Postdiction and the effects of spatial, temporal, and feature compatibility on sensory integration

Thesis by

Dylan Nieman

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

1891

California Institute of Technology

Pasadena, California

2007

(Defended October 24, 2006)

Acknowledgements

First, I would like to acknowledge Shinsuke Shimojo for his mentorship, scholarship and support throughout this lengthy process. I would also like to thank the other scientists who have mentored me during the production of this work: Bhavin Sheth, Romi Nijhawan, Ryusuke Hayashi, and most notably Richard Andersen. For their fellowship, I would like to recognize my lab-mates and colleagues Junghyun Park, Patricia Neil, Daw-an Wu, Claudiu Simion, Beena Khurana, and Katz Watanabe. For assistance with the production of experimental equipment I would like to recognize Mike Walsh and Rick Paniagua. And for experimental assistance with vestibular experiments, I would like to thank Andrew Hsieh.

I would also like to thank my parents, John Nieman, Rick Reder, and Irene and Bruce Lehmann, for their constant support and encouragement throughout my life. Lastly and most importantly, I would like to recognize my wonderful wife, Emily Barrett, whose love and support have made this possible.

Abstract

The brain continually integrates stimuli over multiple sensory modalities and reconciles often disparate information into a unified, consistent representation of the surrounding environment. This process must be robust to differential neural latencies and imperfect alignments of spatial reference frames between sensory modalities. Numerous studies have examined the perception of multisensory stimuli with the presumption that multisensory integration is categorically different from within-modality integration. We looked at a variety of issues related to the updating of sensory reference frames and the integration of unimodal and multimodal stimuli over temporal and spatial disparities. Study 1 found simultaneous, opposite gaze-dependent aftereffects at the same retinal location for both depth and color, demonstrating the degree to which visual-coordinate space is gaze-contingent, not merely retinotopic. Study 2 found that the flash-lag effect, in which a flashed target is perceived as lagging behind a smoothly moving target, generalizes to third-order motion perception of cyclopean stimuli. Study 3 introduced a novel motion illusion which we termed the "turn-point phantom," wherein the position of an abrupt orthogonal direction change is mislocalized backwards along the object's subsequent trajectory. This effect, like flash-lag, can only be adequately explained with postdiction. Study 4 explored the effect of passive head or body turns on spatial perception of visual and auditory stimuli and found systematic mislocalization of pre-turn stimuli in the direction of the turn. This mislocalization decayed with added delay between target and turn onset. Study 5 examined spatial and temporal disparity in visual-motor ventriloquism and found that early visual distracters were essentially equivalent, whereas the influence of late visual distracters diminished with increasing asynchrony. Study 6 found suppression of saccade latency induced by stimulus repetition in certain multisensory experimental contexts. Together, these studies provide numerous examples supporting the idea that sensory perception, both unimodal and multimodal, is postdictive in nature, involving integration of sensory information over a time window that includes, but does not end with, task-relevant stimulus presentation. Additionally, these results provide clues to the character and relevant parameters of the integration process.

Contents

List of Figures

Chapter 1: Coordinate Frames, Multisensory Perception, and Postdiction

The human brain is continually bombarded with an overwhelming depth and breadth of spatial and temporal information that must somehow be synthesized into a common and consistent perceptual experience of the surrounding world. This information comes from multiple senses, in different reference frames, with different inherent neural latencies. In order to effectively reconcile competing and corroborating perceptual reports across time, space, and sensory modality, information must be represented in a common space and integrated over time. One of the more complex problems in the study of multisensory perception and awareness involves understanding how the brain integrates across such diverse information streams in order to reconcile perceptual time and space, what the relevant parameters affecting integration are, and when the time-window is over which integration occurs.

Coordinate Transformation

Information about the surrounding world comes to us in a variety of different coordinate frames. Our eyes monitor the relative intensities of light that impinge upon the retina. This information naturally exists in a retinotopic coordinate space. As our eyes move within the head, changing the direction and location of their gaze, this coordinate map's relation to the outside world shifts as well, and comparison of images before and after eye movements requires accurate information about the relative position of the eye before and after the shift. This information may come from stretch receptors in the ocular muscles, or from the monitoring of efferent motor signals sent to those extraocular muscles. Similarly, as the neck turns and the head moves with respect to the body, the retinotopic coordinate space shifts with respect to the environment, and in order for the brain to account for this, it must utilize information about head and neck position. This information may come from stretch receptors in the neck muscles, or from the monitoring of efferent motor signals sent to those muscles. Finally, as the body moves within the environment, rotating or translating through space, the retinal coordinate space shifts, and the brain must take into account cues from the vestibular system to track these shifts. All of this extra-retinal information—gaze direction, head position, body movement within the environment—must be consistently integrated with the retinal information in order to maintain a consistent representation of visual space over time.

In a similar fashion, our ears monitor air pressure changes and use interaural timing and intensity differences to perceive sound in space. Auditory information naturally exists in a head-centered reference space, and information related to the continued monitoring of head position—neck position, body movement within the environment—must be utilized in maintaining a consistent representation of visual space over time.

Multisensory Integration

Just as perception within each sensory modality must utilize information about how the correlation between its coordinate map and the outside world changes, unified perception across sensory modalities requires information about how each modality's coordinate space correlates. Unless this information about relative coordinate frame mappings is perfect some discrepancy in position representation across modalities will arise, therefore multisensory perceptual systems must be robust to some degree of spatial disparity between modalities.

Multisensory perceptual systems must also be robust to some degree of temporal disparity between modalities in order to properly bind representations of multi-modal stimuli. This is for two reasons: first, visual and auditory information travels through the outside world at vastly different rates, and second, visual and auditory information have different neural latencies in the brain. Light travels through the atmosphere at speeds that are (for these purposes) essentially instantaneous, but sound travels at about 340 meters per second. This means that a synchronized sound and light at a distance of 1 meter arrive at the sensory organs 2.9 ms out of sync, while the same stimulus at 10 meters arrives 29 ms out of sync, and at 100 meters arrives

290 ms out of sync. Yet the brain has no problem experiencing the visual and auditory events as unified at a variety of different relative synchronies. There are, of course, limits to this experience of unity even for singular events (think of the discrete experiences of thunder and lightning). Compounding this, the passage of visual and auditory information through their relevant sensory organs and neural circuits to multimodal regions of the brain involves different inherent latencies. For the superior colliculus, a brain region involved in orienting behaviors and the planning of eye movements, the discrepancy range across 90% of the population is between 20 and 115 ms (Meredith, Nemitz, and Stein 1987).

This consistent unified experience of multisensory stimuli must involve integration across modality that is robust to some degree of discrepancy in both the spatial and temporal alignment of stimulus representation.

Perception and Physiology

While multisensory integration is essential for feature binding of object representations, it is also responsible for a wide variety of perceptual phenomena. Multisensory interactions can modify or change the perception of stimulus properties such as detectability (McDonald, Teder-Salejarvi, and Hillyard 2000), discriminability (Kennett, Taylor-Clarke, and Haggard 2001), subjective brightness (Stein and Wallace, 1996) or duration (Vroomen and de Gelder 2000), subject to the presence of a second stimulus. The processing of speech is greatly aided by the visual correlation of lip movement information. The McGurk effect (McGurk and Macdonald 1976) is a well-known side effect of this auditory and visual speech-perception integration, in which the presentation of an auditory “ba” accompanied by a visual “ga” results in a perceptual experience of “da.” It is interesting to note that in the McGurk effect, the integration of two disparate stimuli does not result in one representation or the other solely winning out, but in the synthesis of a percept distinct from either.

