.32	*SD* = 6.91	*M* = 4.43	*SD* = 6.50
20	*M* = 5.84	*SD* = 7.36	*M* = 4.96	*SD* = 7.01

Table 5.3: Mean and standard deviation of saccade length measured in amplitude (degrees of visual angle) for A/V-concurrent and auditory-first trials across all four set sizes.

Of occasional interest with eye-tracking experiments is an analysis of saccade velocity, the speed of a saccade in any given direction. Although average saccade velocity, but not peak saccade velocity, was found to be significantly different between A/V-concurrent and auditory-first trials across the four set sizes, $f(64) = 5.57$, $p = .004$, there was no discernable pattern observed (see Table 5.4). This is not surprising because the current explanation does not provide any predictions that would produce a pattern in saccade velocity. Neither the proposed hybrid parallel search strategy of A/V-concurrent trials nor the traditional serial search strategy would elicit any sort of pattern that would manifest via saccade velocity, which the results support.

Saccade Velocity for Target-present Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 109.7	*SD* = 154.0	*M* = 116.9	*SD* = 240.7
10	*M* = 102.5	*SD* = 136.6	*M* = 102.5	*SD* = 188.6
15	*M* = 122.4	*SD* = 210.3	*M* = 103.8	*SD* = 169.1
20	*M* = 126.1	*SD* = 186.4	*M* = 111.1	*SD* = 193.9

Table 5.4: Mean and standard deviation of average saccade velocity for A/V-concurrent and auditory-first trials across all four set sizes.

The findings thus far are consistent with the inference made earlier (Chapter 1-4) that search strategies differ drastically between A/V-concurrent and auditory-first trials. Figure 5.4 shows a 100 ms time slice for each set size (5, 10, 15, & 20) shown separately for A/V-concurrent and auditory-first. You can clearly see that for the same time slice fixations in the

A/V-concurrent trials are primarily focused on color-matched objects, while fixations in the auditory-first trials do not appear to exhibit any pattern (see fig. 5.4). This pattern is consistent across all of the trials. The following analyses investigate the assumption that A/V-concurrent target-feature delivery does indeed elicit a different and more efficient search strategy than auditory-first target-feature delivery.

For a target-present trial with a set size of 20 where the target-object is a green horizontal bar, we see that the amount of time (measured in milliseconds) that is spent fixated on an object (dwell time) is longer for green (47.6 ms) objects than for red (19.6 ms), $t(64) = -3.81$, $p < .001$. Meaning that in a trial where the color of the target-object is green, observers spent more time fixating on green objects than red, regardless of when the display was presented in relation to target identification. Moreover, the comparison of dwell time between green and red objects across A/V-concurrent and auditory-first trials, reveal that observers spend more time fixating on green objects than red when target-identity is presented concurrently with display onset in an A/V-concurrent trial (46.5 ms vs. 17.5 ms) than when target-identity is presented prior to display onset in an auditory-first trial (48.4 ms vs. 21.2 ms), $f(64) = 7.33$, $p < .001$. Thus, when presented with target-identity concurrently with the search display observers spend more time in relation (difference of 29 ms) fixated on color-matched objects than non-matched distractors than when presented with target-identity prior to the search display (difference of 27.2 ms).

Up until now our discussion of the eye-tracking results for this experiment have been restricted to target-present trials because, as mentioned before, target-absent trials utilize a different search strategy that includes a mechanism for terminating search before choosing to respond “No” to the query of whether a target object is present or not. There has been notoriously little progress in the area of target-absent visual search. Computational cognitive

scientists have yet to design a model that can consistently and accurately simulate the behavior of a target-absent search process. Nevertheless, the RT-by-set-size search function for target-absent trials, like target-present trials, reveal a more efficient search between A/V-concurrent and auditory-first trials (29.4 ms/item vs. 45.2 ms/item), $t(64) = -10.24$, $p < .001$, as previously mentioned in this chapter and observed in previous experiments of this nature (Spivey et al., 2001; Reali et al., 2006; Chiu & Spivey, 2012).

In a target-absent trial, search must initiate the same way a target-present search does since an observer has no way of knowing that a target is absent or present until completing their search. If it were the case that observers terminated search after exhausting all of the objects in the entire search display, scientists would have already been able to successfully simulate this mechanism, as it would be a simple function of set size and possibly complexity of target features. Unfortunately, this is not the case. Instead, observers terminate search and respond with an absent response before an exhaustive search of every object can be completed. No identifiable pattern has yet been found that fit this phenomenon. Yet it is well documented that target-present trials elicit a faster and more efficient search than target-absent trials. We see this here when target-present reaction time data is compared to target-absent collapsed across the two conditions, the mean across each set size of target-present trials (1580.3, 1553.3, 1632.3, & 1673.1 ms, respectively) is significantly smaller than target-absent trials (1666.7, 1730.0, 1949.5, & 2216.0 ms, respectively), $f(64) = 305.7$, $p < 0.001$.

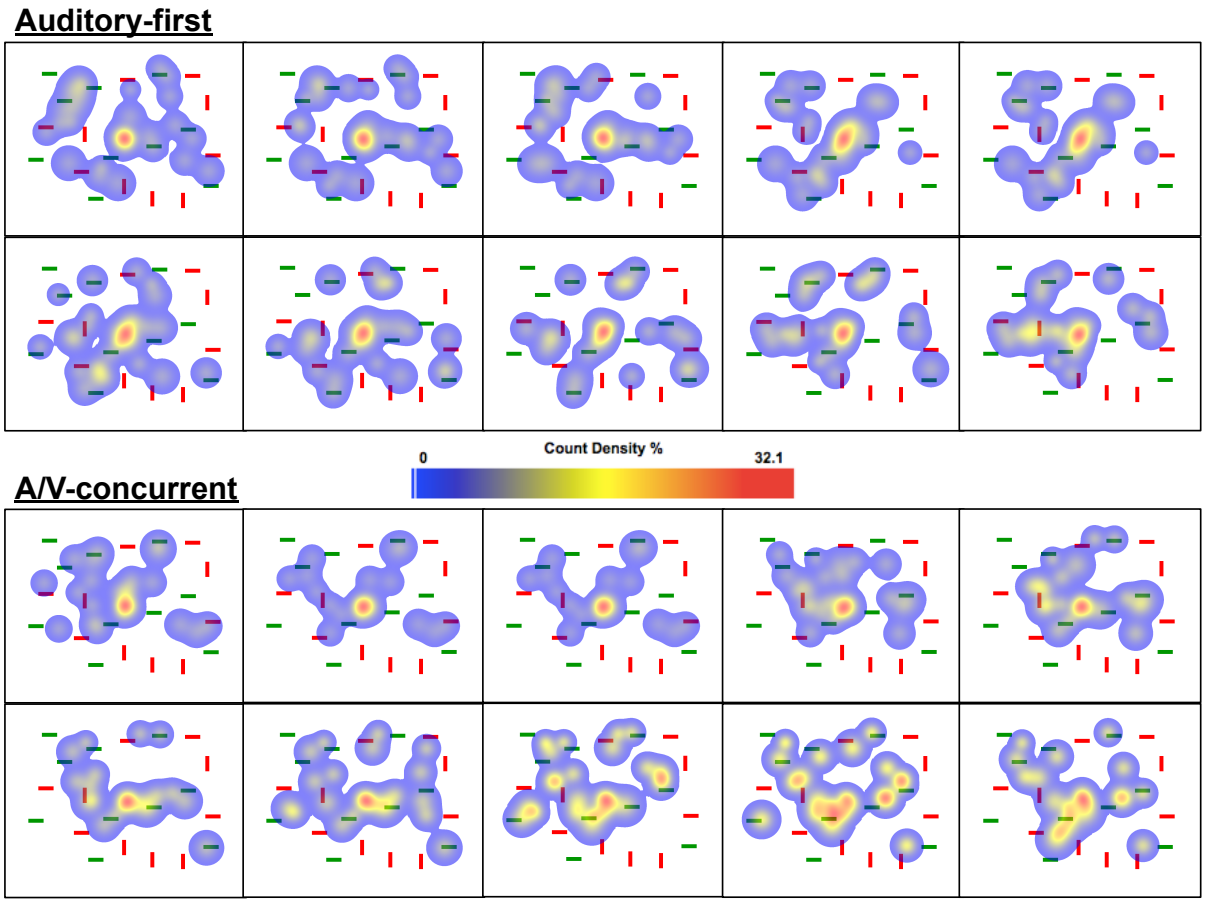

Figure 5.5: Results from Experiment 5 depicting eye-fixations for a target–absent (green vertical bar) trial with a set size of 20, shown separately for auditory-first and A/V-concurrent conditions. Each search display is overlapped with a heat map representing fixation activity (blue = low, yellow = medium, and red = high); each frame represents 100 ms time period.

The eye-tracking data from this experiment supports the claim that target-absent trial begins in a similar fashion as a target-present trial. Figure 5.5 displays ten frames, each with 100 ms of eye-fixations, from a target-absent trial with a set size of 20 shown separately for the A/V-concurrent and auditory-first condition. The ten frames are plucked from the middle of the search process, after search has already been initiated. You can distinguish differences between search-strategies at the beginning of the ten-frame set for this trial. The main distinction between

the two conditions is that fixations in the A/V-concurrent condition are primarily focused on objects that share the color of the inquired target, a green vertical bar (see fig. 5.5), and fixations for the auditory-first condition appear to be further dispersed and not particularly focused on merely green objects. As the trial progresses though we see that fixations in the A/V-concurrent condition begin to spread to non-green objects, mimicking the auditory-first search strategy. At this point, after exhaustively searching through the smaller subset of color matched objects and failing to identify an unique target object, it is possible that observers switch tactics, which is either due to or caused by the decreasing saliency of the color matched subset.

Some benefits of a concurrent linguistic delivery of target-identifying features still remain but to what amount? As with target-present trials, we find significantly fewer number of fixations in target-absent trials when presented in the A/V-concurrent condition (M = 14.64, SD = 6.89) than in the auditory-first condition (M = 15.53, SD = 23.76), $t(64) = 10.17$, $p < .001$ (see Table 5.5 for individual values). Thus A/V-concurrent feature delivery continues to generate fewer fixations than presenting both features prior display onset during a target-absent trial.

Number of Fixations for Target-absent Trials

Set Size	A/V-concurrent		Auditory-first	
5	M = 12.8	SD = 5.7	M = 13.8	SD = 5.9
10	M = 15.1	SD = 7.4	M = 14.7	SD = 6.5
15	M = 14.5	SD = 5.7	M = 15.1	SD = 5.2
20	M = 15.8	SD = 5.1	M = 17.8	SD = 7.8

Table 5.5: Mean and standard deviation values for the number of fixations in target-absent trials for the A/V-concurrent and auditory-first condition across all four set sizes.

Fixation durations trends in target-absent trials are consistent with target-present trials showing that the average fixation duration for target-absent trials are significantly shorter for the A/V-concurrent condition, 312.75 ms (*SD* = 389.3), than for the auditory-first condition, 365.55 ms (*SD* = 507.58), $t(64) = 8.65$, $p < .001$. As with target-present trials this effect continues to be significant for target-absent trials across all four set sizes, $f(64) = 60.88$, $p < .001$ (see Table 5.6). This consistency is not surprising since observers in an auditory-first trial, regardless of a present or absent target object, continue to receive both target-identifying adjectives before the onset of the search display, which is apparently used to judge each fixated object, producing a process that requires a relatively lengthier fixation compared to an A/V-concurrent trial. The process for an A/V-concurrent condition would be speeder in comparison because objects that match the first identified feature (color) would already be isolated, delegating only one feature (orientation) to judge the remaining subset of objects. This search assistance, with an A/V-concurrent feature presentation, appears to persist with the target-absent trials as evident by the smaller number of fixations and shorter fixation duration.

Fixation Duration for Target-absent Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 351.2	*SD* = 406.3	*M* = 398.1	*SD* = 529.2
10	*M* = 313.9	*SD* = 384.3	*M* = 374.0	*SD* = 522.1
15	*M* = 305.3	*SD* = 402.3	*M* =360.8	*SD* = 501.4
20	*M* = 288.2	*SD* = 366.0	*M* = 339.4	*SD* = 483.5

Table 5.6: Mean and standard deviation values of fixation durations, measured in milliseconds, of target-absent trials for the A/V-concurrent and auditory-first condition across all four set sizes.

The saccade amplitude effect is also highly significant here with the target-absent trials revealing average saccade amplitudes are larger for A/V-concurrent trials ($M = 5.24$, $SD = 6.88$), than for auditory-first trials ($M = 4.73$, $SD = 7.03$). This is also true across all four set sizes, $f(64) = 14.19$, $p < .001$ (see Table 5.7). As with target-present trials, no other significant patterns emerged from the remaining analysis. Although RT-by-set-size functions slope coefficients are smaller, indicating a more efficient search strategy, for target-present trials compared to target-absent trials for both the A/V-concurrent (slope: 5.5 vs. 29.4 ms/item, respectively) and auditory-first (slope: 8.7 vs. 45.2 ms/item, respectively) conditions, the eye-tracking data finds that A/V-concurrent delivery of target features continues to produce more efficient search strategies than auditory-first feature delivery regardless of whether the target is present or not.

Saccade Amplitude for Target-absent Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 5.33	*SD* = 7.00	*M* = 5.22	*SD* = 8.07
10	*M* = 4.45	*SD* = 6.14	*M* = 4.38	*SD* = 6.55
15	*M* = 5.32	*SD* = 6.91	*M* = 4.43	*SD* = 6.50
20	*M* = 5.84	*SD* = 7.36	*M* = 4.96	*SD* = 7.01

Table 5.7: Mean and standard deviation values of saccade amplitude, measured degrees of visual angle, of target-absent trials for the A/V-concurrent and auditory-first condition across all four set sizes.

The findings here are consistent with the inferences made in the previous chapters (1-4) as well as in prior linguistically mediated visual search studies (Spivey et al., 2001; Reali et al., 2006; Chiu & Spivey, 2012). The eye-tracking data found search strategies differ drastically between A/V-concurrent and auditory-first trials. The significantly fewer fixations, shorter fixation durations, and larger saccade amplitudes observed when auditory linguistic target features are delivered concurrent with display onset, in the A/V-concurrent condition, compared to when target features were delivered prior to display onset, in the auditory-first condition, provides further evidence supporting the notion that observers employ distinctive search strategies when display onset timing is altered in relation to feature identity delivery. Furthermore, the longer dwell time observed with color-matched objects than non-matched distractors in A/V-concurrent trials provides further evidence of an initial more efficient parallel process that does not occur in auditory-first trials.

The novel discoveries here further promote the assumption that upon hearing the first-mentioned adjective in a spoken query, visual attention is able to begin the search with only that single feature. Whereby the process is initiated with a highly efficient single-feature search such that when the second adjective is delivered, several hundred milliseconds later, the target can be quickly found among the attended subset of objects. Conversely, trials presented in the auditory-first condition appear to exhibit a search strategy representative of a traditional series search processes, by which each object in the search display is compared to the aforementioned target-object one at a time until the target-object is located in a target-present trial. This study provides us with significant insight into the mechanisms of auditory language mediated visual search and continues to provide strong evidence supporting a large body of research that finds a dynamic and immediate integration of auditory information with visual information.