In the stream-bounce illusion (Sekuler, Sekuler, and Lau 1997), the perception of an ambiguous, spatial coincidence event is categorically changed by the timing of an auditory stimulus. Two identical disks, starting in opposite top corners of a computer screen and moving diagonally down to opposite corners such that they cross in the center, can be perceived as crossing and streaming past each other, or reflecting and bouncing off of each other. Presenting an auditory stimulus with sharp rise and fall times around the time of the disks' spatial coincidence results is a strong bias towards the "bounce" percept, whereas presentation of the same auditory stimulus presented with sufficient temporal disparity before or after the moment of coincidence results in the opposite bias towards the "stream" percept. The maximal effect of the accompanying sound occurs not with perfect synchrony, but when the sound precedes the spatial coincidence by 150 ms. In a similar effect, auditory stimuli accompanying a flashed visual stimulus have been found to generate additional perceived illusory flashes (Shams, Kamitani, and Shimojo, 2000). That is to say, a single visual flash presented with a brief double beep is perceived as two separate visual flashes.

In localization tasks, responses to audio-visual targets have faster response latencies and improved accuracy compared to auditory-only or visual-only targets (Hughes *et al.*, 1994). When the auditory and visual stimuli are spatially separated, auditory targets presented with disparate visual distracters are perceived displaced in the direction of the visual distracter (Hairston *et al.*, 2003). This "ventriloquism" effect has a number of correlates. When spatially aligned auditory and visual stimuli are presented with sufficient temporal disparity (150–200 ms) they can be perceived as spatially separated (Jack and Thurlow, 1973; Radeau and Bertelson, 1977; Slutsky and Recanzone, 2001). Whereas in the standard ventriloquism effect, synchronous presentation of spatially disparate stimuli causes the stimuli to be perceived closer in space, in "temporal ventriloquism" spatially aligned presentation of asynchronous stimuli causes them to be perceived as temporally closer (Bertelson and Aschersleben, 2003).

There is also a role of uncertainty in the integration of auditory and visual stimuli. For clearly defined auditory and visual stimuli, the high spatial acuity of vision tends to bias auditory localization towards a visual target (ventriloquism). Conversely, auditory targets do little to bias the localization of visual targets. However, it has been found that with less well-defined stimuli this is not always case. Auditory stimuli with a Gaussian temporal envelope that does not affect the localization of a small moving Gaussian blob with low spatial uncertainty, can bias the localization of a larger Gaussian blob with higher spatial uncertainty (Heron, Whitaker, and McGraw 2007).

Physiological substrates for audio-visual integration have been at the level of individual neurons, particularly in the superior colliculus (Meredith and Stein, 1983, 1986; Wallace, Wilkinson, and Stein 1996). Characteristic features of these multi-modal neurons include maximal excitation with lower stimulus intensity levels than those associated with unimodal stimuli and enhanced activation in response to synchronous, spatially coincident multimodal stimuli. The enhanced activity associated with synchronous stimuli was found to decrease monotonically to zero as the overlap decreased between peak discharge periods of activation evoked by each modality, and even result in depression of activity for sufficiently large temporal disparities (Meredith, Nemitz, and Stein 1987).

Thus, in the integration of auditory and visual stimuli, spatial and temporal alignments clearly play important roles in determining the degree and type of interaction between modalities. How exactly to model this interaction has proven confounding given the variety of known perceptual and visual-motor effects.

Time-Window-of-Integration

Multisensory integration does not require perfect synchrony, but occurs over some time window surrounding stimulus events. This often means that subsequent events can affect the perception of prior stimuli. Affects of this nature are not limited to multisensory perception, but occur within modality as well. In backwards masking, the perception of a stimulus can be modulated or even ablated when it is followed in close succession by a second stimulus (Bachmann, 1994). In the color phi phenomenon, two colored targets presented sequentially across small spatial and temporal disparities appear to change gradually from one color and position to the other (Kolers and von Grunau, 1976). The perception of the first stimuli during the time between the two presentations is influenced by the presentation of the second, though the latter has not yet occurred. This is the basic effect underlying apparent motion and video display.

In the flash-lag effect, a flashed visual target is perceived to lag behind a smoothly moving target (Nijhawan, 1994). This effect is demonstrated for conditions in which the flash occurs at the same time as the motion target's onset, but does not occur when the flash is timed with the disappearance of the motion target, implying that it is the visual stimuli subsequent to the presentation of the flash that are responsible for this effect (Eagleman and Sejnowski, 2000).

The "cutaneous rabbit" effect even demonstrates this in the haptic domain (Geldard and Sherrick, 1972). In this effect, reminiscent of the phi phenomenon, a rapid succession of taps to the wrist followed by a rapid succession of taps to the elbow is experienced as a series of equidistant taps regularly spaced from one to the other. The spatial perception of wrist taps is influenced by subsequent elbow taps just as the spatial perception of elbow taps is influence by prior wrist taps. The latter is hardly surprising if we imagine that the series of wrist taps predict future wrist taps, and this prediction biases the subsequent experience of elbow taps. But for the subsequent elbow taps to bias the prior wrist taps, we must imagine that the series of elbow taps postdicts the wrist taps.

There are a variety of known postdictive phenomena both within and across modalities. A variable time-window-of-integration assumes that stimuli occurring in close temporary proximity should exert some influence on each other's perception. There is no reason to believe that integration should be limited to, or even biased in favor of prior events as opposed to those subsequent to the relevant perceptual stimuli, or that this sort of integration should only occur across sensory modalities and not within sensory domains.

Summary of Experimental Studies

The following chapters represent two published papers, three manuscripts in preparation, and work from one ongoing project. The first experimental study, on gaze-dependent aftereffects, demonstrates the ability of the visual system to maintain gaze-specific representations of visual space at the same retinotopic position (Nieman, *et al*., 2005). The second study, on cyclopean flash-lag illusion, demonstrates a novel flash-lag presentation and represents an example of a known effect that is likely postdictive in nature (Nieman *et al*., 2006). The third study, introducing the turn-point phantom effect, presents a new visual motion illusion with evidence suggesting that it is postdictive in nature. The fourth study, on the displacement of the perceived positions of visual and auditory targets subject to immediately subsequent passive head or body turns, explores a postdictive effect resultant from vestibular sense integration. The fifth study, on temporal factors related to visual-motor ventriloquism, explores the some of the spatial and temporal factors affecting the integration of auditory and visual stimuli for orienting. And the sixth study, on the suppression of saccadic response by stimulus repetition within a multisensory context, introduces another postdictive phenomenon in the visual-motor domain with surprising contextual effects that suggest how overall perceptual context affects the time-window of sensory integration. Taken together, all of this suggests a general model of stimulus integration that is task specific and operates both within and across sensory modalities.

Chapter 2

Study 1: Gaze Direction Modulates Visual Aftereffects in Depth and Color

(Previously published as Nieman *et al.* 2005)

Abstract

Prior physiological studies indicate that gaze direction modulates the gain of neural responses to visual stimuli. Here, we test gaze modulation in the perceptual domain using color and depth aftereffects. After confirming retinotopy of the effects, we employed a balanced alternating adaptation paradigm (adaptation alternates between opponent stimuli) to demonstrate that opposite color and depth aftereffects can codevelop at the same retinal location for different gaze directions. The results provide strong evidence for (a) gaze modulation of aftereffects, (b) generality of gaze modulation across two visual attributes, and (c) perceptual correlates of the modulation of neural activity by gaze direction.