CHAPTER SIX

Discussion

Summary of Results

Chapter 1 began by presenting the traditional description of visual search that posits two separate cognitive mechanisms, serial and parallel processing, for visual search (Treisman & Gormican, 1988). This was followed by more recent evidence that illustrates a more complex picture of visual search mechanisms, which taken as a whole, could in principle be seen as consistent with the differential occurrence of a parallel process, a serial process, and then a combination of the two (Maioli et al., 2001, Olds et al., 2000a, 2000b, 2000c). However, rather than a hybrid model that proposes the merging of two separate cognitive mechanisms for search processing, the research here continues to propose a single purely parallel mechanism for processing real-time auditory linguistic input during a conjunction search task (Spivey et al., 2001; Reali et al., 2005; Chiu & Spivey, 2012).

Chapters 2 through 5 substantiate this proposal. Chapter 2 reveals, in an assortment of experimental conditions (Experiments 1 & 2), that because of the incremental nature of spoken language comprehension, observers in the A/V-concurrent condition from Experiment 1 (along with observers in the 400 ms and 600-ms SOA semi-concurrent conditions from Experiment 2) can selectively attend to, for some partial degree, the subset of objects that exhibit the target feature that is mentioned first in the speech stream. A possible explanation could be that as linguistic information is processed continuously, with the visual display concurrently available,

search processes are able to partially enhance the salience of the group of items sharing the feature first mentioned and partially suppress the salience of the other now-irrelevant objects. This is indeed what happens in the localist attractor model simulation that is implemented and detailed in Chapter 2.

Chapter 3 demonstrated with various conditions (Experiments 3A & 3B) that search efficiency does not increase with simultaneous delivery of target features in a conjunction-search task despite relatively lengthy previews of search display, 1500 ms in some conditions. Chapter 3 also introduced a novel unimodal visual search paradigm that is purely visual in both search task and delivery of target identity features. Chapter 4 presents a series of conditions that explore the effects of incremental non-linguistic information delivery on visual search (Experiment 4A, 4B, & 4C) by expanding the unimodal paradigm introduced in Chapter 3 to visually simulate the incrementality of an auditory linguistic query. Experiment 4A discovered search efficiency is improved when visual non-linguistic delivery of target features is concurrent with search display onset, but not when the target features are delivered prior to display onset. This indicates that the improvement in search efficiency is not specific to linguistic delivery of target identity, but instead is due to the incrementality of informing the observer of one target feature before informing them of the second target feature. Notably, when the rate of this non-linguistic information delivery was increased (300 vs. 500-ms for color and 600 vs. 1000-ms for orientation) -- in the same way speech rate was increased (3.0 vs. 4.8 syllables/second) in the study by Gibson et al. (2005) for their auditory linguistic queries -- Experiment 4B revealed that previously observed improvements in search efficiency, when target features are delivered incrementally and concurrently with search display onset, were no longer present. Thus, the improvement in search efficiency requires some time to process the first target feature before

delivery of the second target feature. Lastly, in Experiment 4C when the order of feature presentation was reversed using the original delivery rate for target identification (500 ms for color and 1000 ms for orientation), it became apparent that when orientation was delivered first, concurrent delivery of incremental target-identifying cues with display onset also did not produce improvements in search strategies as observed in Experiment 4A. Therefore, the greater efficacy of color-then-orientation delivery, which was previously observed with linguistic cues, is also observed with visual cues.

Chapter 5 detailed an eye-tracking experiment (Experiment 5) that allows us, for the first time, the ability to observe real-time eye-movement and -fixation patterns during a linguistically mediated visual search. Experiment 5, first replicated the language mediated visual search findings initially designed and detailed by Spivey et al (2001). Then, secondly the eye-tracking results provided evidence that supports the claim that search strategies do indeed differ depending on when a search display is made available in relation to target identification. The dense-sampling data indicates that observers utilize a strategy akin to a traditional serial search scheme when target-identity is delivered prior to search display onset. However, when target identity is presented incrementally via spoken language and concurrently with the onset of the search display, observers utilize a more efficient approach comparable to a parallel search process, which has been observed to be greatly less affected by the number of distractor objects (Spivey et al., 2001; Reali et al., 2006; Chiu & Spivey, 2012).

An analysis of the eye-tracking data finds support that a purely parallel mechanism, central to the localist attractor network introduced in Chapter 2 that accurately simulated RT-by-set-size functions, underlies the visual search process mediated by language for the A/V-concurrent and auditory-first condition. Similarly, when Stephen and Mirman (2010) analyzed

the overall distributions of saccade lengths over the course of many trials in a visual search task, they found evidence of a single underlying process for both single-feature and conjunction search (e.g., Spivey & Dale, 2004). Evidence of a lognormal and power-law distribution was also found that implies a self-organized interaction-dominant dynamics in visual cognition (Aks, Zelinsky, & Sprott, 2002) rather than additive encapsulated components (Cavanagh, 1988). Thus, by treating the series of cognitive events as a single continuous process and statistically analyzing as so, the patterns reveal properties of the phenomenon that are not well accommodated by traditional linear box-and-arrow accounts of cognition (Spivey, 2007).

The importance of successfully modeling findings, such as the ones detailed here (Experiment 1-5), is vital for cognitive science and attempting to fulfill the eternal pursuit to comprehend human cognition. The generalizability of a model to a wide range of paradigms and ensuing discoveries will genuinely advance our understanding of the relationship between perceptual systems, as with vision and audition, in human cognition as well as the interaction between perceptual processing and motor action. A key to understanding the foundation of crossmodal interactions lies in examining the evolution of its models, such as with the simulation of audio-visual interactions like language mediated visual search, because as models evolve and improve in generalizability and accuracy, certain theoretical perspectives may effectively be ruled improbable. The next section describes the progression from which a modular completely feed-forward process, that has difficulty modeling the rapid and immediate interaction between the visual system and auditory system, develops into a network that employs continuous interactions between layers of processing to achieve the effects observed in complex real world relationships.

Models of Crossmodal Interaction

Traditional perspectives of cognition were characterized by completely unimodal, and autonomous, modules processing information in a purely feed-forward fashion. Accordingly, the first types of artificial neural networks invented reflected this perspective and were simply feed-forward. In this type of network, information moves in only one direction, from input to output. Moving from the input nodes, through the hidden nodes, and finally to the output nodes the information does not cycle or loop in this type of network. An example of an early purely feed-forward artificial neural network comes from work by Massaro (1999) that utilized a purely feed-forward network to simulate audio-visual speech perception findings. The model described a graded and immediate effect of visual speech perception for when observers judged speakers' mouths for one of two phonemes, "ba" or "da," as faces and sound files were altered digitally along a "ba-da" continuum. In this model, three layers (input, hidden, and output) trained with a simple back-propagation algorithm used weights that were calculated by multiplying the error of the network with the delta rule, a machine learning rule based on convergence of a function that is use to updating the weight of the artificial neurons in a layer of nodes, to minimize overall error. Despite the limitations, this model was described to be "fairly good" at learning the task and generalizing to other similar conditions (Massaro, 1999).

Unfortunately, these early networks do not account well for nonlinear relationships and were unable to simulate the immediate interactions common in many complex real world interactions that dense sample methods, such as eye-tracking and reach-tracking, had more recently revealed. This is not to say that purely feed-forward multi-layered neural networks trained with backpropagation are unable to model nonlinear interactions. Rather these networks have been found to be a valuable tool for modeling and forecasting nonlinear time series where

linear and nonlinear matrix regression methods were insufficient, although large samples were necessary to constrain these models during learning (Blank & Brown, 1992; Zhang, Patuwo, & Hu, 2001). Nevertheless, the latest dense-sampling results suggest a very different account of cognition, inexplicable by traditional feed-forward networks, that incorporates both feed-forward and feedback projections of information alongside constant interaction between different stages of processing. Appropriately, artificial neural network processes evolved to utilize continuous competition and feedback projections in order to accommodate the rapid and immediate processing observed with dense-sampling approaches. In these more physically and computationally plausible models the data are fed forward but also permitted to feedback allowing it to more easily behave in a nonlinear fashion (Spivey, 2007), which is more representative of observed perception and action loops (Hommel, 2004).

The TRACE model of speech perception by McClelland and Elman (1986), detailed in Chapter 1 (p. 19), is an example of a step towards this more plausible simulation of cognition. Another successful example of modeling real-time phenomenon comes from the previously mentioned work on linguistically mediated visual search (Spivey et al., 2001; Spivey & Dale, 2004; Reali et al., 2006; Chiu & Spivey, 2012). Spivey and Dale (2004) and Reali and colleagues (2006) implemented a simple localist attractor network model that was introduced in Chapter 2 (p. 39). This model, inspired by Desimone and Duncan's (1995) biased competition framework, effectively provided the structure for the localist attractor network that was used to simulate results from Experiment 1 and forecast outcomes for Experiment 2 specified in Chapter 2 (p.38).

The next step in modeling multimodal interactions, such as language-mediated vision, appears to be with research in human-robot interactions (HRI) (Roy, 2002; 2005; Cantrell,

Krause, Scheutz, Zillich, & Potapova, 2012; Krause, Cantrell, Potapova, Zillich, & Scheutz, 2013). When interacting with another person, speakers expect listeners to rapidly and immediately integrate perceptual information (Clark & Marshall, 1981). Therefore achieving HRI akin to ordinary human-human interaction (HHI) is a challenging but noble objective because a robot model with the capacity to rapidly and immediately integrate perceptual information would need to exhibit some demanding characteristics. For instance, we expect robots to, as humans do, incrementally process spoken references to visually perceivable objects in an environment as the referents are verbally described. In order to implement efficient and naturalistic HRI one must pose tight timing requirements on visual as well as language processing. Cantrell et al. (2012) and Krause et al. (2013) did just this with a model that uses an integrated robotic architecture capable of integrating novel visual input incrementally by using natural language processing to refine attentional focus. Consistent with the human data detailed here (Experiment 1-5), Cantrell et al. (2012) and Krause et al. (2013) have found significantly better performance of robot vision systems when using incremental linguistic constraints than when using a non-incremental visual processing approach. The field of HRI is emerging to be the future of modeling cognitive processes, especially for perceptual integration and crossmodal interactions, because of the requirement for human-like behavior (Cantrell et al., 2012).

Multisensory Integration and Crossmodal Interaction

Initially compelled by Bishop Berkeley's theory that visual perception of space is acquired on the basis of tactile experience, early research on multisensory perception demonstrated vision's near total dominance over proprioception (Hay, Pick, & Ikeda, 1965). However, recent research has illustrated a more complex picture revealing that visual bias of

proprioception only accounts for part of the discrepancy with the inverse, proprioceptual bias of vision, occurring more frequently than originally believed (Spence & Driver, 1996). As a result the complexity of multisensory integration and crossmodal interactions was made apparent and the importance of exploring its mechanisms emerged as a vital aspect of mapping out cognition in the mind.

Attention often precedes action and is, thus, an essential piece of multimodal research. As discussed in Chapter 1, the study of endogenous spatial attentional subsystems is populated by three main potential architectures: a completely supramodal perspective, a modality-specific perspective, and a hybrid merger of the two theories. In a review of current research Spence (2010) draws what he describes as the most "parsimonious" conclusion imaginable, concluding that exogenous spatial cuing effects are indeed supramodal. Meaning as long as auditory, visual, and tactile cues are presented in the same temporal (300 ms or less SOA) and spatial location they will all give rise to a shift of spatial attention that facilitates observers' responses to auditory, visual, and tactile targets. Findings that have identified spatial and temporal distance between targets and cues as a critical determinant of whether or not a crossmodal spatial cuing effect will be observed support Spence's (2010) conclusion (Prime, McDonald, Green, & Ward, 2008). Although further research is necessary to validate Spence's (2010) conclusion, the visual search data observed in experiments 1 – 5, especially the eye-tracking data from Experiment 5, exhibit a strong intermodal that is sensitive to relatively minute changes in SOA of both auditory linguistic processing and visual information processing with visual attention, which is consistent with a supramodal account of exogenous spatial attention.

General Discussion

Humans are inherently limited-capacity creatures and as a result crossmodal capabilities bestow considerable behavioral advantages. More than just having the capacity to use sensory information interchangeably or being able to recognize objects when deprived of a sense it is the capacity to combine sensory inputs across modalities that is the true wonder, because by doing so one can considerably enhance the discrimination and detection of stimuli as well as dramatically speed up response time (Spence & Ho, 2008). Although redundant information can be beneficial (Selcon, Taylor, & McKenna, 1995) input from the different sensory modalities are typically complementary, thus crossmodal integration of multiple sensory inputs more often that not provides a percept of the environment or event that is unobtainable from any single sense alone. For instance, we have all experienced food tasting bland when our noses are stuffy from a cold, which effectively blocks olfactory input. The sensation of blandness occurs because our experience of taste is derived from the integration of information from both the gustatory and olfactory system and without olfactory stimulation, our percept of the event is not as robust. The experience of taste without smell is actually quite dull because the sum of the multisensory interaction exceeds the sum of the parts.

The brain's capacity to converge these multisensory cues is imperative for guiding us through our environment and directing attention. This single percept of the world constructed from multisensory cues can also improve object detection, localization, identification, as well as improve event response speed and accuracy (Welch & Warren, 1986; Stein & Meredith, 1993). Given the ubiquitous and indispensable nature of crossmodal processing for human experience, knowledge of the underlying neurophysiology seems key to our understanding of human brain function. We must consider the continuous feed-forward and feedback process of crossmodal

interactions when designing and engineering HRIs. As well as remain ever-conscious of what produces both the strongest activation of salient targets and suppression of distractors when designing other circumstances of human technology interaction such as with various alerting signals (e.g., alarm clocks, warning indicators in cars, etc.), computer and cell phone controls, and even video game interfaces; being careful to take advantage of the human cognitive ability to immediately and rapidly integrate sensory information.

There has been a rapid growth of interest in the application of laboratory-based studies of crossmodal attention to improve real-world interface design, focusing particularly on the design of multisensory warning signals for automobile and their operators (Spence & Ho, 2008). The need to identify the best modality or combination of modalities that produce the most intuitive, non-redundant, and least annoying mode for information delivery is imperative to both this line of research and also the general understanding of the human cognition. Unfortunately, there is a lot more that needs to be done to generalize this line of research to real-world events as supposed to ersatz laboratory scenes, but the progress made so far has been enlightening and promising.

Conclusion

In conclusion, the findings detailed in Experiments 1 – 5 suggest that with concurrent delivery of target-identity it is the incremental nature of target-delivery, whether via speech comprehension or visual processing, that allows the visual search process to begin when only a single feature of the target-identity has been revealed. When the initial feature is identified, the search proceeds in an efficient parallel-like fashion. This process increases saliency of matching targets and suppresses distractors such that when the second target-feature is presented, a substantial amount of the target identification process has already been completed. As a result, the presence of

multiple distractors is less disruptive, effectively producing a more efficient search strategy. Furthermore, simulations of data in Experiment 1 and 2 as well as dense-sampling analyses support the proposal that the process associated with language mediated visual search may be purely parallel in nature. These results add to a mounting collection of evidence that demonstrates a dynamic and robustly interactive account of language comprehension and visual attention (Spivey, 2007). Although the findings in this examination add to our understanding of the bond between two modalities, it also serves to add to the complexity of the relationship. Thus, investigating the intricacies of how language comprehension and visual attention interact in real-time is important because it will not only benefit our understanding of both multisensory integration and the interaction between modalities, but it will in turn add to our general knowledge of human cognition as a whole.

References

Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science*, *2*(3), 153-157.

Aks, D. J., Zelinsky, G. J., & Sprott, J. C. (2002). Memory across eye-movements: 1/f dynamic in visual search. *Nonlinear dynamics, psychology, and life sciences*, *6*(1), 1-25.

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

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, *38*(4), 419-439.

Anderson, S. E., Chiu, E. M., Huette, S., and Spivey, M. L. (2010). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica, 137*(2), 181-189.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(04), 723-742.

Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1289.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617-645.

Blank, T. B., & Brown, S. D. (1993). Nonlinear multivariate mapping of chemical data using feed-forward neural networks. *Analytical chemistry*, *65*(21), 3081-3089.

Boucart, M., & Humphreys, G. W. (1997). Integration of physical and semantic information in object processing. *PERCEPTION-LONDON-*, *26*, 1197-1209.

Boulenger, V., Décoppet, N., Roy, A. C., Paulignan, Y., & Nazir, T. A. (2007). Differential effects of age-of-acquisition for concrete nouns and action verbs: Evidence for partly distinct representations?. *Cognition*, *103*(1), 131-146.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application electrophysiological criteria to the BOLD effect. *NeuroImage, 14*, 427-438.

Cantrell, R., Krause, E., Scheutz, M., Zillich, M., & Potapova, E. (2012, July). Incremental Referent Grounding with NLP-Biased Visual Search. In*Proceedings of AAAI 2012 Workshop on Grounding Language for Physical Systems*.

Cavanagh, P. R. (1988). *U.S. Patent No. 4,771,394*. Washington, DC: U.S. Patent and Trademark Office.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge: MIT press.

Chiu, E. M. & Spivey, M. J. (2012). The role of preview and incremental delivery on visual search. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 216-221). Austin, TX: Cognitive Science Society.

Chun, M. M. and Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology, 30*, 39-78.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, *45*(5), 801-814.

Clark, H., and Marshall, C. (1981). Definite reference and mutual knowledge. In Joshi, A. K.; Webber, B. L.; and Sag, I. A., eds., *Elements of discourse understanding*. Cambridge:

Cambridge University Press. 10–63.

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15-28.

de Sa, V. R. & Ballard, D. H. (1998). Category learning through multimodality sensing. *Neural Computation, 10*, 1097-1117.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1245.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

De Valois, R. L., & De Valois, K. K. (1993). A multi-stage color model. *Vision research*, *33*(8), 1053-1065.

Driver, J., & Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current Biology*, *10*(20), R731-R735.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological review*, *87*(3), 272.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological review*, *96*(3), 433.

Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attention processing, *Psychological Science, 9*, 111-118.

Elazary, L., & Itti, L. (2010). A bayesian model for efficient visual search and recognition. *Vision Research, 50*(14), 1338-1352.

Farah, M. J., Wong, A. B., Monheit, M. A., & Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal?. *Neuropsychologia*, *27*(4), 461-470.

Farmer, T. A., Anderson, S. E., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language*, *57*(4), 570-595.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, Mass.: MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178-210.

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*(4), 673.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory?. *Trends in cognitive sciences*, *10*(6), 278-285.

Gibson, B. S., Eberhard, K. M., & Bryant, T. A. (2005). Linguistically mediated visual search: The critical role of speech rate. *Psychonomic Bulletin and Review*, *12*(2), 276.

Giray, M., & Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(6), 1278.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*(6776), 390-394.

Hay, J. C., Pick, H. L., & Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic Science*.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in cognitive sciences*, *8*(11), 494-500.

Howard, I., and Templeton, W. (1966). *Human spatial orientation*. Oxford: Wiley.

Hubel, D. H. (1988). Eye, brain, and vision (Scientific American Library). *New York*.

Huettig, F., & Altmann, G. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*,*96*(1), B23-B32.

Huettig, F., & Altmann, G. T. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*(8), 985-1018.

Humphreys, G. W. & Riddoch, M. J. (1993). Interactions between object and space systems revealed through neuropsychology. In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV* (pp. 183-218). Cambridge, MA: MIT Press.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, *2*(3), 194-203.

Jones, J. J., Kaschak, M. P., & Boot, W. R. (2011). Language mediated visual search: The role of display preview. *Cognitive Science Proceedings*, 2739-2744.

Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology: General*, *136*(4), 551.

King, A. (1999). Sensory experience and the formation of a computational map of auditory space in the brain. *BioEssays, 21*, 900-911.

Krause, E., Cantrell, R., Potapova, E., Zillich, M., & Scheutz, M. (2013). Incrementally biasing visual search using natural language input. In*Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems* (pp. 31-38). International Foundation for Autonomous Agents and Multiagent Systems.

Lewald, J. (1997). Eye-position effects in directional hearing. *Behavioural brain research*, *87*(1), 35-48.

Lewald, J. (1998). The effect of gaze eccentricity on perceived sound direction and its relation to visual localization. *Hearing research*, *115*(1), 206-216.

Lewald, J., & Ehrenstein, W. H. (1996). The effect of eye position on auditory lateralization. *Experimental brain research*, *108*(3), 473-485.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization3* (Vol. 8). Oxford University Press.

Lupyan, G., & Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology*, *18*(10), r410–412.

Lupyan, G. & Spivey, M.J. (2010). Now you see it, now you don't: Verbal but not visual cues facilitate visual object detection. *PLoS ONE*.

Maioli, C., Benaglio, I., Siri, S., Sosta, K., & Cappa, S. (2001). The integration of parallel and serial processing mechanisms in visual search: Evidence from eye movement recording. *European Journal of Neuroscience. 13*(2), 364-372.

Marslen-Wilson, M. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71-102.

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

Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual review of neuroscience*, *10*(1), 363-401.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1-86.

McCluskey, A., & Lalkhen, A. G. (2007). Statistics II: Central tendency and spread of data. *Continuing Education in Anaesthesia, Critical Care & Pain*, *7*(4), 127-130.

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

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22-24.

Nakayama, K., & Joseph, J. S. (1998). Attention, pattern recognition, and pop-out in visual search. *The attentive brain*, 279–298.

Olds, E. S., Cowan, W. B., & Jolicoeur, P. (2000a). Partial orientation pop-out helps difficult search for orientation. *Perception & psychophysics*, *62*(7), 1341–1347.

Olds, E. S., Cowan, W. B., & Jolicoeur, P. (2000b). The time-course of pop-out search. *Vision Research*, *40*(8), 891–912.

Olds, E. S., Cowan, W. B., & Jolicoeur, P. (2000c). Tracking visual search over space and time. *Psychonomic Bulletin and Review*, *7*(2), 292–300.

Olds, E. S., & Fockler, K. A. (2004). Does previewing one stimulus feature help conjunction search?. *PERCEPTION-LONDON-*, *33*(2), 195-216.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision research*, *42*(1), 107-123.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in cognitive sciences*, *1*(2), 56-61.

Posner, M. I. (1978). *Chronometric explorations of mind*. Lawrence Erlbaum.

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

Posner, R. A. (1990). *The problems of jurisprudence*. Harvard University Press.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and performance X: Control of language processes*, *32*, 531-556.

Posner, M. I., Walker J. A., Friedrich, F. J., & Rafal R. D. (1984). Effects of parietal ingjury on covert orienting of attention. *Journal of Neuroscience, 4*, 1863-1974.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: General*, *109*(2), 160-174.

Prime, D.J., McDonald, J.J., Green, J., and Ward, L.M. (2008). When crossmodal attention fails: a controversy resolved? *Can. Journal of Experimental Psychology, 62*, 192–197.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576-582.

Reali, F., Spivey, M. J., Tyler, M. J., & Terranova, J. (2006). Inefficient conjunction search made efficient by concurrent spoken delivery of target identity. *Perception and Psychophysics*, *68*(6), 959.

Reynolds, J., & Desimone, R. (2001). Neural mechanisms of attentional selection. *Visual attention and cortical circuits (Braun J, Koch C, Davis JL, eds)*, 121–136.

Roy, D. K. (2002). Learning visually grounded words and syntax for a scene description task. *Computer Speech & Language*, *16*(3), 353-385.

Roy, D. (2005). Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence*, *167*(1), 170-205.

Selcon, S.J., Taylor, R.M., and McKenna, F.P. (1995). Integrating multiple information sources: using redundancy in the design of warnings, *Ergonomics, 38*, 2362-2370.

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

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current opinion in neurobiology*, *11*(4), 505-509.

Soto-Faraco, S., Foxe, J. J., & Wallace, M. T. (2005). Multisensory processes. *Experimental Brain Research*, 166, 287-288.

Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, *1191*(1), 182-200.

Spence, C. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(3), 555.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 1005.

Spence, C., & Driver, J. (Eds.). (2004). *Crossmodal space and crossmodal attention*. Oxford University Press.

Spence, C. and Ho, C. (2008). Multisensory warning signals for event perception and safe driving. *Theoretical Issues in Ergonomics Science, 9*(6), 523-554.

Spence, C., Nicholls, M. E., Gillespie, N., & Driver, J. (1998). Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. *Perception & Psychophysics*, *60*(4), 544-557.

Spivey, M. (2007). *The continuity of mind*. New York: Oxford University Press.

Spivey, M. J., & Dale, R. (2004). On the continuity of mind: Toward a dynamical account of cognition. In B. Ross (Ed.), *The psychology of learning and motivation*. *Volume 45* (pp. 87-142). San Diego: Elsevier.

Spivey, M. J., Dale, R., Knoblich, G., & Grosjean, M. (2010). Do curved reaching movements emerge from competing perceptions? A reply to van der Wel et al.(2009).

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393-10398.

Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, *12*(4), 282-286.

Spivey-Knowlton, M. J. (1996). *Integration of visual and linguistic information: Human data and model simulations* (Doctoral dissertation, University of Rochester. Department of Brain and Cognitive Sciences).

Spratling, M.W., & Johnson, M.H. (2004). A feedback model of visual attention. *Journal of Cognitive Neuroscience, 16*(2), 219-237.

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

Stephen, D. G., & Mirman, D. (2010). Interactions dominate the dynamics of visual cognition. *Cognition*, *115*(1), 154-165.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science, 268*, 1632-1634.

Tipper, S. P., Howard, L. A., & Jackson, S. R. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition*, *4*(1), 1-38.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research*, *23*(8), 775-785.

Todd, J. W. (1912). *Reaction to multiple stimuli* (No. 25). Science Press.

Treisman, A. (1990). Variations on the theme of feature integration: Reply to Navon (1990).

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

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15–48.

Tyler, C. W., & Hamer, R. D. (1990). Analysis of visual modulation sensitivity. IV. Validity of the Ferry-Porter law. *JOSA A*, *7*(4), 743-758.

Van Der Wel, R. P., Eder, J. R., Mitchel, A. D., Walsh, M. M., & Rosenbaum, D. A. (2009). Trajectories emerging from discrete versus continuous processing models in phonological competitor tasks: A commentary on Spivey, Grosjean, and Knoblich (2005).

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*(3), 331.

Viemeister, N. F., & Plack, C. J. (1993). Time analysis. In *Human psychophysics* (pp. 116-154). Springer New York.

Watson, M. R., Brennan, A. A., Kingstone, A., & Enns, J. T. (2010). Looking versus seeing: Strategies alter eye movements during visual search. *Psychonomic Bulletin & Review, 17*(4), 543-549.

Welch, R., and Warren, D. (1986). Intersensory interactions. In K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of perception and human performance, Vol. 1: Sensory processes and perception* (pp. 1-36). New York: Wiley.

Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science*, *20*(11), 1428-1435.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202-238.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*, 33-39.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1.

Zhang, G. P., Patuwo, B. E., & Hu, M. Y. (2001). A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers & Operations Research*, *28*(4), 381-396.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32*, 25-64.

Tables

Table 5.1

Number of Fixations for Target-present Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 12.1	*SD* = 5.1	*M* = 13.4	*SD* = 7.3
10	*M* = 12.7	*SD* = 4.9	*M* = 14.0	*SD* = 5.7
15	*M* = 14.2	*SD* = 10.0	*M* = 14.5	*SD* = 6.2
20	*M* = 13.2	*SD* = 5.9	*M* = 14.5	*SD* = 5.4

Table 5.1: Number of fixation mean and standard deviation values in target-present trials for the A/V-concurrent and auditory-first condition across all four set sizes.

Table 5.2

Fixation Duration for Target-present Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 361.6	*SD* = 414.4	*M* = 421.6	*SD* = 547.5
10	*M* = 334.6	*SD* = 376.2	*M* = 375.5	*SD* = 488.0
15	*M* = 327.2	*SD* = 397.6	*M* = 368.3	*SD* = 484.9
20	*M* = 320.8	*SD* = 391.2	*M* = 367.9	*SD* = 494.5

Table 5.2: Mean and standard deviation values of fixation durations, measured in milliseconds, for A/V-concurrent and auditory-first trials across all four set sizes.

Table 5.3

Saccade Amplitude for Target-present Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 5.33	*SD* = 7.00	*M* = 5.22	*SD* = 8.06
10	*M* = 4.45	*SD* = 6.14	*M* = 4.38	*SD* = 6.55
15	*M* = 5.32	*SD* = 6.91	*M* = 4.43	*SD* = 6.50
20	*M* = 5.84	*SD* = 7.36	*M* = 4.96	*SD* = 7.01

Table 5.3: Mean and standard deviation of saccade length measured in amplitude (degrees of visual angle) for A/V-concurrent and auditory-first trials across all four set sizes.

Table 5.4

Saccade Velocity for Target-present Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 109.7	*SD* = 154.0	*M* = 116.9	*SD* = 240.7
10	*M* = 102.5	*SD* = 136.6	*M* = 102.5	*SD* = 188.6
15	*M* = 122.4	*SD* = 210.3	*M* = 103.8	*SD* = 169.1
20	*M* = 126.1	*SD* = 186.4	*M* = 111.1	*SD* = 193.9

Table 5.4: Mean and standard deviation of average saccade velocity for A/V-concurrent and auditory-first trials across all four set sizes.

Table 5.5

Number of Fixations for Target-absent Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 12.1	*SD* = 5.1	*M* = 13.4	*SD* = 7.3
10	*M* = 12.7	*SD* = 4.9	*M* = 26.1	*SD* = 44.6
15	*M* = 14.2	*SD* = 10.0	*M* = 14.5	*SD* = 6.2
20	*M* = 13.2	*SD* = 5.9	*M* = 14.5	*SD* = 5.4

Table 5.5: Mean and standard deviation values for the number of fixations in target-absent trials for the A/V-concurrent and auditory-first condition across all four set sizes.

Table 5.6

Fixation Duration for Target-absent Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 361.6	*SD* = 414.4	*M* = 421.6	*SD* = 547.5
10	*M* = 334.6	*SD* = 376.2	*M* = 375.5	*SD* = 488.0
15	*M* = 327.2	*SD* = 397.6	*M* = 368.3	*SD* = 484.9
20	*M* = 320.8	*SD* = 391.2	*M* = 367.9	*SD* = 494.5

Table 5.6: Mean and standard deviation values of fixation durations, measured in milliseconds, of target-absent trials for the A/V-concurrent and auditory-first condition across all four set sizes.

Table 5.7

Saccade Amplitude for Target-absent Trials

Set Size	A/V-concurrent		Auditory-first	
5	*M* = 5.33	*SD* = 7.00	*M* = 5.22	*SD* = 8.07
10	*M* = 4.45	*SD* = 6.14	*M* = 4.38	*SD* = 6.55
15	*M* = 5.32	*SD* = 6.91	*M* = 4.43	*SD* = 6.50
20	*M* = 5.84	*SD* = 7.36	*M* = 4.96	*SD* = 7.01

Table 5.7: Mean and standard deviation values of saccade amplitude, measured degrees of visual angle, of target-absent trials for the A/V-concurrent and auditory-first condition across all four set sizes.