Introduction

Due to frequent spontaneous saccades, the retinal image undergoes repeated and continual relocation. The problem of spatial constancy—how we derive and maintain a highly stable visual world despite this highly dynamic visual stream—has long been one of the central concerns of vision science. As the position and orientation of the retina changes due to movements of the eye within its orbit, the head atop the shoulders, or the body with respect to its surroundings, the spatial mapping necessary to correctly and consistently interpret a visual scene is in constant flux. Information regarding the position of the eye must provide a context for the transformation of retinal coordinates into environmental locations.

Physiological evidence suggests that a critical basis for this computation may be provided by adjusting the gain of visual neurons in a gaze-dependent manner. It has been shown with monkeys that presenting the same visual stimulus at the same retinal position can give rise to very different neuronal responses depending on an animal's gaze direction (Andersen and Mountcastle, 1983). A given neuron will still respond maximally when the stimulus is presented at its tuned retinal location, but the amplitude of activity at that location and presentations elsewhere in the visual field will systematically vary with gaze. These gaze-modulated receptive fields were first observed in parietal cortex (Andersen *et at*., 1985) and have been shown to be associated with performance on visually guided motor tasks (Andersen and Mountcasle, 1983). "Gain field" neurons, synthesized from these kinds of gaze-modulated receptive fields, could be the basis for a head-centered representation of visual space and may indeed be the neural substrate of such coordinate transformations (Zipser and Andersen, 1988; Pouget and Sejnowski, 1997). These findings have not been limited to visio-motor systems. Recent physiological studies have shown similar neural modulation in areas V1 (Trotter and Celebrini, 1999), V3 (Galletti and Battaglini, 1989), and MST (Squatrito and Maioli, 1997; Shenoy, Bradley, and Andersen., 1999).

Are these types of gaze-modulated neural responses in perceptual areas reflective of perception, and if so are they psychophysically detectable? To explore this possibility, we utilize visual aftereffects, which have the benefit of being both generally retinotopic, and sensitive to spatial-visual context across gaze shifts. To put this another way, aftereffects are convenient for this study because they tend to be focused at a specific retinal location (the adapted location) and are also somewhat robust to eye movements.

For our studies we used a depth-ordering aftereffect (DOAE) and a color aftereffect (CAE). In the DOAE, adaptation to overlapping regions simultaneously presented at different disparity-cued depth planes causes a neutral test stimulus (both regions presented at equal depth) to appear opposite their adapted orientation. This is a variation of Blakemore and Julesz's original depth aftereffect (Blakemore and Julesz, 1971) in which adaptation to a depth plane nearer than fixation caused a neutral depth plane to appear farther away while adaptation to a depth plane beyond fixation caused a neutral plane to appear closer. Our preliminary observations suggested that our depth-ordering version of stimuli tended to yield clearer percepts, thus providing a more consistent aftereffect. The stimuli used for these experiments were overlapping horizontal and vertical rectangles composed of random dots whose depths were defined by disparity without monocular cues (figure 2.1).

We also tested the color aftereffect (CAE), in which prolonged adaptation to a patch of one color causes a gray test stimulus to appear in the opponent color (adaptation to green produces a red aftereffect). It is known that the CAE is not subject to interocular transfer (Coltheart, 1973)—adaptation in one eye does not produce an aftereffect for a stimulus viewed only with the other. This suggests that the effect occurs somewhere prior to the integration of binocular information. The DOAE is based upon adaptation to depth planes defined by binocular disparity cues, necessitating information from both eyes. While both effects are likely to occur relatively early in the visual system (V1 or V2) the depth effect involves a greater degree of neural integration and probably occurs later.

Even before the physiological data suggested a mechanism for it, Mayhew reported a compelling example of a gaze-dependent motion aftereffect in humans (Mayhew, 1973). After alternate their gaze between a clockwise rotating disc on the left and a counterclockwise rotating disc on the right while maintaining a fixed head position, observers reported the direction of motion aftereffect (MAE) observed for a static disc presented in each gaze direction. Mayhew found that the direction of the aftereffect varied with gaze direction, each always opposite to that of its adapting stimulus. Two simultaneous, opposite aftereffects at the same retinal position could not be explained by a purely retinotopic mechanism and the physiology to suggest a mechanism for gaze integration was unknown at the time. A recent study, enlightened by physiological findings, has demonstrated gaze modulation of motion, tilt, and size aftereffects, although the size of the modulation was modest (approximately 15%; Nishida *et al.*, 2003). Unlike Mayhew's study, gaze-dependent opposite aftereffects were not demonstrated.

We revisit gaze alternation between opponent adapting stimuli as a means of examining the gaze modulation of aftereffects in depth and color. These effects likely occur in different neural pathways of visual processing (Livingstone and Hubel, 1984) and may reflect on the generality (or specificity) of gaze modulation in visual processing. While Mayhew's original demonstration provided a compelling example of gaze modulation, the spatial extent and selectivity of the effect were never quantified. Here we reexamine and augment the original paradigm to measure the degree of modulation for each effect. After measuring the spatial tuning of these aftereffects, we conducted experiments in which observers adapted by repeatedly alternating fixation at fixed time intervals between a location in one direction (right), which contained an adaptation stimuli, and a location in the other (left), which contained only a fixation point. Observers were then asked to make a saccade to a neutral location (center) before shifting their gaze to a test pattern presented at one of the two adapted locations, or the neutral location. We found significant gaze-dependent modulation of both color and depth aftereffects.

In other experiments, observers regularly switched their gaze location between opponent stimuli at the two locations (red on the right and green on the left, or horizontal-in-front on the right and vertical-in-front on the left), alternately adapting to opposite stimuli at the same retinal location (in the fovea). Again, after being asked to saccade to a neutral central location, observers were presented with a test stimulus at one of the three locations (figure 2.2). We found evidence of gaze modulation of aftereffects, i.e. opposite directions of aftereffect in two gaze directions, in the case of both color and depth.

Experiments

Before directly examining the gaze modulation of these aftereffects, we first sought to establish a means of measuring the size of each effect under fixed gaze. Methodology for probing of the two aftereffects is discussed in detail below.

Depth apparatus: All experiments were performed on Macintosh computers running MATLAB (MathWorks Inc., Natick, MA) and the Psychophysics Toolbox extensions (Brainard, 1997). Stimuli were displayed on an LCD screen (60 Hz refresh; 32.3 cm × 21.6 cm) at a viewing distance of 57 cm (1 cm on screen = 1° of visual angle). Observers wore red-green stereo-glasses and room lights were turned off. The apparatus was the same in all depth experiments.

General notes on depth stimuli: The adaptation stimuli consisted of 840 random dots arranged in two 5.0° X 1.67° overlapping bars (one horizontal, one vertical; see the top of figure 2.1 for an example). Each dot was displayed in both red and blue, and the horizontal disparity between each dot's red and blue components was adjusted to convey depth. All dots appeared at one of three depths corresponding to horizontal crossed disparities of 6.6, 10.0, and 13.2 arcmin (far, middle, and near). For the adaptation stimuli the two bars were presented at different depth planes (100% of dots composing the horizontal bar appearing near, 100% of dots composing the vertical bar appearing far, and vice versa). Throughout the experiment, observers fixated a 0.29° circle presented at the far depth plane (6.6 arcmin red-blue horizontal disparity) that was always visible through the adaptation stimuli.

In order to quantify this effect, we defined a "depth index" for the test stimuli corresponding to the percentage of dots in the stimulus that are not at the middle (10.0 arcmin) disparity. Positive depth index is assigned to stimuli in which the vertical bar appears nearer than the horizontal bar, negative depth index is assigned to stimuli in which the horizontal bar appears nearer than the vertical bar. For example, a stimulus with a depth index of +60 would have 60% of the dots that compose its vertical bar at the near-depth disparity (the other 40% at the middle-depth disparity) and 60% of the dots that compose its horizontal bar at the far-depth disparity (the other 40% at the middle disparity). Conversely, a stimulus with a depth index of −60 would have 60% of the dots that compose its vertical bar at the *far*-depth disparity, 60% of the dots that compose its horizontal bar at the *near*-depth disparity, and the other 40% of both bars at the middle disparity. The vertical-near/horizontal-far adaptation stimulus has a depth index of +100 (100% of the vertical-bar dots near, 100% of the horizontal-bar dots far) and the horizontal-near/vertical-far adaptation stimulus has a depth index of -100 (100% of the vertical-bar dots far, 100% of the horizontal-bar dots near). For a perfectly neutral test stimulus, 100% of the dots representing both the vertical and horizontal bars were presented at the middle disparity. This stimulus receives a depth index of zero (figure 2.1).

Color apparatus: All experiments were performed on Macintosh computers running MATLAB (MathWorks Inc., Natick, MA) and the Psychophysics Toolbox extensions (Brainard, 1997). Stimuli were displayed on a CRT Monitor (LaCie electron21, 60 Hz refresh; screen dimensions: 38.4 cms X 29.0 cms) at a viewing distance of 55 cm. The apparatus was the same in all color experiments.

General notes on color stimuli: The adaptation stimuli consisted of 3.47° x 3.47° squares of red (RGB(255, 0, 0), CIE xyY(0.6357, 0.3411, 11.19)) and blue green(RGB(0, 255, 255), CIE xyY(0.2145, 0.3015, 37.20)). To avoid issues with edge overlap, test stimuli consisted of smaller 2.08° x 2.08° color squares. Throughout the experiment, observers fixated a 0.28°X 0.28° square of white.

In order to quantify this effect along a single dimension, we constructed a "color index" for the values used in the test stimuli. The palette of test colors was chosen along the RGB interpolation of the red and blue green adaptation stimuli. All colors used as test stimuli were then measured by X-Rite Color Monitor Optimizer, plotted in xy-space of the CIE xyY color system, and fitted with a linear regression. Each test color was then orthogonally projected onto the regression line and assigned a color index corresponding to its distance from the scale's origin, which was set at the orthogonal projection point of monitor white (RGB(255, 255, 255)). Positive color indexes signified redder values and negative indexes signified more blue green values. The red adaptation stimulus corresponded to a color index of +0.3366 while the blue green adaptation stimulus corresponded to a color index of -0.0805.

Analysis: All experiments used the interleaved fixed-step-size staircase method. Eye movements were not monitored. Observers responded after each trial by pressing one of two adjacent keys. Each key specified a fixed judgment (horizontal/vertical near or red/blue green) and the stimulus was then adjusted accordingly to achieve cancellation of the effect in subsequent trials. For each experimental condition observers ran two staircases, one beginning at each adaptation stimulus. Staircases terminated after observers had reversed their responses five times. The point of subjective equality for each condition was determined by the average of six points (the last three reversal points from each staircase). PSEs were compared using two-tailed, non-paired t-tests. A typical experiment lasted between 15 and 20 minutes. Reaction times were not measured.

Experiment 1

Our first task was to quantify the effects and to demonstrate their spatial tuning on the retina without a gaze-shift. Observers maintained central fixation throughout the experiment as they were presented with an adaptation stimulus in the fovea and test patterns at five retinal locations.

Participants: Two authors and five naive observers with normal or corrected to normal vision were used in all experiments.

Depth stimuli: Observers began each experimental session by adapting to the vertical-in-front stimulus at the center of the screen for 20s. At the beginning of each trial, that stimuli would reappear for 3s, then disappear and 0.5s later, a test stimulus would appear at one of five locations (8.4° left, 4.2° left, 0°, 4.2° right, 8.4° right). The test stimuli would be visible on the screen for 0.7s.

Depth task: Observers were asked to judge, in a 2 AFC task, whether the horizontal bar of the test stimulus appeared in front of or behind the vertical bar. They were instructed to hold their gaze on the central fixation circle throughout the experiment.

Depth results: All observers exhibited strong retinal tuning for the depth ordering aftereffect, with tightly peaked tuning curves centered at the fovea (figure 2.3 top). As a rough probe for gaze modulation, in another experiment we had observers adapt to a stimulus in the fovea, and then saccade to a new location before presentation of the test pattern in one of five locations around the new fixation point. This showed strong retinal tuning as well, and the size of the measured effect at the fovea after refixation was diminished (figure 2.3 top, gray line).

Color stimuli: Observers began each experimental session by adapting to the red stimulus at the center of the screen for a prolonged 20s. At the beginning of each trial, that stimuli would reappear for 5s, then disappear and 0.5s later, a test stimuli would appear at one of five locations (6.9° left, 3.5° left, 0°, 3.5° right, 6.9° right). The test stimuli would be visible on the screen for 0.7s.

Color task: Observers were asked to judge, in a 2 AFC task, whether the test stimulus appeared reddish or blue greenish. They were instructed to hold their gaze on the central fixation square throughout the experiment.

Color results: All observers exhibited strong retinal tuning for the color aftereffect, with tightly peaked tuning curves centered at the fovea (figure 2.3 bottom).

Experiment 2

To quantify gaze modulation of the DOAE, observers alternated gaze direction during adaptation between a single stimulus (horizontal field-in-front) at one position (right) and a fixation point at the other (left). The strength of the aftereffect (vertical field-in-front) was then tested at the two familiar locations and a neutral location (center).

Depth stimuli: Observers began each experimental session with an adaptive period of gaze alternations between the horizontal-in-front stimulus (with fixation circle) 9.7° to the right of center, and the fixation circle alone 9.7° to the left of center (3s right, 3s left, 10 alternations). Only one stimulus (horizontal-in-front with fixation circle, or fixation circle alone) was visible at a time. At the beginning of each trial, observers underwent a shorter period of gaze alternating adaptation (1.5s right, 1.5s left, 6 alternations), after which the fixation circle appeared at the center of the screen for 0.5s (prompting a saccade). The fixation circle then moved to one of three locations (9.7° right, 0°, 9.7° left) for 0.5s before the test stimulus appeared at that location. The test stimulus was visible on the screen for 1s.

Depth task: Observers were asked to judge, in a 2 AFC task, whether the horizontal bar of the test stimulus appeared in front of or behind the vertical bar. They were instructed to follow the fixation circle and keep it fixated throughout the experiment.

Depth results: All observers showed significant attenuation of the effect at the non-adapted location (on the left; the average effect at this location was only 45% as large as in the adapted gaze direction; $p < 0.05$). The attenuation of the effect at the neutral location (in the center; nearer to the adapted location) was less than in the nonadapted direction, but still significant relative to the effect at the adapted location (the average effect at this location was 53% as large as in the adapted direction; $p < 0.001$), and not significantly different from the aftereffect at the non-adapted gaze location (figure 2.4 top).