Figures

Figure 1.1

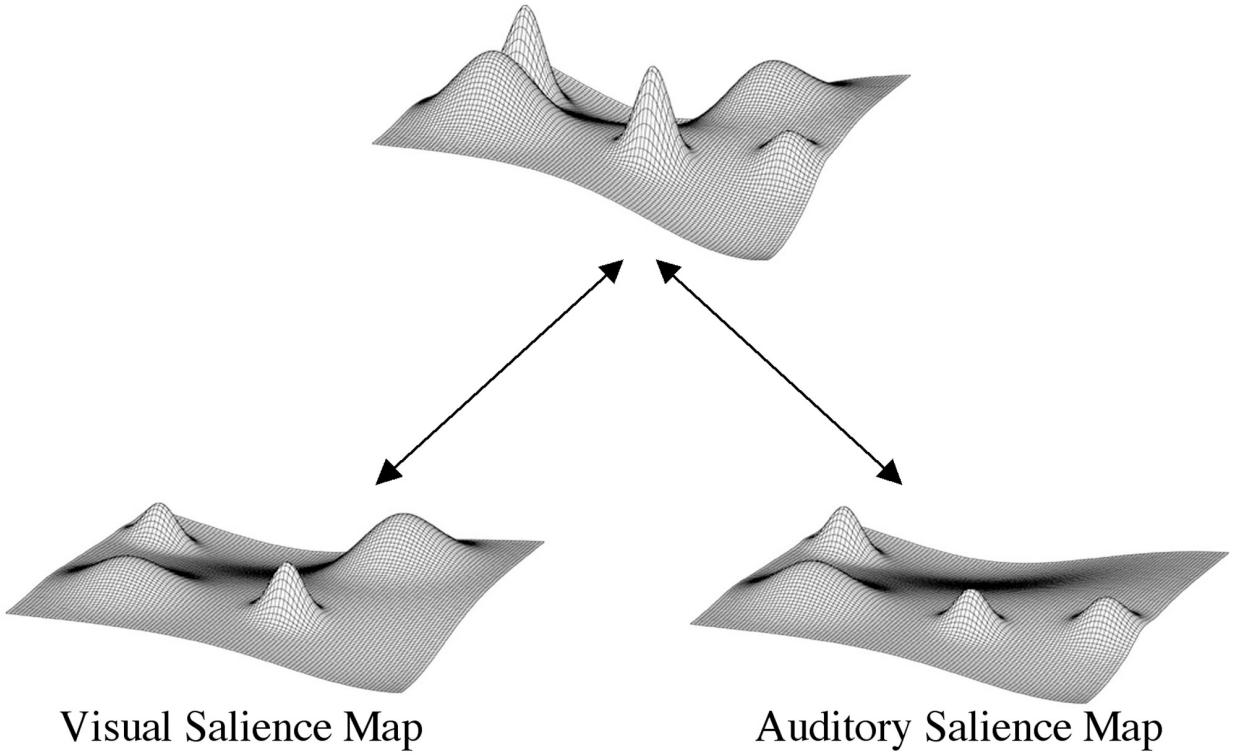

Figure 1.1: Depiction of a supramodal hybrid theory of attentional saliency. A supramodal salience map receives input from and sending feedback to unimodal salience maps. Note that areas with overlapping activation in the unimodal maps would produce a subsequently larger activation in the supramodal map (Figure adapted from Spivey (2007) with permission.)

Figure 2.1

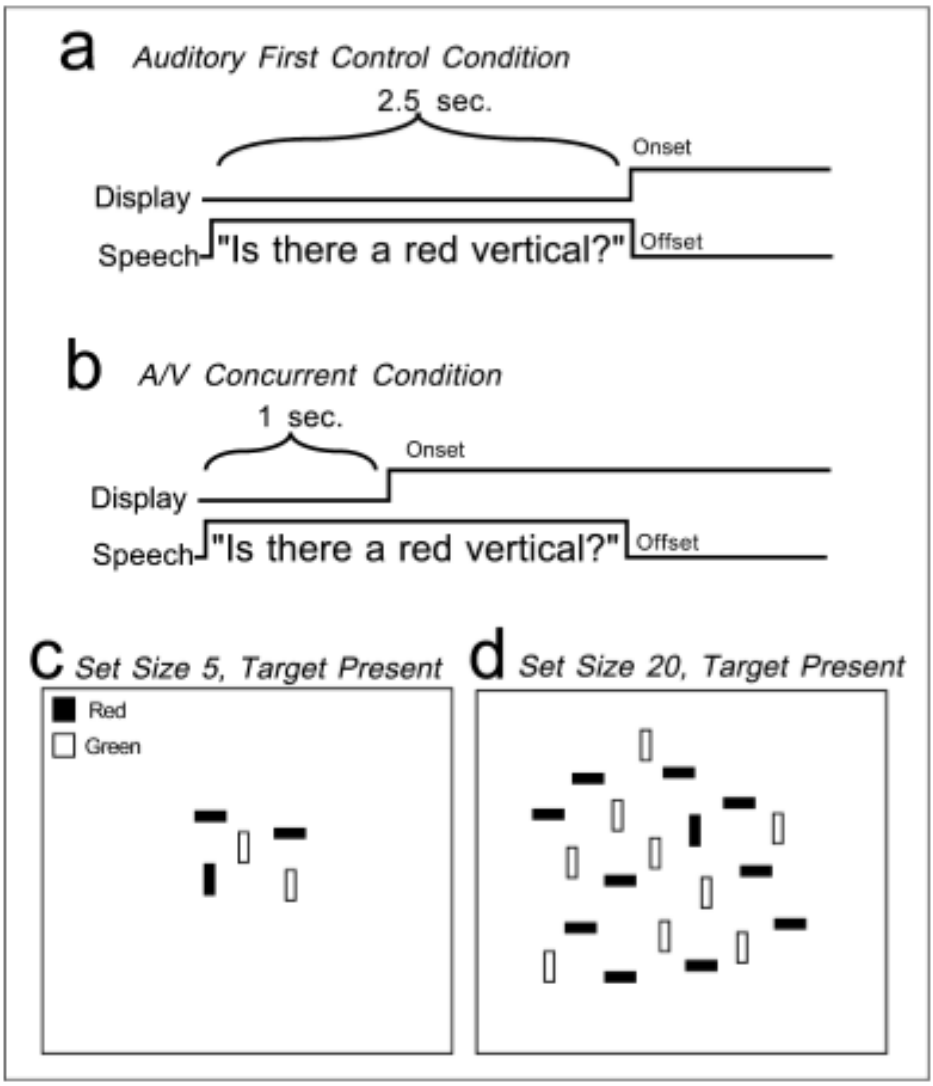

Figure 2.1: Examples of the auditory and visual stimuli. In the auditory-first control condition (a) the onset of the visual display coincided with the offset of the spoken target query. In the audiovisual-concurrent (A/V-concurrent) condition (b), the onset of the visual display coincided with the onset of the first target-feature word in the spoken query. The example displays show target-present trials with a set size of 5 (c) and 10 (d). In these displays, the target is a red vertical bar, which is accompanied by vertical green distractor bars and horizontal red distractor bars. (Figure adapted from Spivey et al. (2001))

Figure 2.2

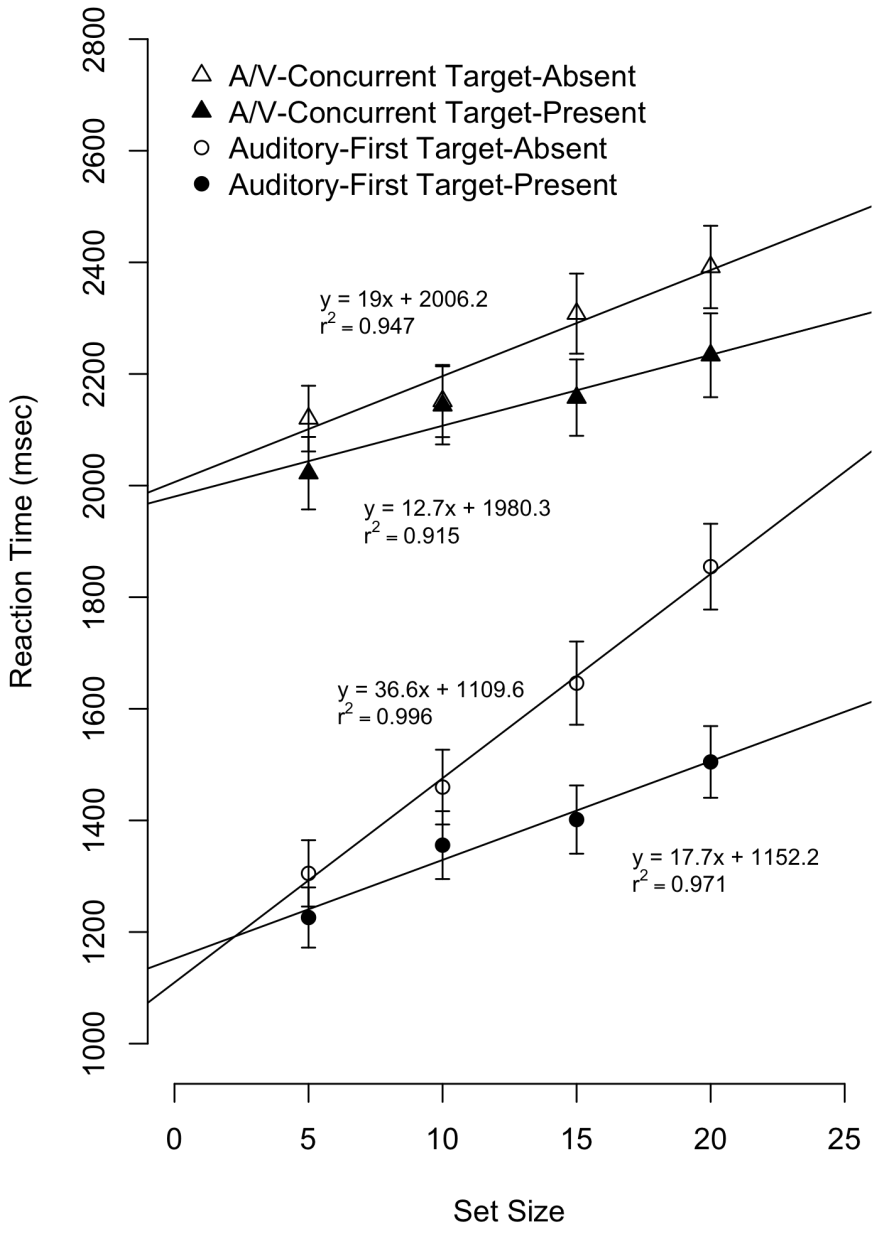

Figure 2.2: Results from Experiment 1. Shown separately for target-present (filled symbols) and target-absent (open symbols) trials for both the auditory-first control (circles) and the A/V-concurrent conditions (triangles). Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

Figure 2.3

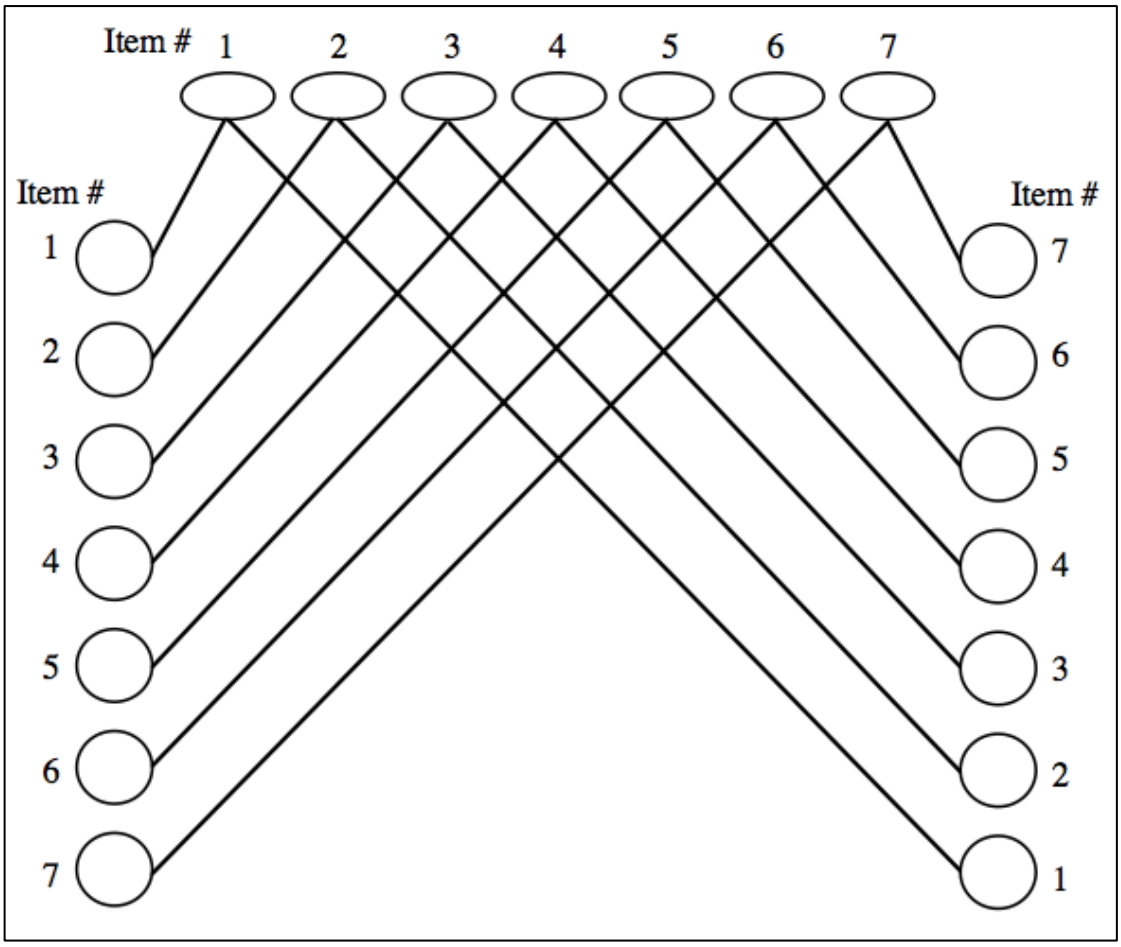

Figure 2.3: Integration-competition model of visual search. A localist attractor network model that simulates a potential mechanism by which the visual search process may be influenced by incremental linguistic input. One feature vector of nodes represented the target property redness (positive activation) and non-redness (zero activation). Another feature vector represented the target property verticalness (positive activation) and non-verticalness (zero activation). Finally, an integration vector (top) receiving input from those feature vectors represented each object's likelihood of being the target. The lengths of these vectors vary depending on set size, 7 in this example.

Figure 2.4

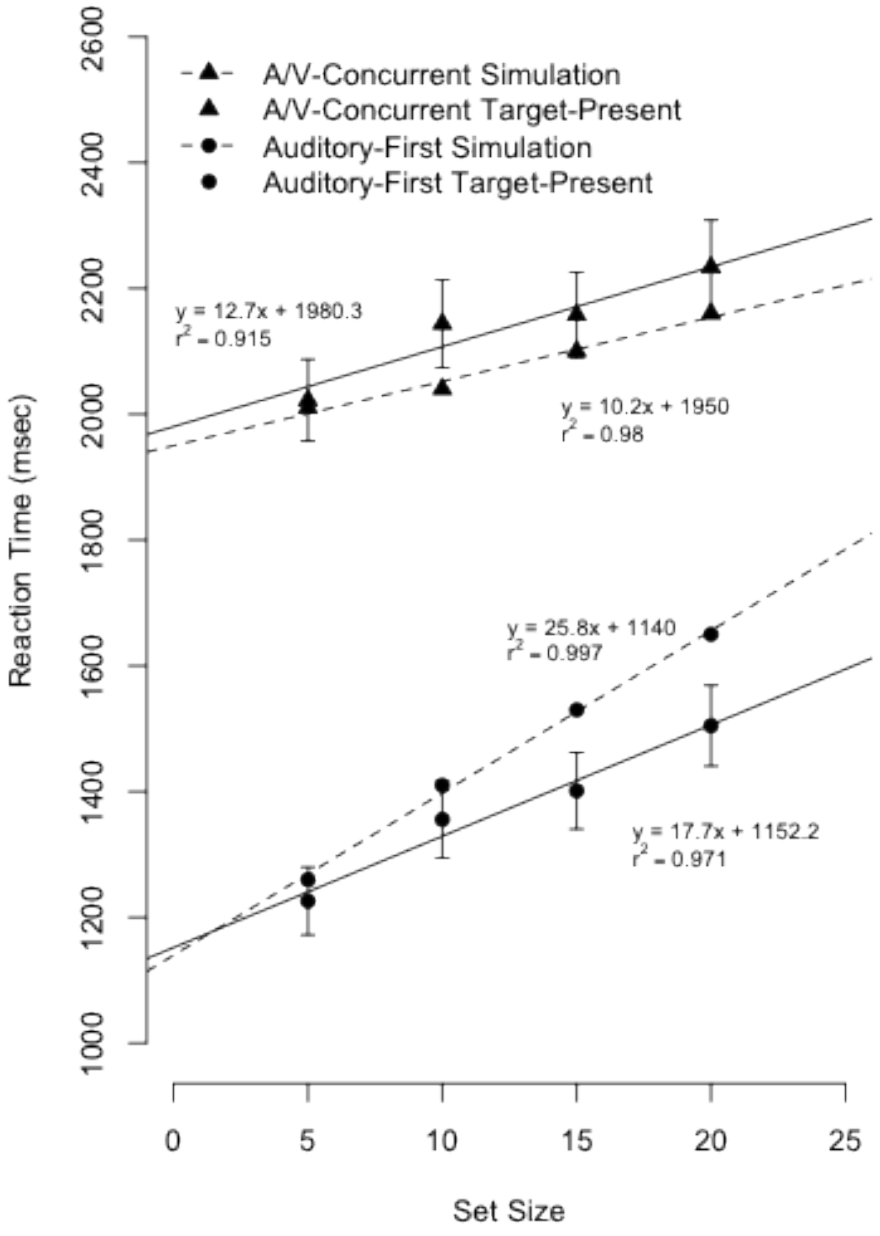

Figure 2.4: Results from the localist attractor network simulation. Dashed lines show the simulation with human data (solid lines) from Experiment 1 for target-present trials. Each line is accompanied by the best-fit linear equation. The results of Experiment 1 are accompanied by the accounted proportion of variance (r^2). The error bars indicate standard error of the mean.

Figure 2.5

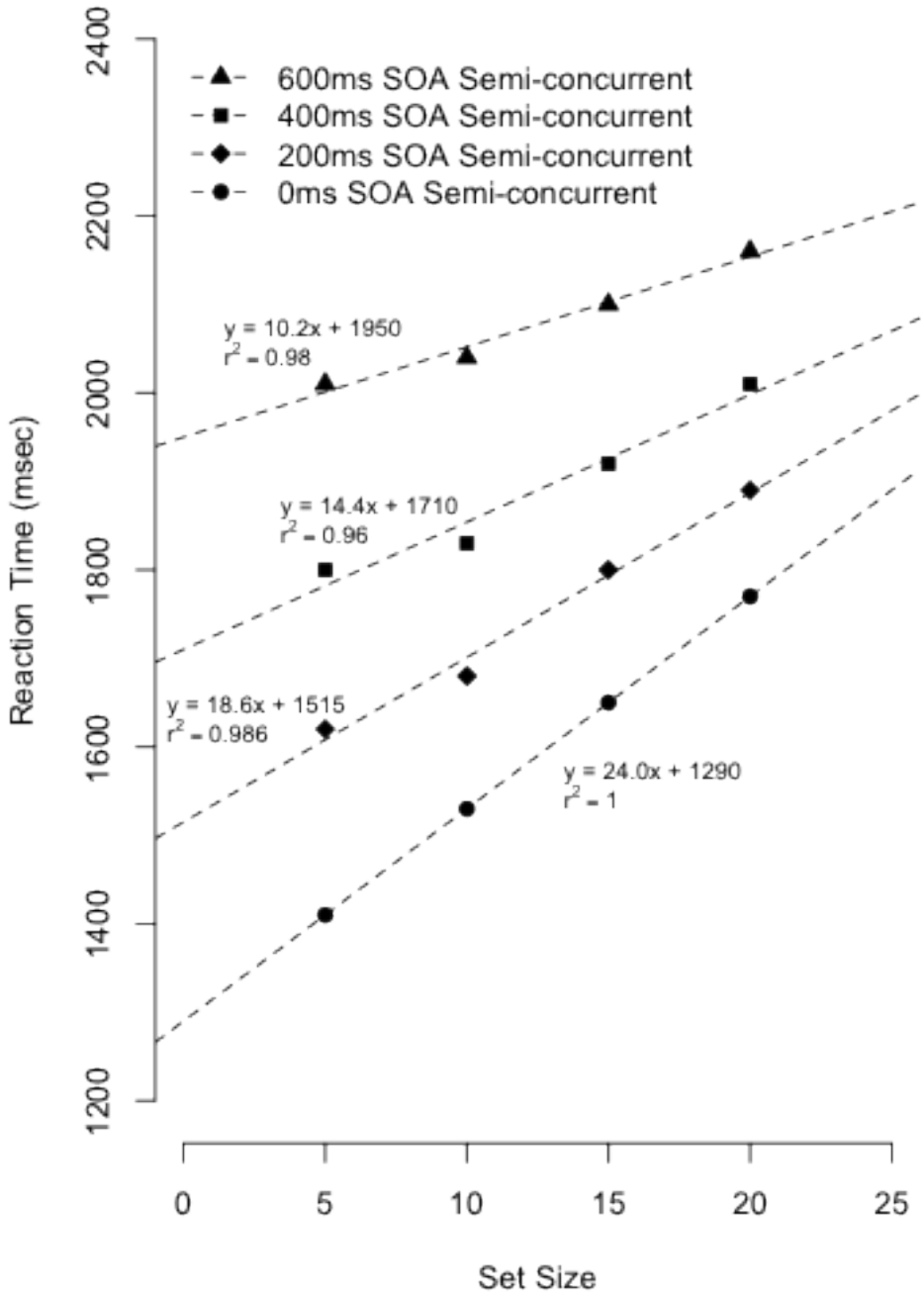

Figure 2.5: Localist attractor network predictions for semi-concurrent conditions. Each line is accompanied by the best-fit linear equation.

Figure 2.6

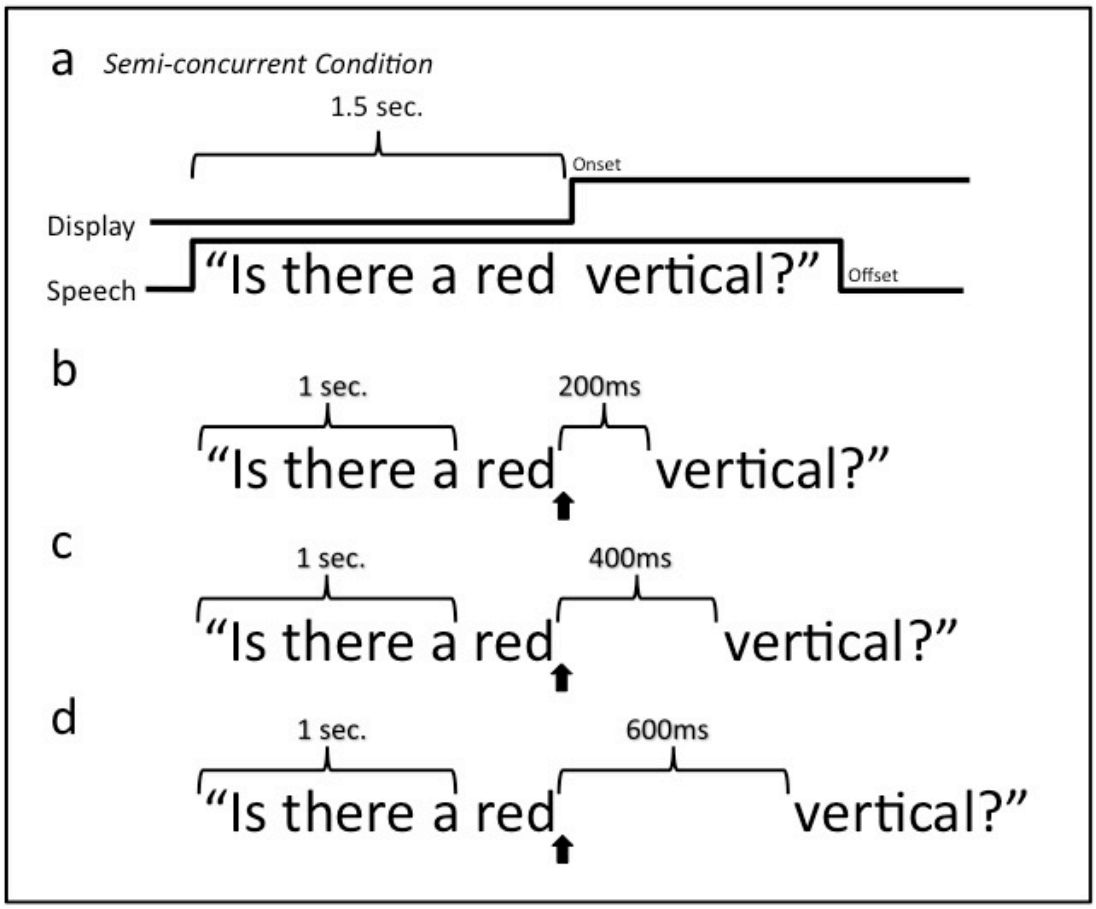

Figure 2.6: Examples of auditory stimuli for semi-concurrent conditions. In the 0-ms SOA semi-concurrent (a) condition, which is similar to the A/V-concurrent condition of Experiment 1, the onset of the visual display coincided with the end of the first descriptive adjective (color). The arrows indicate display onset for the 200-ms SOA semi-concurrent (b), 400-ms SOA semi-concurrent (c), and the 600-ms SOA semi-concurrent (d) conditions.

Figure 2.7

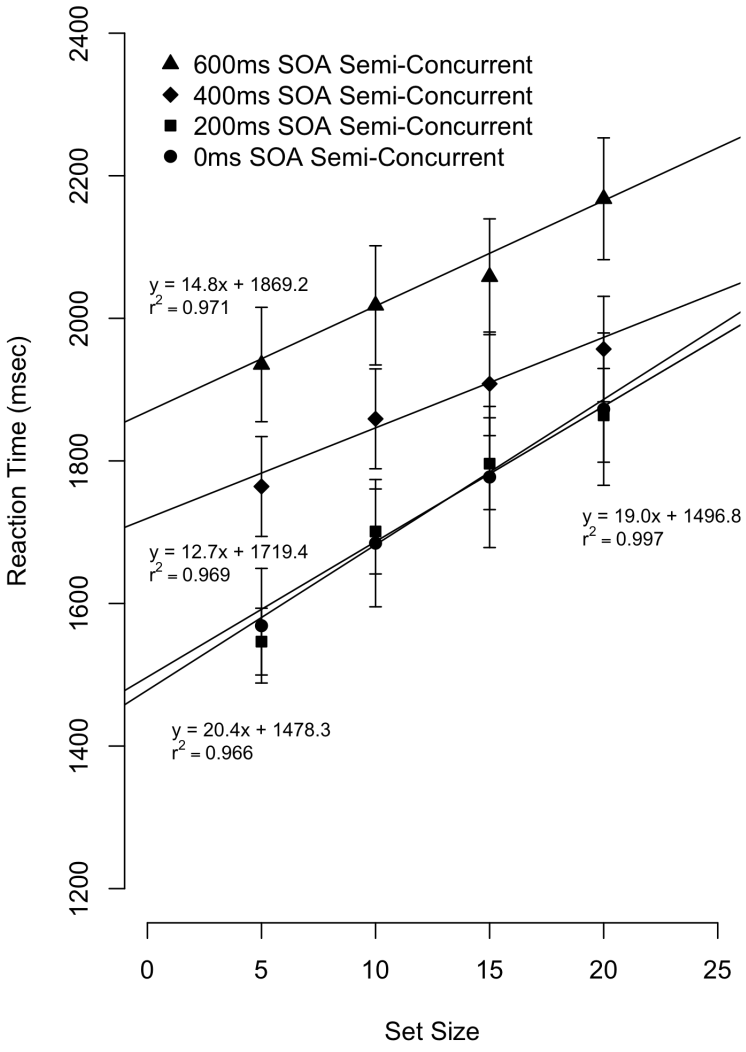

Figure 2.7: Results from Experiment 2. Shown are target-present trials for the 0-ms SOA semi-concurrent conditions (circle), 200-ms SOA semi-concurrent conditions (square), 400-ms SOA semi-concurrent conditions (diamond), 600-ms SOA semi-concurrent conditions (triangle). Each line is accompanied by the best-fit linear equation. The results from Experiment 2 are accompanied by the accounted for proportion of variance (r^2). Error bars indicate standard error of the mean.

Figure 3.1

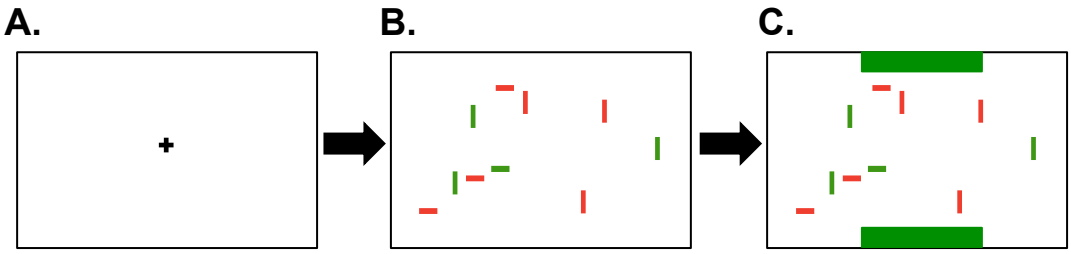

Figure 3.1: Example of nonlinguistic visual cues trial presentation for Experiment 3. Duration of search display (B) varied between 0, 350, & 750 ms in Experiment 3A and 0 & 1500 ms in Experiment 3B before the target identifying visual cues appeared (C).