Color stimuli: Observers began each experimental session with an adaptive period of gaze alternations between the red stimulus (with fixation square) 13.8° to the right of center, and the fixation square alone 13.8° to the left of center (4s right, 4s left, 15 alternations). Only one stimulus (red adaptation stimulus with fixation square, or fixation square alone) was visible at a time. At the beginning of each trial, observers underwent a shorter period of gaze alternations (2s right, 2s left, 5 alternations), after which the fixation square appeared at the center of the screen for 0.5s (prompting a saccade). The fixation square then moved to one of three locations (13.8° right, 0°, 13.8° left) for 0.5s before the test stimulus appeared at that location. The test stimulus was visible on the screen for 1s.

Color task: Observers were asked to judge, in a 2 AFC task, whether the test stimulus appeared reddish or blue greenish. They were instructed to follow the fixation circle and keep it fixated throughout the experiment.

Color results: As with the DOAE, observers showed a significant, but in this case much smaller, degree of attenuation of the effect at the non-adapted location (on the left; average effect was 90% as large as in the adapted gaze direction; $p < 0.05$). This degree of gaze modulation is not unlike that observed for other simple visual aftereffects previously (Nishida *et al*, 2003). The neutral location showed less attenuation, and the measured effect there was not significantly different from either adapted location (figure 2.4 bottom).

Experiment 3

Is gaze modulation strong enough to induce opposite aftereffects in different gaze directions? A positive result would be the strongest conceivable evidence for gaze-dependent modulation. In our third experiment, observers alternated gaze between two locations with opponent adapting stimuli (horizontal field in front on the right, vertical field in front on the left) before responding to a test stimulus at one of three locations (left, right or center).

Depth stimuli: Stimuli were the same as in experiment 2, except that during the adaptive periods of gaze alternation, the vertical in front stimulus (with fixation circle) appeared on the left instead of the fixation circle alone. Thus observers alternated their gaze between the horizontal in front stimulus (with fixation circle) 9.7° to the right of center, and the vertical in front stimulus (with fixation circle) 9.7° to the left of center (3s left, 3s right, 10 alternations). At the beginning of each trial, observers underwent a shorter period of gaze alternations as in Experiment 2 (shown schematically in figure 2.2).

Depth task: As before, observers were asked to judge, in a 2 AFC task, whether the horizontal bar of the test stimulus appeared in-front of or behind the vertical bar. They were instructed to follow the fixation circle and keep it fixated throughout the experiment.

Depth results: On average, observers showed opposite and significantly different DOAE at the two adapted locations (+22.7% ± 8.8 (V-in-front) on the right vs. -22.3% ± 9.22 (H-in-front) on the left; $p < 0.001$; figure 2.5 top). Thus adaptation to opponent stimuli at the same retinal position but in different gaze directions produced simultaneous, opposing aftereffects in the same retinal location.

Color stimuli: Stimuli were the same as in experiment 2, except that during the adaptive periods of gaze alternation, the blue green adaptation stimulus (with fixation square) appeared on the left instead of the fixation square alone. Thus observers alternated their gaze between the red stimulus (with fixation square) 13.8° to the right of center, and the blue green stimulus (with fixation square) 13.8° to the left of center (4s left, 4s right, 15 alternations). At the beginning of each trial, observers underwent a shorter period of gaze alternations as in Experiment 2.

Color task: As before, observers were asked to judge, in a 2 AFC task, whether the test stimulus appeared reddish or blue greenish. They were instructed to follow the fixation circle and keep it fixated throughout the experiment.

Color results: After alternating adaptation between opponent color stimuli, observers showed considerably less overall color adaptation (0.011 vs. 0.058) with greater gaze-specific effects (average difference between the magnitude of effect at the two locations: 0.0093 vs. 0.0068) than in the single stimulus case as predicted. Observers again showed significantly different effects at the two adapted locations ($p < 0.0001$), though physically opposite aftereffects were not observed (Figure 5 bottom). This may be due to net color adaptation in the fovea. If the level of gaze modulation for CAE is only moderate (as observed in experiment 2), a mismatch in the subjective intensity of the red vs. blue green adaptation stimuli may have resulted in the stronger red adaptation on the right overwhelming the blue green adaptation on the left, and producing net red adaptation in both locations. Nonetheless, the relative aftereffects at the three measured locations clearly demonstrate gaze modulation.

Discussion

Gaze alternation between opponent adaptation stimuli marks a significant departure from traditional aftereffect experimental paradigms, and may provide decisive evidence for gaze modulation as well as a more sensitive measure of spatial tuning. By adapting the same retinotopic location alternately to opposite stimuli, strictly retinotopic adaptation should average out, leaving primarily adaptation to effects that exist in non-retinal coordinate systems (such as head-centered or environmental systems). While it is impossible to insure that the competition between opposite stimuli indeed results in complete cancellation, at the simplest levels of visual processing the net adaptation should be very small, and even with net retinal adaptation, observing simultaneous opposite aftereffects would be impossible without gaze modulation. Rather than measuring the combination of retinotopic and gaze-dependent adaptation under conditions of varied gaze, this paradigm seeks to measure the gaze-dependent component directly.

In all of these experiments, the strength of the aftereffect was measured to be greatest at the adapted gaze location and diminished at other gaze locations. In the single-sided alternation paradigm, it is unlikely that retinal mismatch of adaptation and test stimuli induced by gaze shift could have accounted for any of the modulation observed, since all trials involved multiple saccades prior to testing and should have been equally susceptible to mismatch effects. By alternating gaze directions (but only one stimulus) throughout adaptation, observers in our experiments adapted to both gaze locations prior to each test. Since stimuli were always presented at the same retinal location, it would be difficult to attribute these findings to anything other than gaze modulation. It seems likely that humans share the gaze modulation observed in V1, V4, and MST of monkeys, and that perception reflects this neural modulation.

In the past, aftereffects have been shown to be contingent on a variety of visual attributes—orientation contingent on color (McCullough, 1965), motion on color (Favreau, Emerson, and Corballis, 1972), motion on depth (Nawrot and Blake, 1989), and depth on motion

(Regan and Beverley, 1972; Anstis and Harris, 1974). All of this suggests that the presence of these aftereffects depends on more than simply the retinal location of visual stimuli and exists within some more complicated framework. While we can postulate how the contingent relationships between various features tie into issues of feature binding, gaze dependence serves a more obvious purpose. It may provide a critical basis for the transformation from retinal to body-centered, and eventually environmental coordinates.

Physiological studies have demonstrated the existence of visually responsive neurons that adjust their gain in a gaze-dependent manner. Neurons of this type could provide a mechanism for the gaze modulation of visual aftereffects. Consider a population of color-tuned neurons (some responding to red, some responding to green) with receptive fields centered in the fovea and gain fields (the locations at which the gain of its tuning curve is the greatest) centered at the left adaptation location. The balance of output from this population of cells will represent color balance on the red-green axis. A red adapting stimulus presented at the left location will cause a large increase in firing of red-tuned cells, whereas a green stimuli presented at the right adaptation location will evoke only a small increase in firing rate in the green-tuned cells (lower gain). The red-tuned cells will adapt out their responses to a greater degree than the green-tuned cells and, after adaptation, a neutral stimuli (equal parts red and green) presented at the left adaptation location will evoke a lesser response from the adapted red cells than the less adapted green cells. The neutral stimulus will appear green. Conversely, a similar population of cells with their gain fields centered at the right adaptation location, presented with the same adaptation stimuli (red-left, green-right), will represent a neutral stimulus on the right as being more red.