Figure 3.2

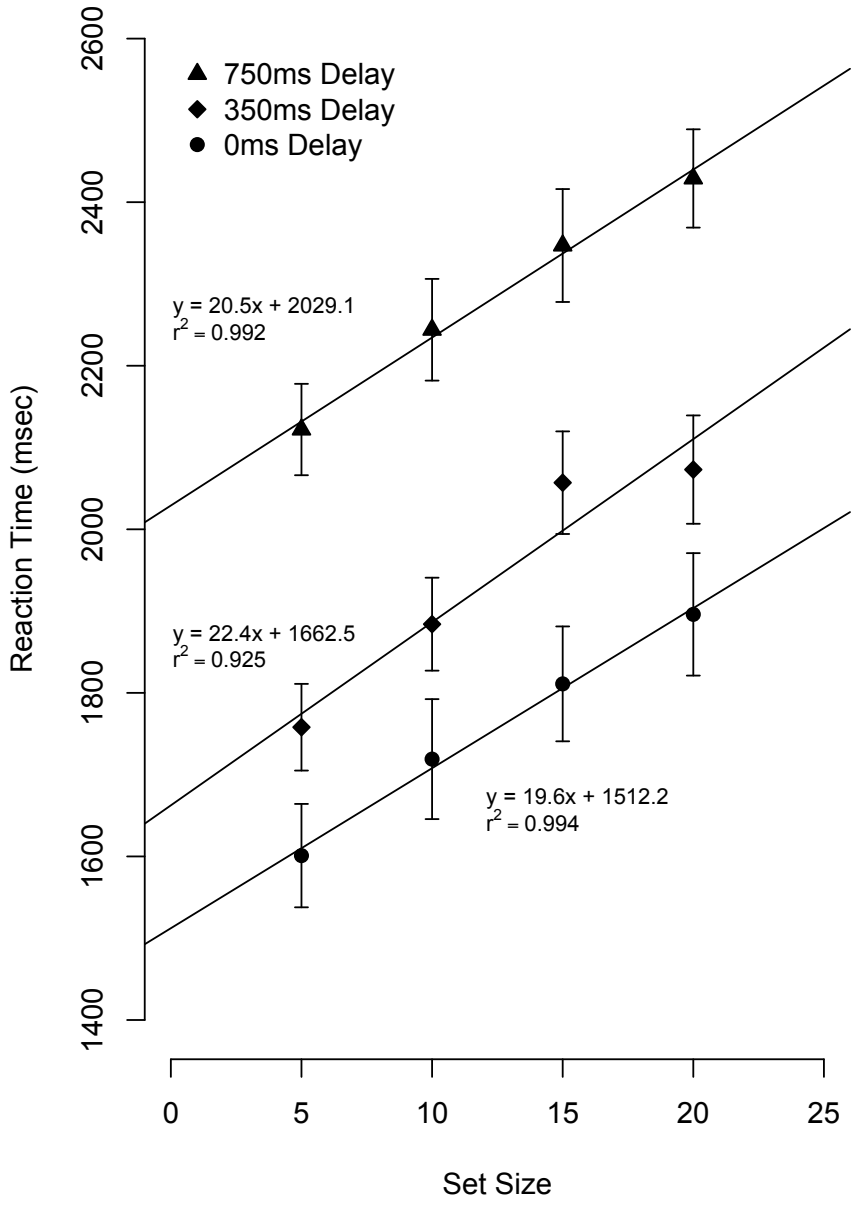

Figure 3.2: Results from Experiment 3A for target-present trials (filled symbols). Shown for the 0 ms delay condition (circle), 350 ms delay condition (diamond), and 750 ms condition (triangle). Each line is accompanied by the best-fit leaner equation and the accounted proportion of variance (r^2). Error bars indicate standard error of the mean.

Figure 3.3

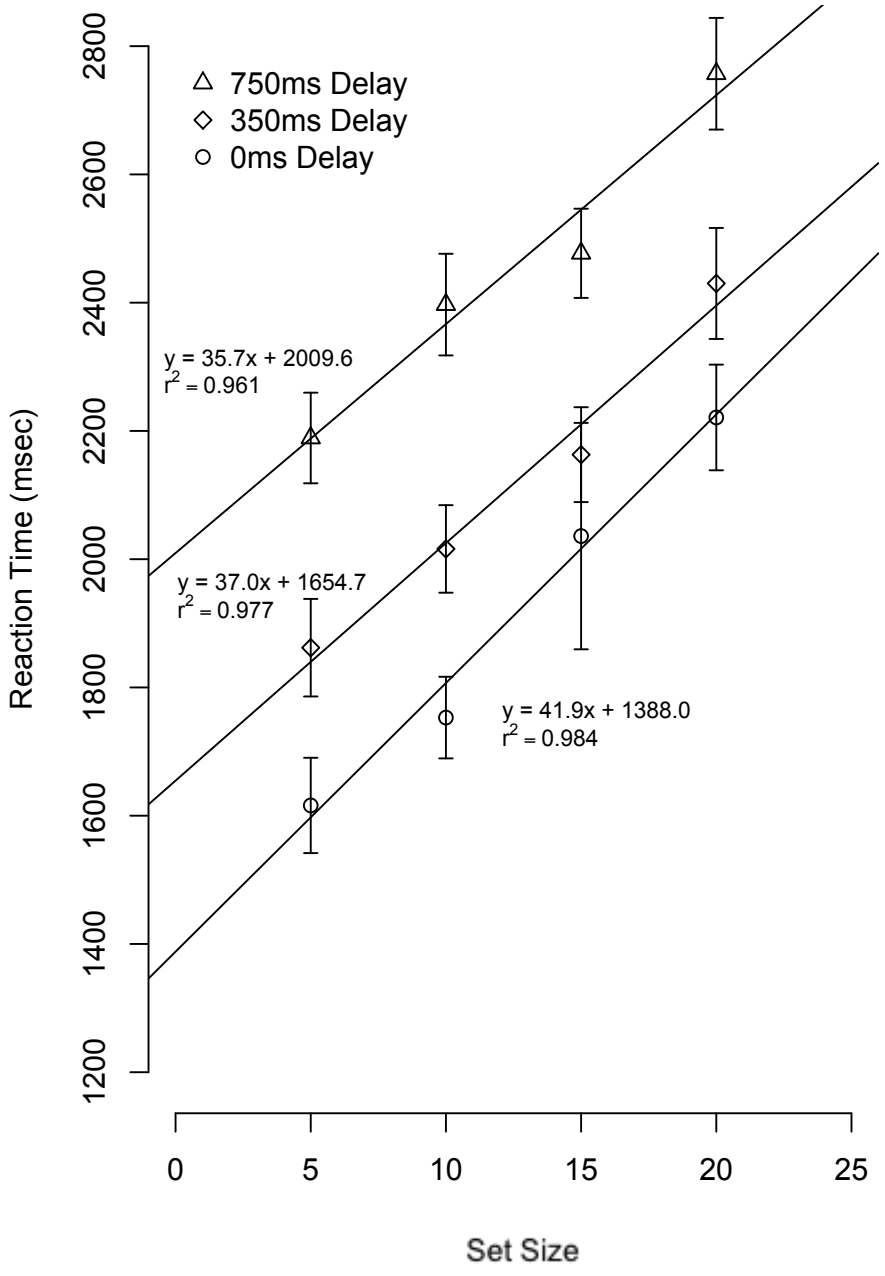

Figure 3.3: Results from Experiment 3A for target-absent trials (open symbols). Shown for the 0 ms delay condition (circle), 350 m delay condition (diamond), and 750 ms condition (triangle). Each line is accompanied by the best-fit leaner equation and the accounted proportion of variance (r^2). Error bars indicate standard error of the mean.

Figure 3.4

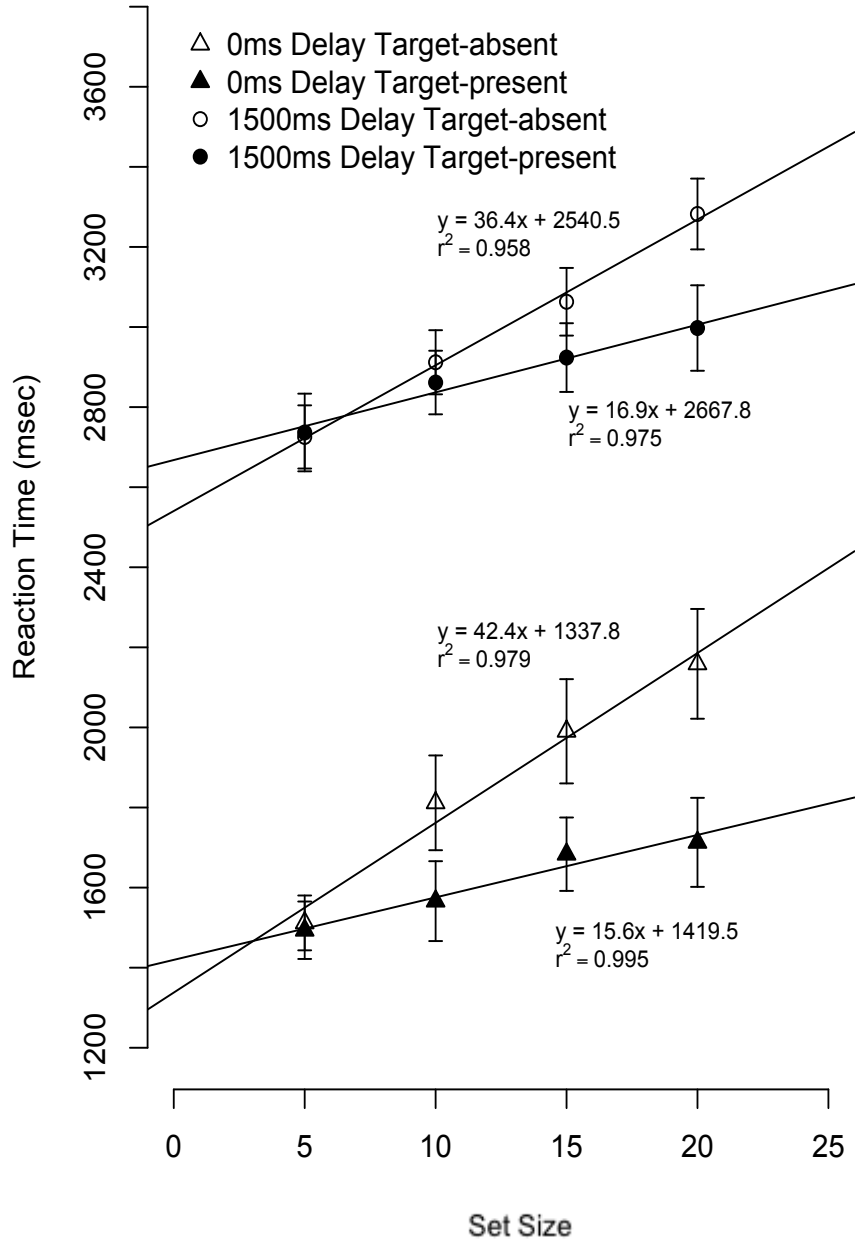

Figure 3.4: Results from Experiment 3B. Shown for target-present trials (filled symbols) and target-absent (open symbols) for the 0 ms delay conditions (triangle) and 1500 ms delay conditions (circle). Each line is accompanied by the best-fit linear equation, the accounted proportion of variance (r^2). Error bars indicate standard error of the mean.

Figure 4.1

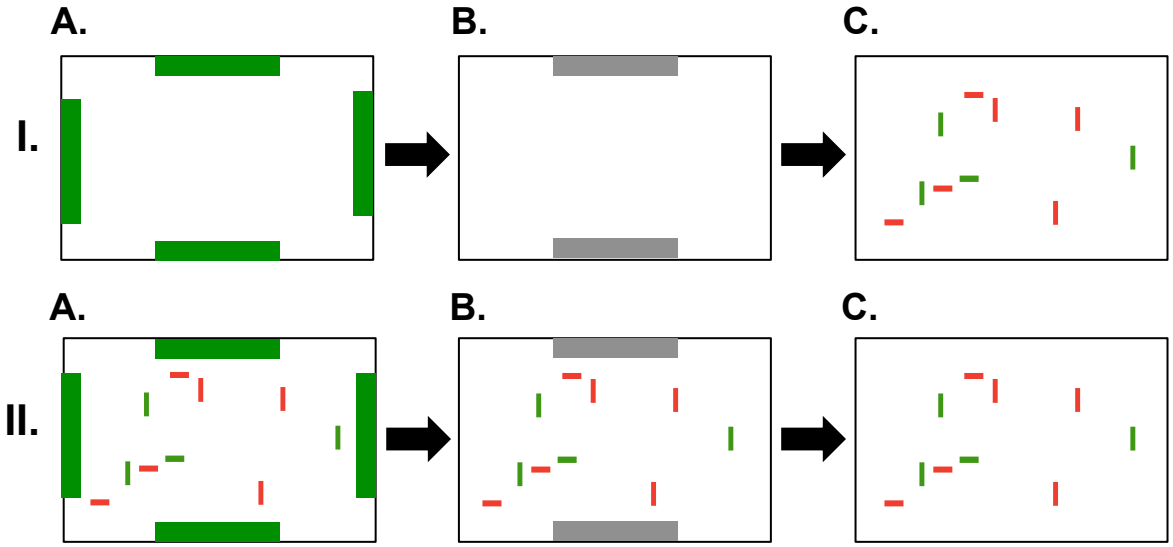

Figure 4.1: Example of nonlinguistic visual cue trial presentation for Experiment 4. Shown separately for cue-first (I) and cue-concurrent (II) conditions. Trial presentation for Experiment 4B are identical to Experiment 4A (500 ms for color & 1000 ms for orientation) with the exception that the duration of the color cue (A) lasted for 300 ms and the duration of the orientation cue (B) lasted for 600 ms. Experiment 4C uses the same cue timing as Experiment 4A but present orientation first (B to A to C)

Figure 4.2

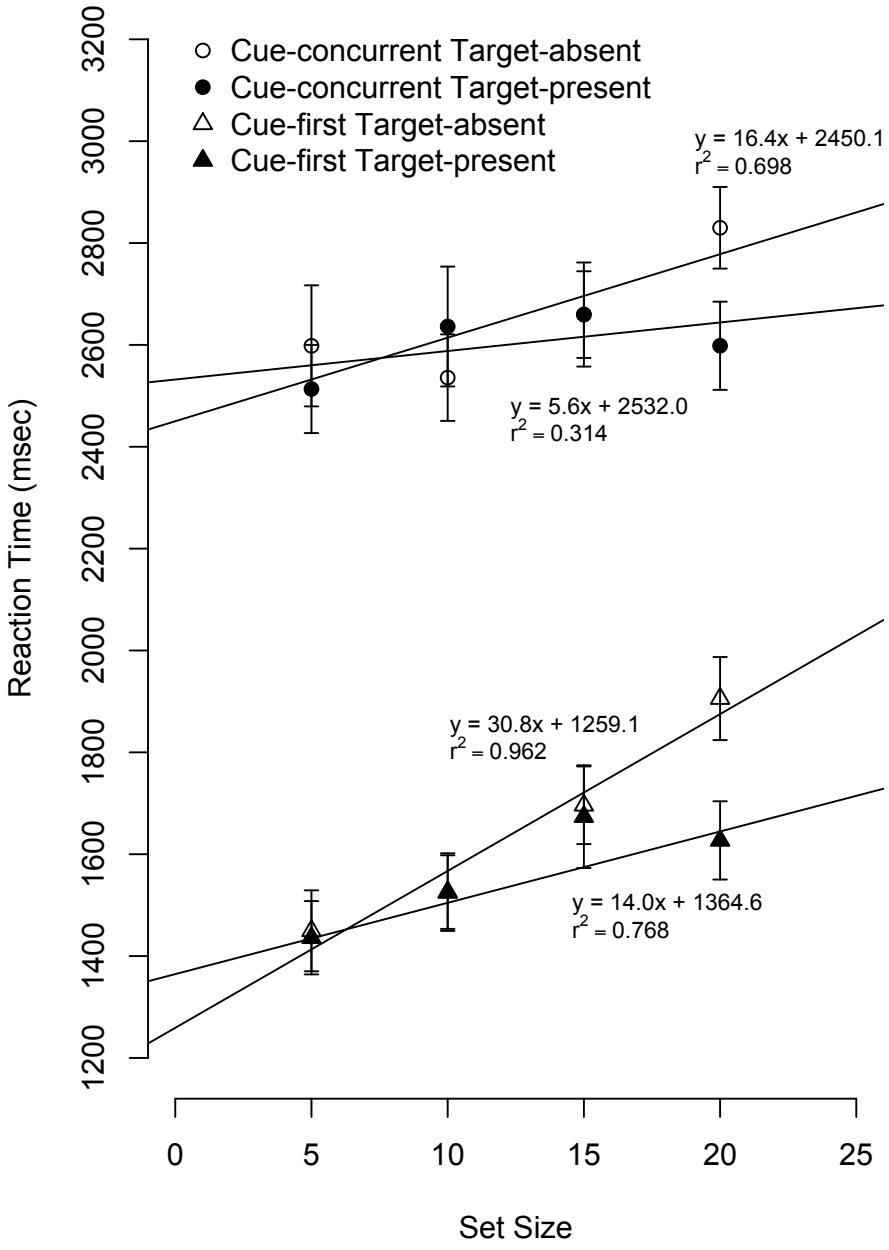

Figure 4.2: Results for Experiment 4A. Shown separately for target-present (filled symbols) and -absent trials (open symbols) for both cue-first (triangles) and cue-concurrent (circles) conditions. Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