If we presume that perception is indeed reflective of underlying neural modulation such as that found in physiological studies, what does this imply about gaze modulation and possibly coordinate representation in the visual system? As mentioned above, the CAE is not subject to interocular transfer, localizing it very early in the visual pathways. It seems likely that the CAE is due to a reduction in the response of color selective neurons of V1 (or earlier) following adaptation. The results of these experiments would then suggest that very early on there occurs some integration of gaze information which modulates, albeit modestly, the response properties of color selective neurons. This is in line with recent findings of modest gaze modulation (approximately15%) of a variety of aftereffect (motion, tilt, size) that also likely occur as early as V1 or V2 (Nishida *et al*, 2003).

The depth ordering aftereffect probably involves a higher degree of visual processing, but could also occur as early as V1. In physiological studies in monkeys, stereoptic, depth-tuned neurons were found in V1, V2 , V3, MT and MST (Poggio, 1994). The much greater degree of gaze-dependent modulation seen here with depth is not entirely surprising considering the degree of correlation and integration of position information at higher cortical levels that is involved in stereopsis.

The evidence here indicated that gaze direction modulates perception of aftereffects in both color and depth. These effects likely occur early in human visual processing and may be the perceptual correlates of the modulation of neuronal response in human visual cortex.

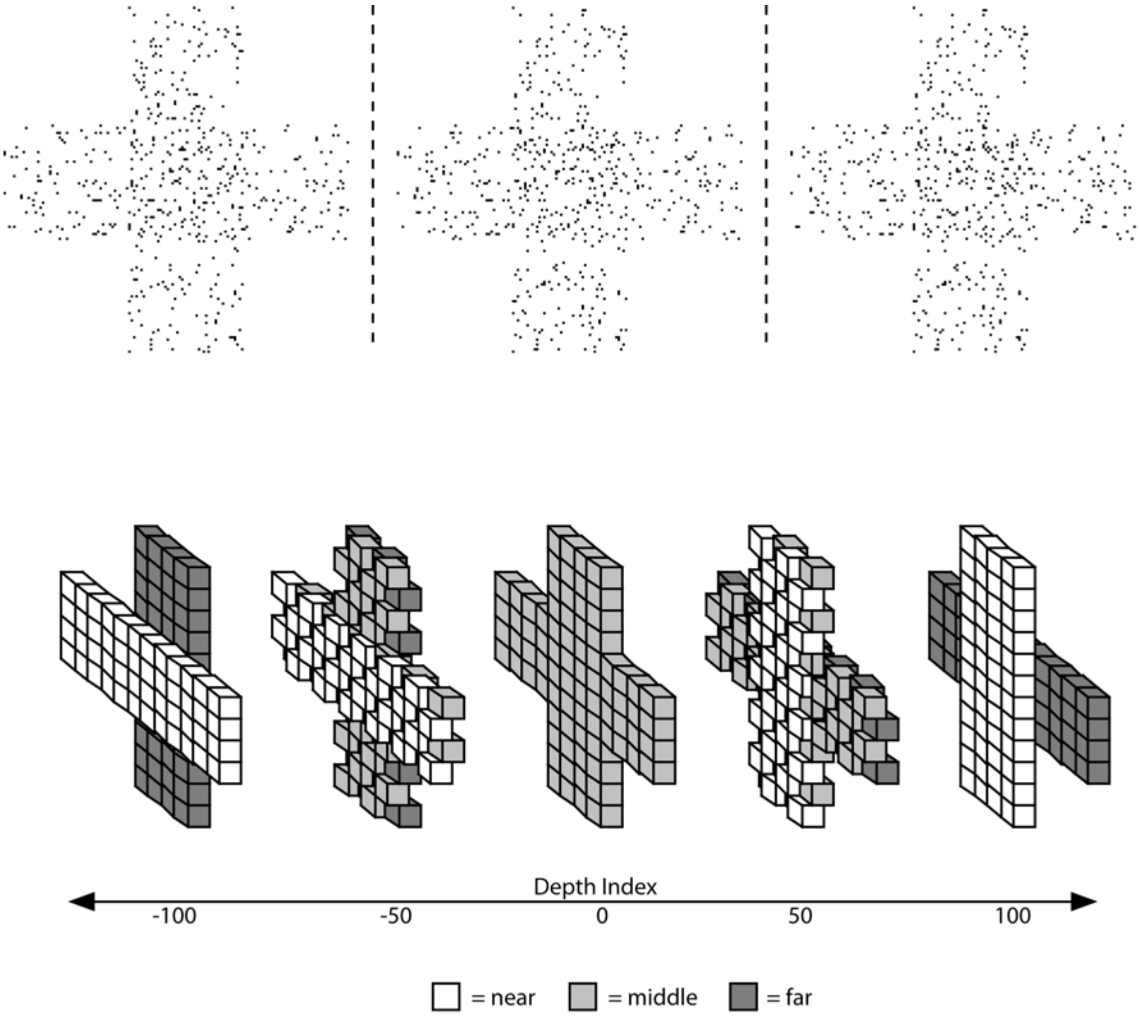


Figure 2.1: Stimuli for the Depth Ordering Aftereffect. (Top) Cover the image on the right and fuse the image on the left with image in the center. One of the bars should appear nearer than the other. Now cover the image on the left and fuse the image on the right with the image in the center. The bars should appear to have reversed their depth ordering. This is provided as a demonstration. For the actual experiments, images were presented in red-blue anaglyph. (Bottom) The depth index specifies what percentage of random dots specifies a particular depth orientation. A negative index corresponds to the vertical bar appearing farther, a positive index means the vertical bar is nearer.