Figure 4.3

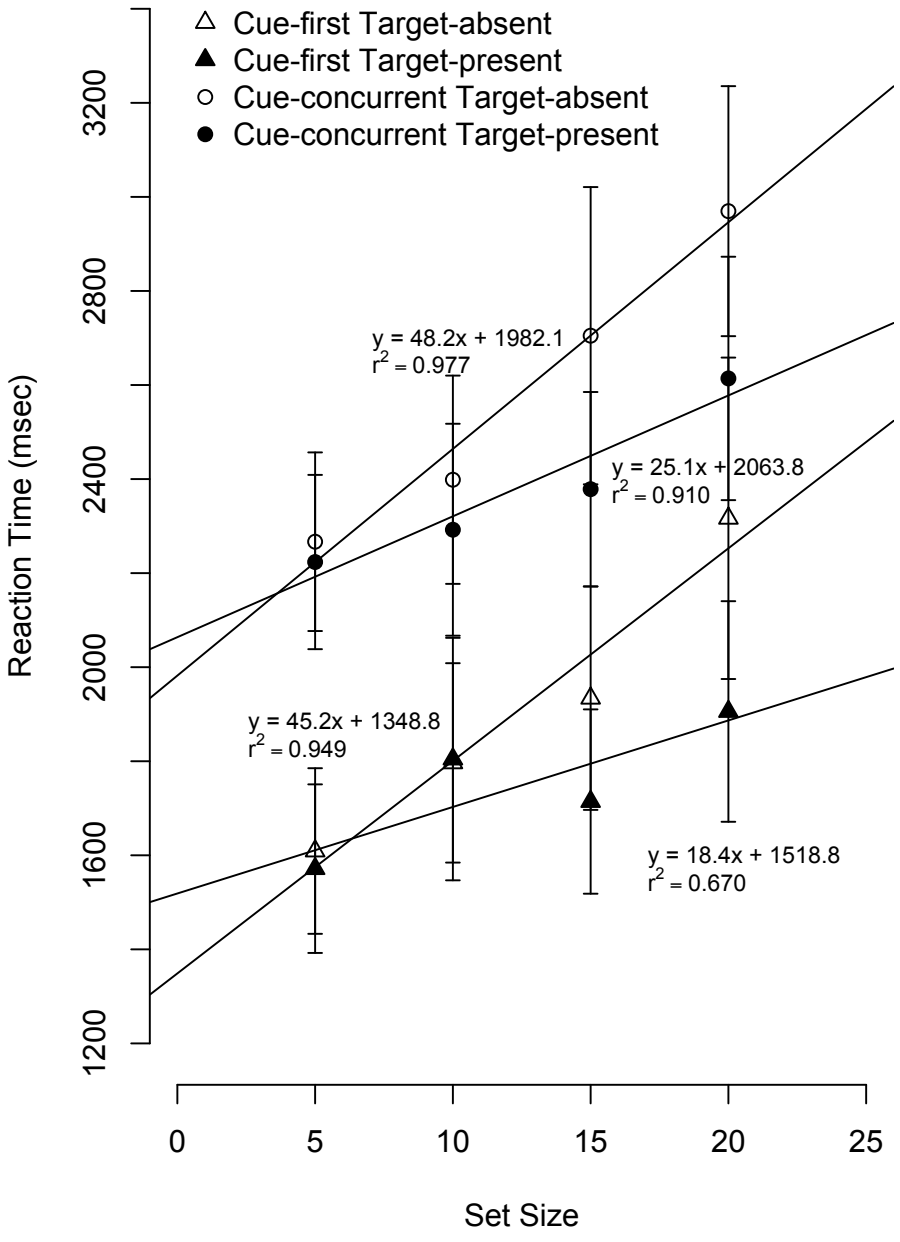

Figure 4.3: Results for Experiment 4B. Shown separately for target-present (filled symbols) and -absent trials (open symbols) for both cue-first (triangles) and cue-concurrent (circles) conditions. Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

Figure 4.4

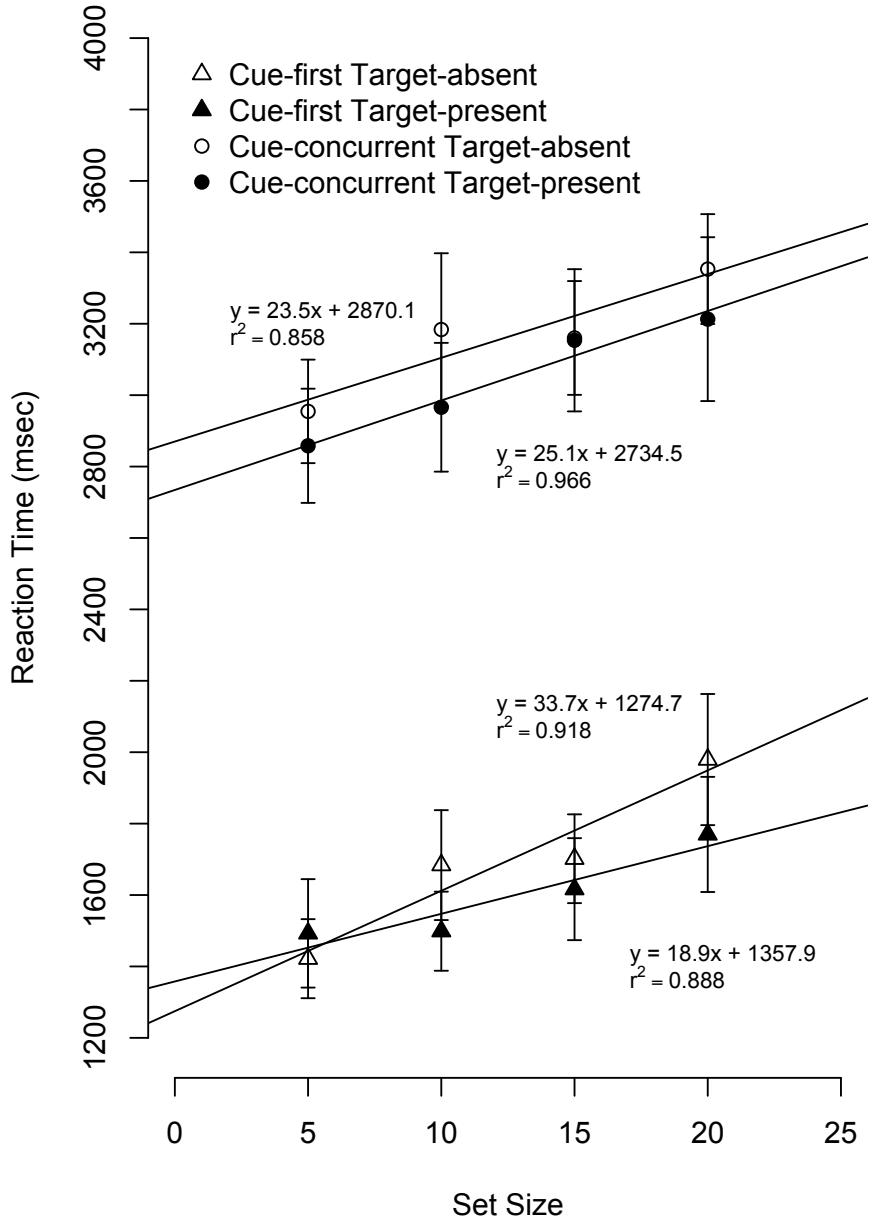

Figure 4.4: Results from Experiment 4C. Shown separately for target-present (filled symbols) and –absent trials (open symbols) for cue-first (triangles) and cue-concurrent (circles) conditions. Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

Figure 5.1

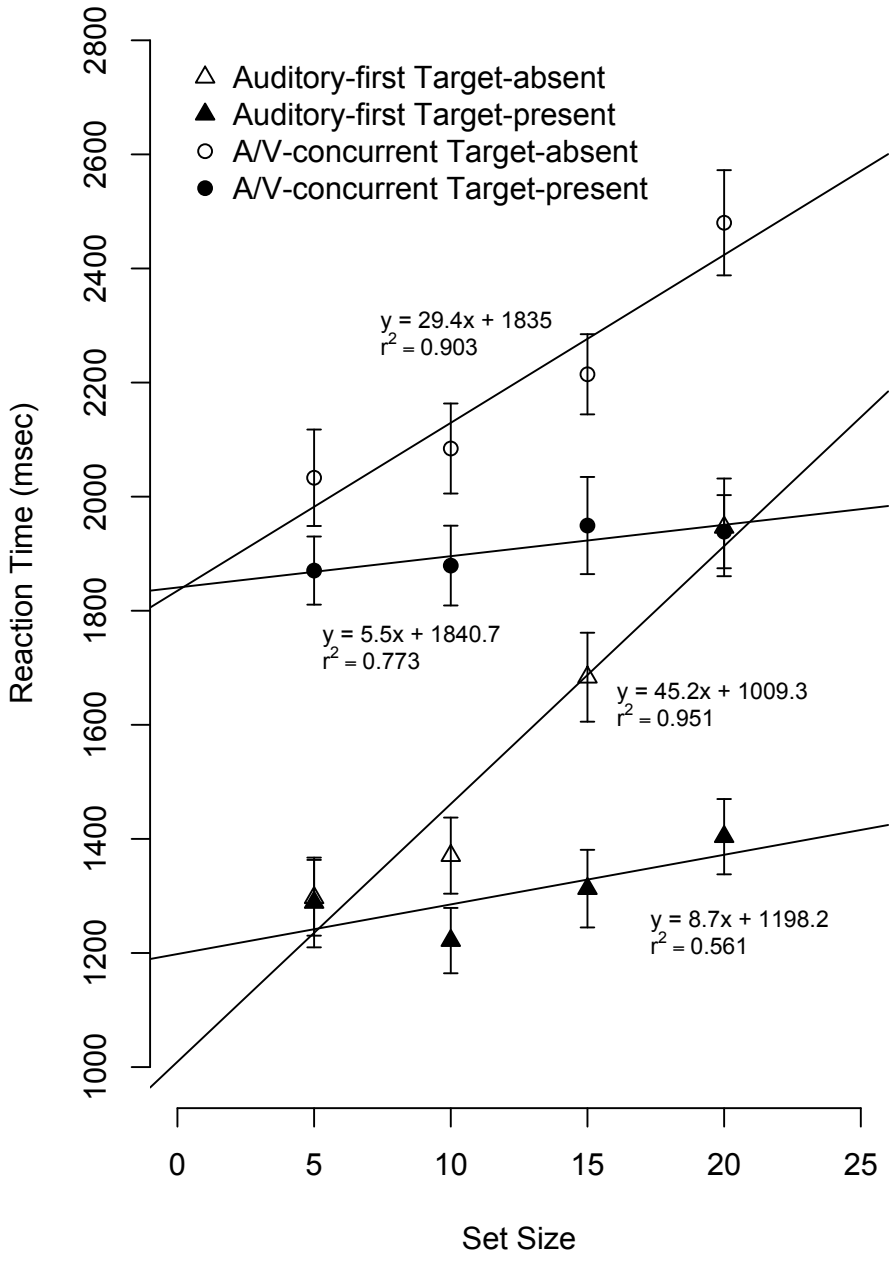

Figure 5.1: Results from Experiment 5. Shown separately for target-present (filled symbols) and –absent trials (open symbols) for cue-first (triangles) and cue-concurrent (circles) conditions. Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

Figure 5.2

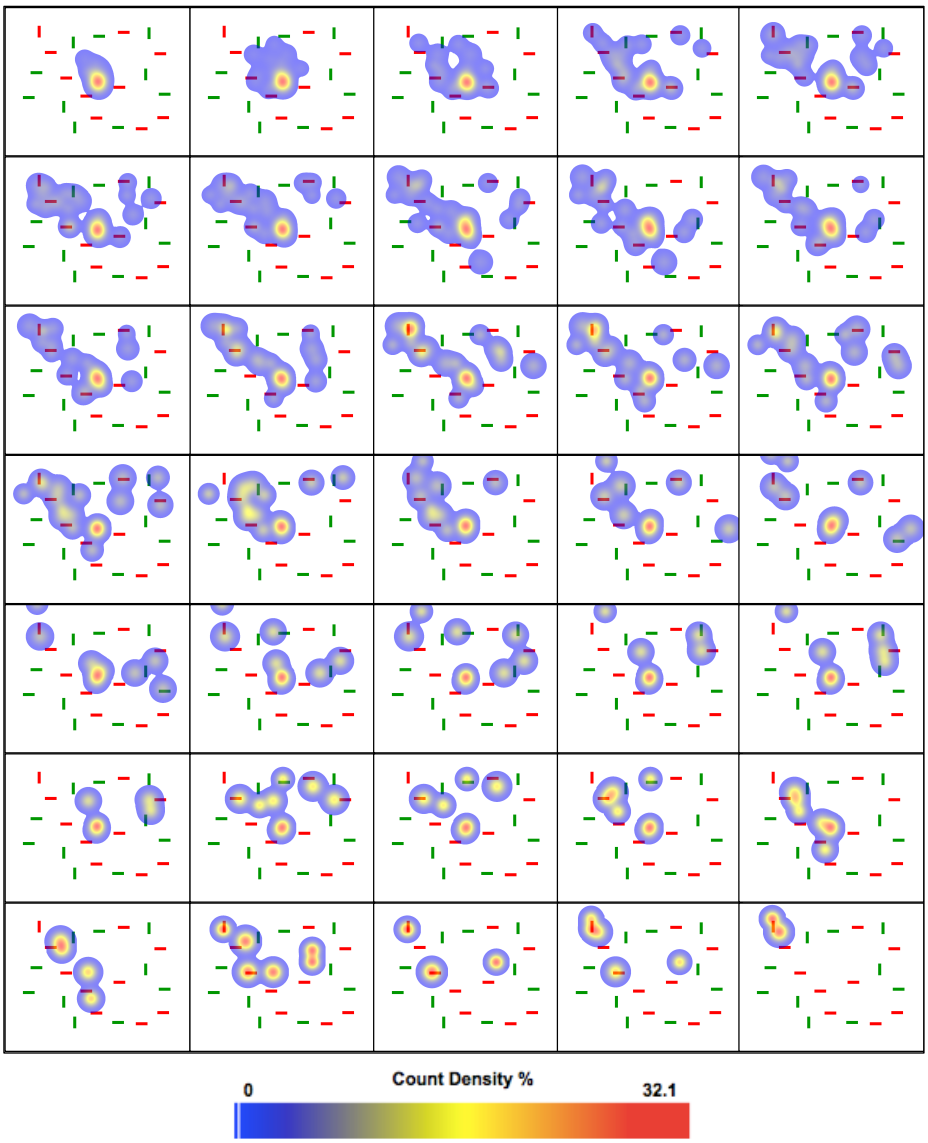

Figure 5.2: Eye-tracking results for Experiment 5. Search displays are overlapped with a heat map representing fixation activity (blue = low, yellow = medium, and red = high) for a target-present trial with a set size of 20. The target in this trial is a red vertical bar located in the top-left of the search display. Fixations for this figure are comprised of participants in Group B who received this search display in the auditory-first condition. Each frame represents 100 ms timesteps.

Figure 5.3

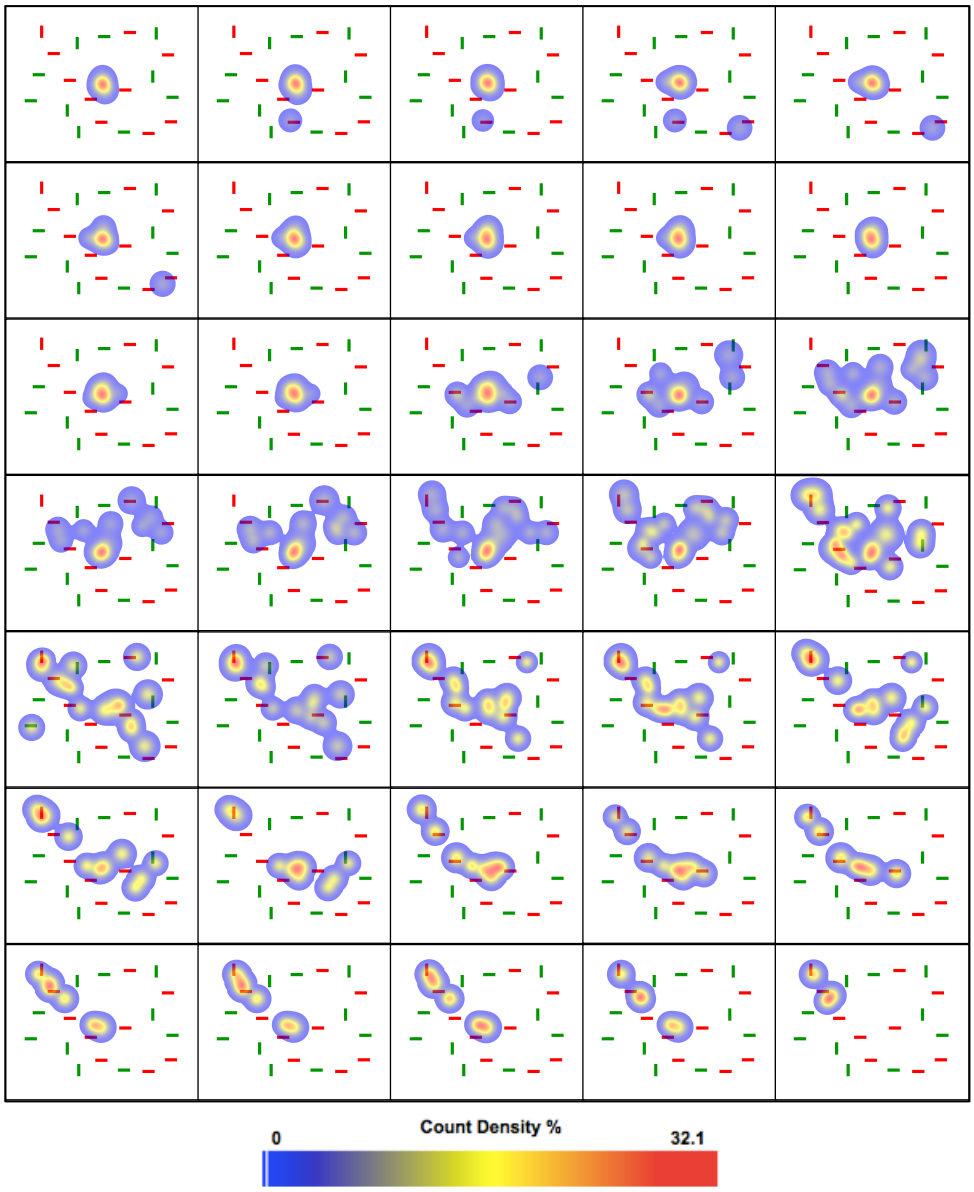

Figure 5.3: Eye-tracking results for Experiment 5. Search displays are overlapped with a heat map representing fixation activity (blue = low, yellow = medium, and red = high) for the same target-present (red vertical bar) trial with a set size of 20 depicted in Figure 5.2. Fixations for this figure are comprised of participants in Group A, who received this search display in the A/V-concurrent condition. Each frame represents 100 ms timesteps.

Figure 5.4

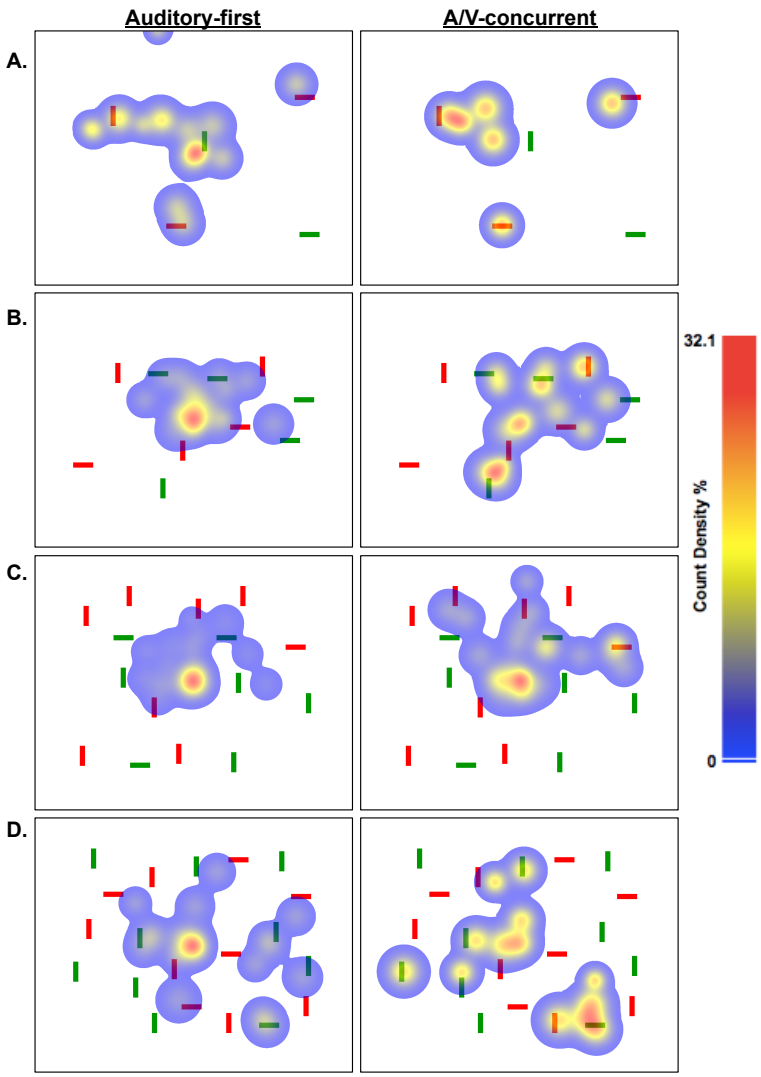

Figure 5.4: A closer look at the eye-tracking results from Experiment 5. Target-present trials are shown separately for auditory-first control and the A/V-concurrent trials. Search displays are overlapped with a heat map representing fixation activity (blue = low, yellow = medium, and red = high). A single 100 ms time period is depicted for each set size: 3100-3200 ms for 5 (A), 2400-2500 ms for 10 (B), 1700-1800 ms for 15 (C), and 2600-2700 ms for 20 (D). Targets for each trial are as follows: 5 = red vertical, 10 = green vertical, 15 = red horizontal, and 20 = green horizontal.

Figure 5.5

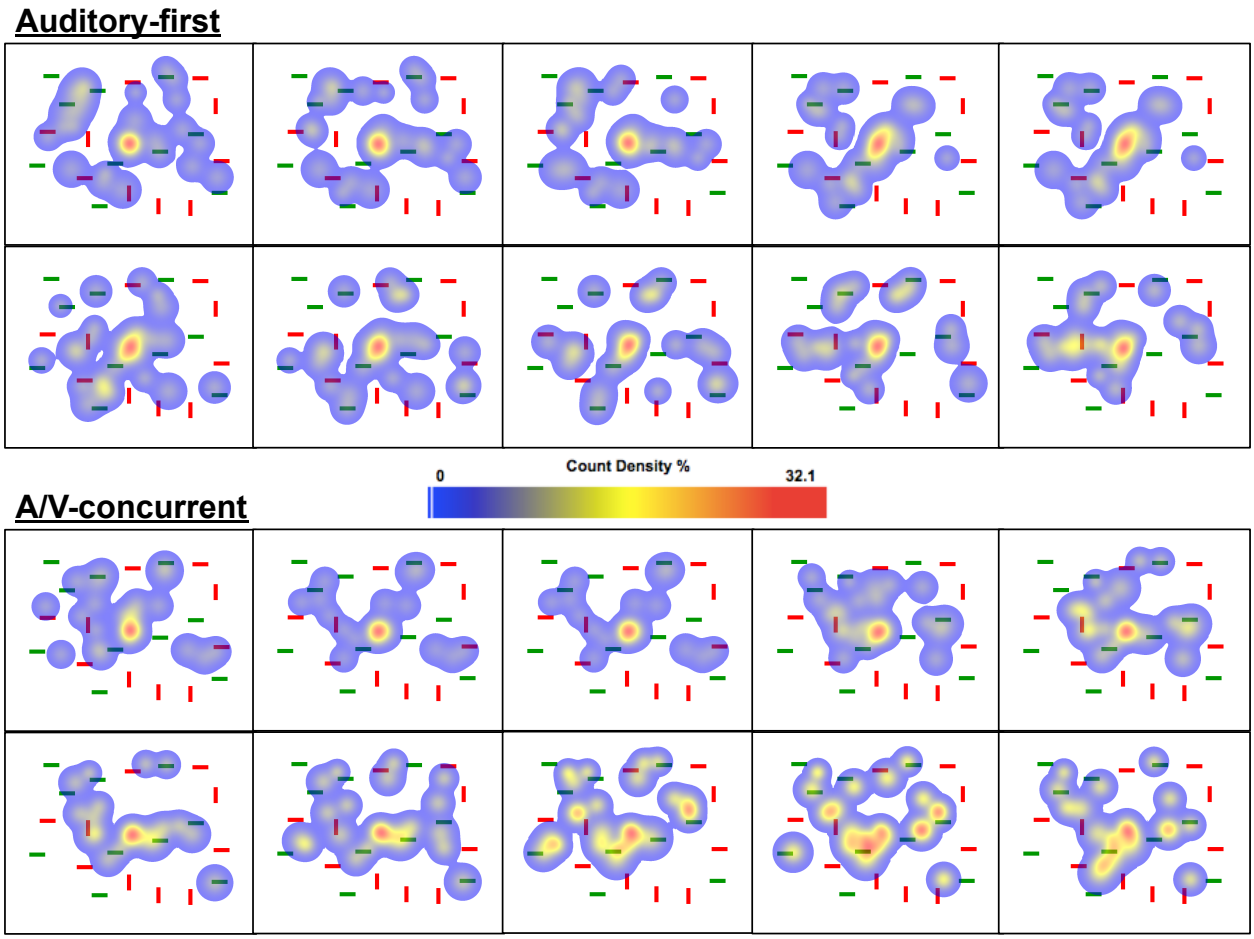

Figure 5.5: Results from Experiment 5 depicting eye-fixations for a target–absent (green vertical bar) trial with a set size of 20, shown separately for auditory-first and A/V-concurrent conditions. Each search display is overlapped with a heat map representing fixation activity (blue = low, yellow = medium, and red = high); each frame represents 100 ms time period.

Appendices

Appendix A:

Localist Attractor Model Matlab Code

```
%--------A/V Concurrent Visual Search-----------
clear

clf

hold off

n = 0;

for ss = 5:5:20                    % set sizes 5 to 20 in intervals of 5

    n = n + 1;

    integ = zeros(1,ss);           % object array set to 0
    red = (ones(1,ss))/ss;         % redness array set to 1
    vert = (ones(1,ss))/ss;        % verticalness array set to 1

    % redness were half of objects are red
    redin = [1 ones(1,(floor((ss-1)/2))) zeros(1,round((ss-1)/2))];

    % verticalness were half of objects are vertical
    vertin = [1 zeros(1,(floor((ss-1)/2))) ones(1,round((ss-1)/2))];

    t = 0;

    while max(integ) < 0.95

        t = t + 1;

        if t == 1                  % if t equals to 1 (beginning)

            red = red.*redin;      % setting up redness array
            redin;

        end

        if t == 10                 % 1 for Auditory-first 30 or 35
                                   % for A/V-Concurrent [or 40 at UCM]
                                   % for semi-concurrent: 10, 17, 23, and 30

           vert = vert.*vertin;    % setting up verticalness array

        end

        red = red/sum(red);        % normalizing redness
        vert = vert/sum(vert);     % normalizing verticalness

        integ = red + vert;        % integrating the two
```

```
                                    % (weights are 1 and don't add to 1)
        integ = integ/sum(integ);   % normalizing intergration

        if ss == 10                 % if setsize is 10

            integacts(t,:) = integ; % integ act for timestep=integ
            redacts(t,:) = red;
            vertacts(t,:) = vert;

        end

        red = red + integ.*red;     % integ feedback to rednes
        vert = vert + integ.*vert;  % integ feedback to verticalnes

    end

    % no reaction time added
    rts(n) = t*30 + 900;            % NOTE THE SLIGHT CHANGES ON BOTH OF
                                    % THESE COMPARED TO REALI ET AL. (for UCM
end                                 % students)


    subplot(4,1,1)                  % slow activation of first feature array
    plot(redacts(:,1),'k*-')        % the target object #1
        hold on
    plot(redacts(:,2:5),'k^-')
    plot(redacts(:,6:10),'ko-')
        title('Redness Activations')
        ylabel('Activation')
        xlabel('Time Step')


    subplot(4,1,2)                  % fairly steep activation of integration
    plot(integacts(:,1),'k*-')      % the target object #1
        hold on
    plot(integacts(:,2:5),'k^-')    % red horizontals; nodes 2-5
    plot(integacts(:,6:10),'ko-')   % green verticals; nodes 6-10
        title('Integration Activations')
        ylabel('Activation')
        xlabel('Time Step')


    subplot(4,1,3)                  % steep increase of second feature array
    plot(vertacts(:,1),'k*-')       % the target object #1
        hold on
    plot(vertacts(:,2:5),'k^-')
    plot(vertacts(:,6:10),'ko-')
        title('Verticalness Activations')
        ylabel('Activation')
        xlabel('Time Step')

    subplot(4,1,4)
        hold on
    plot([5:5:20],rts)
    axis([0 25 500 2000])           % x-axis (setsize) = [0 40]
```

```
                                % y-axis (ms) = [0 1200]
ylabel('Settling Time')
xlabel('Set Size')

slope = (rts(4)-rts(1))/15      % Slope
slope

rts
```