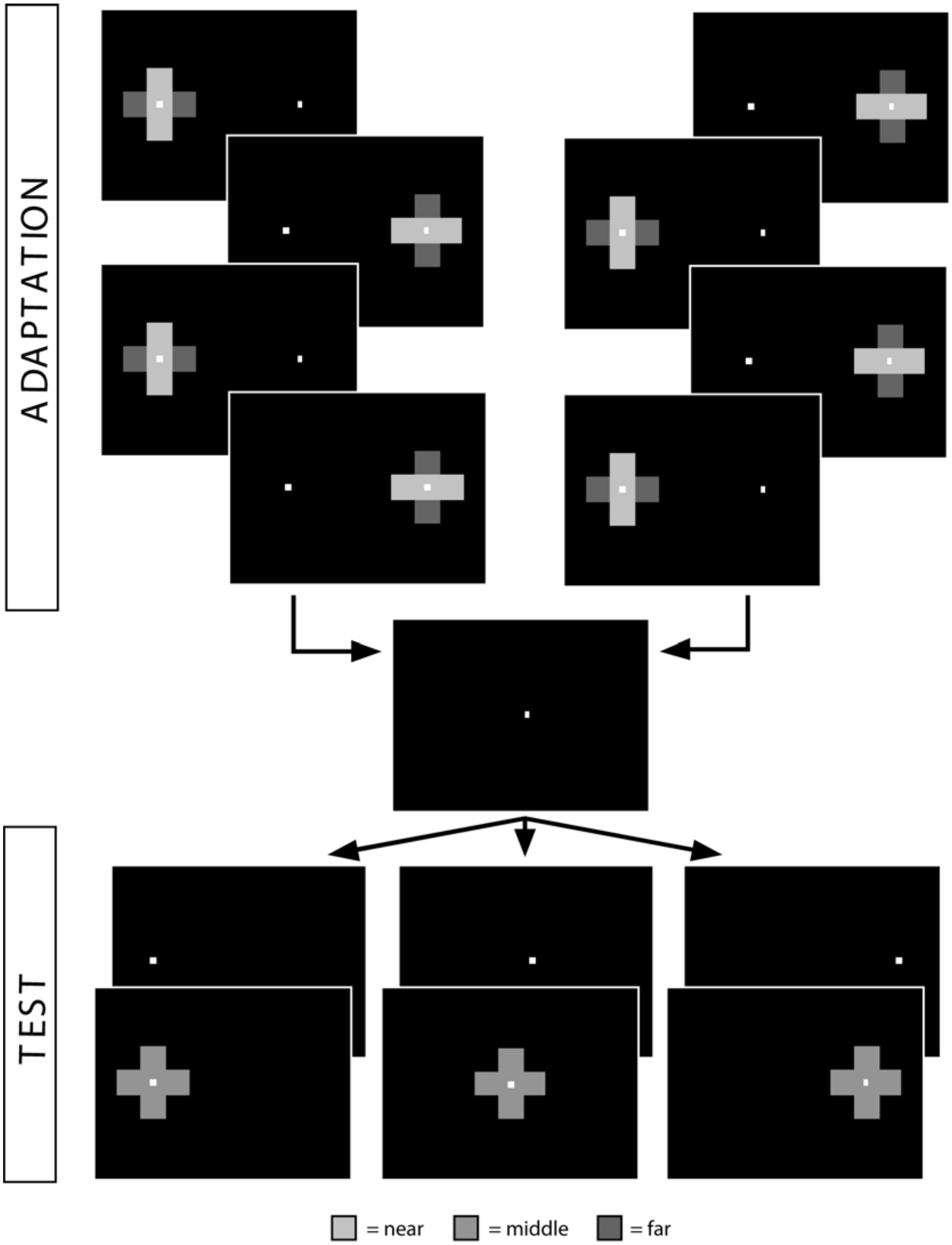


Figure 2.2: The experimental paradigm. The stimulus begins on either the right or the left. The subject alternates gaze location, following the stimulus throughout adaptation. Before testing, observers saccade to a neutral location at the center, and are then tested at one of three locations.

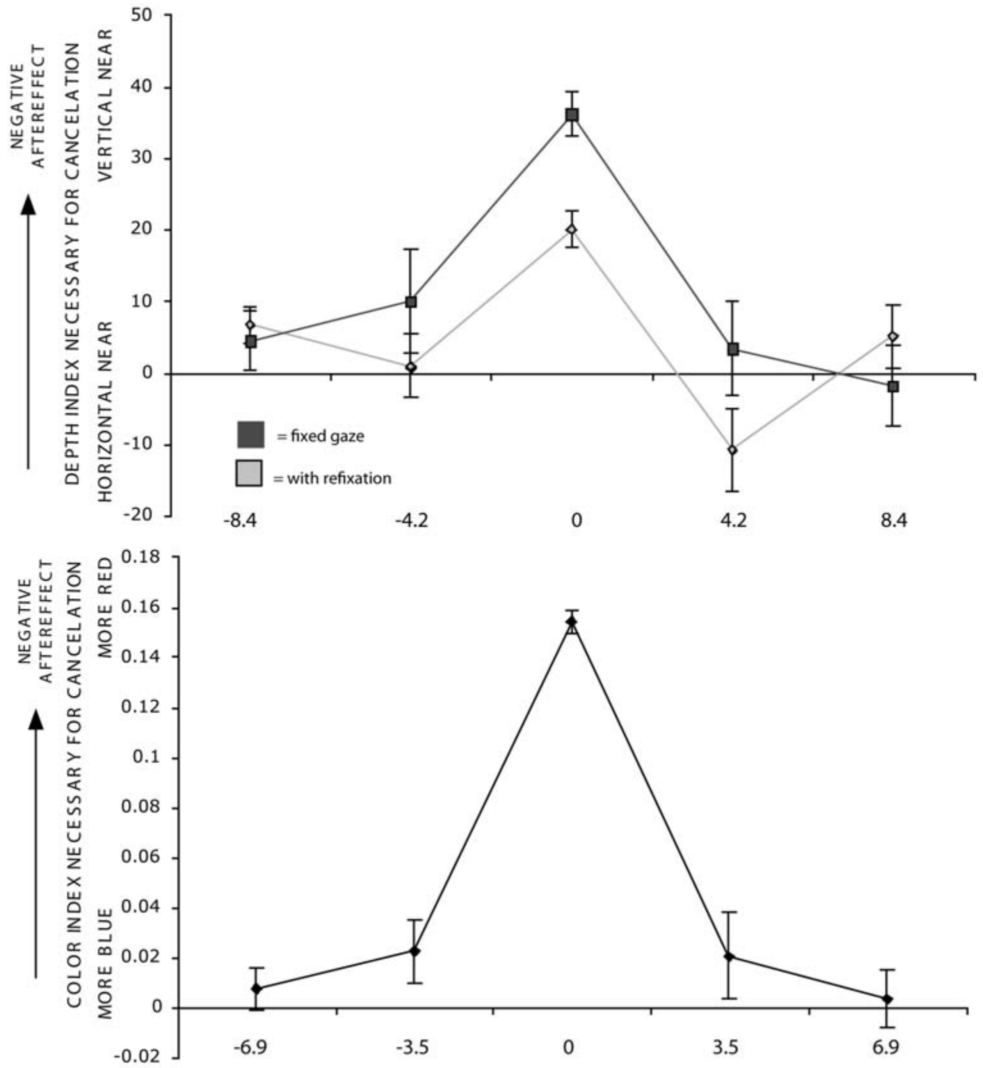


Figure 2.3: <u>Retinal Tuning of the Color and Depth Ordering Aftereffects</u>. (Top) Observers adapted to a horizontal-in-front stimulus in the fovea after which they maintained fixation (black) or shifted to a new fixation point 12.6° to the right or left (gray), and a test stimulus was presented at one of five retinal locations. (Bottom) Observers adapted to a red stimulus in the fovea after which a test stimulus was presented at one of five retinal locations. For both effects, the magnitude was greatest at the adapted retinal location and dropped off quickly to the left and right. Error bars represent standard error of the mean for the six reversal points used in determining the point of subjective equality.

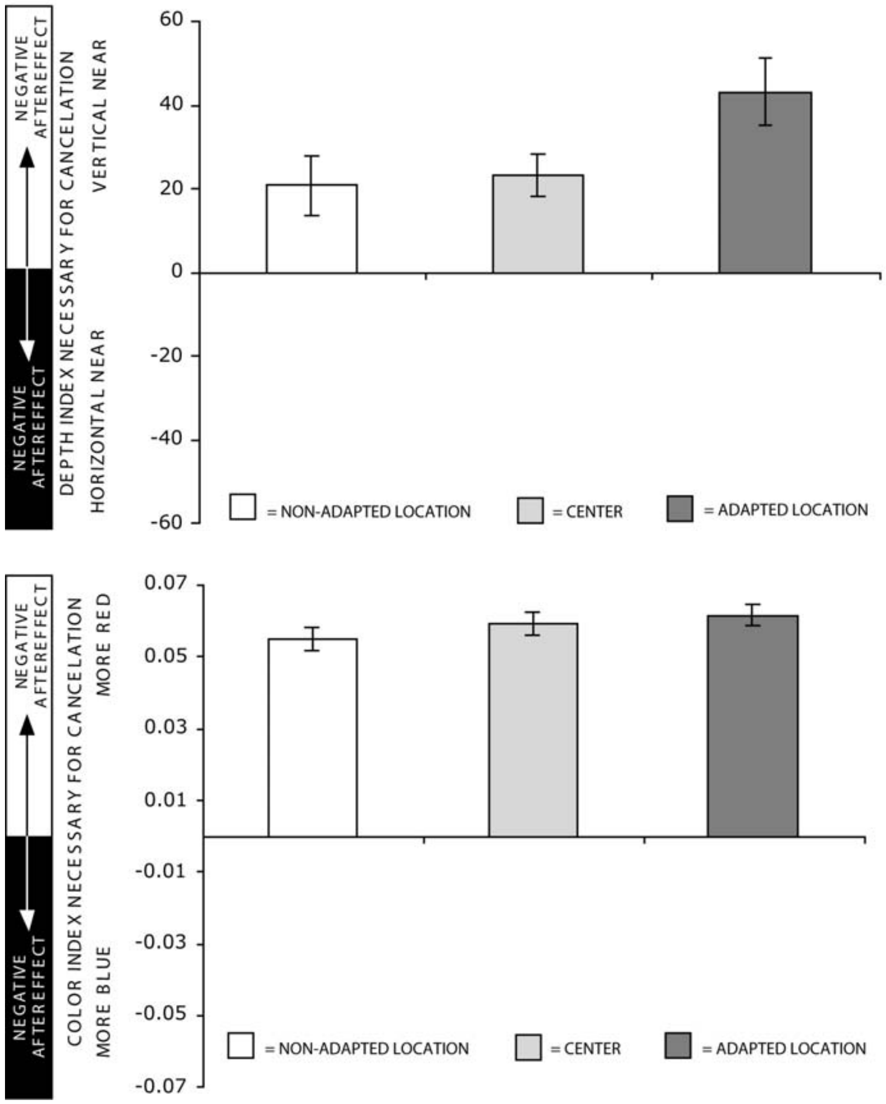


Figure 2.4: Gaze modulation of aftereffects. (Top) Observers adapted to a vertical-in-front stimulus in the fovea on the right side of the screen. (Bottom) Observers adapted to a red adaptation stimulus in the fovea on the right side of the screen. In both experiments, after shifting gaze to the center of the screen, and then to one of three testing locations (left, center, right) observers were presented with a test stimulus. With both effects, observers showed the strongest aftereffect at the adapted gaze location and diminished effects at other gaze locations. Error bars represent standard error of the mean for the six reversal points used in determining the point of subjective equality.

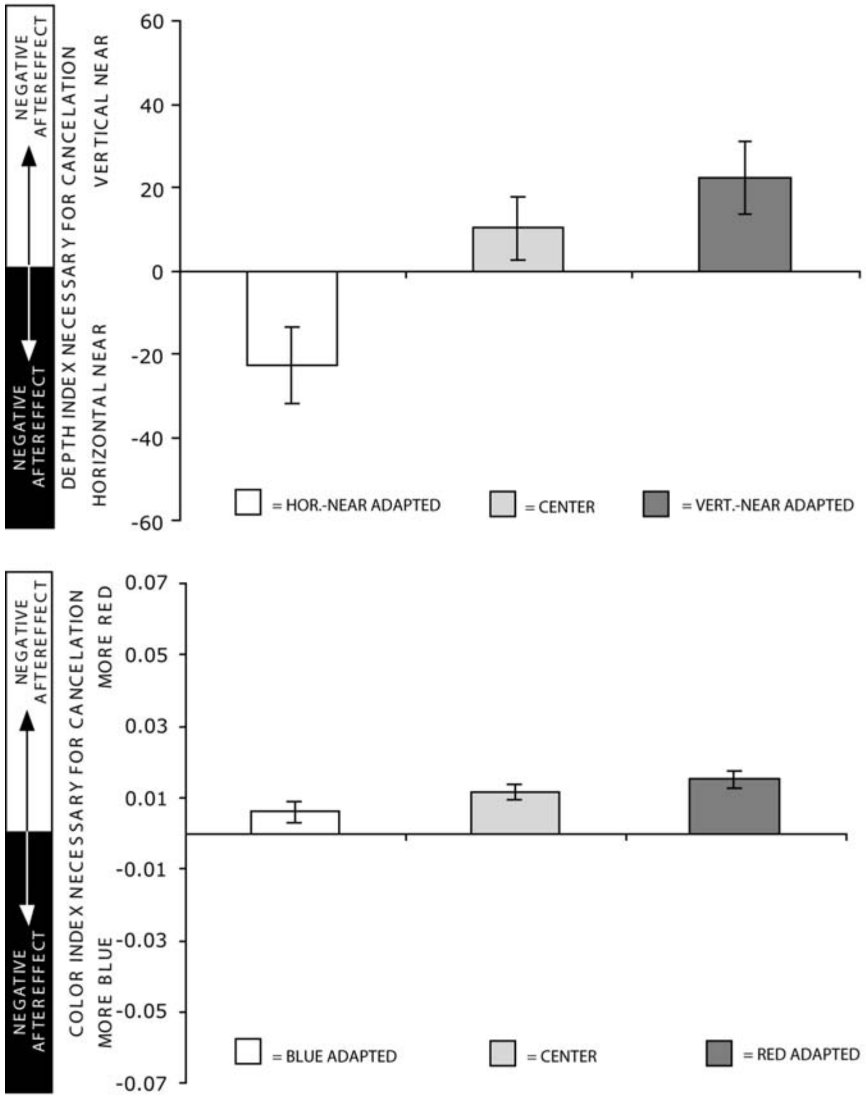


Figure 2.5: Gaze Dependent Aftereffects. (Top) Observers alternated adaptation between a vertical-in-front stimulus on the right and a horizontal-in-front stimulus on the left. (Bottom) Observers alternated adaptation between a red stimulus on the right and a blue stimulus on the left. In both experiments, subsequent to adaptation, gaze was shifted to the center of the screen and then to one of three testing locations (left, center, right) and a test stimulus appeared. Error bars represent standard error of the mean for the six reversal points used in determining the point of subjective equality.

Chapter 3

Study 2: Cyclopean Flash-Lag Illusion

(Previously published as Nieman *et al.* 2006)

Abstract

Possible physiological mechanisms to explain the flash-lag effect, in which subjects perceive a flashed item that is colocalized with a moving item as trailing behind the moving item, have been found within the retina of lower species, and in the motor pathways of humans. Here, we demonstrate flash-lag employing "second-order" moving and flashed stimuli, defined solely by their binocular disparity, to circumvent any possible "early" contributions to the effect. A significant flash-lag effect was measured with cyclopean stimuli composed entirely of correlated random dot patterns. When the disparity-defined moving stimulus was replaced with a luminance-defined one, potentially engaging retinal mechanisms, the magnitude of the measured effect showed no significant change. Thus, in primates, though retinal mechanisms may contribute, flash-lag must be explained through cortical processes.

Introduction

Neurophysiological findings on the retina in lower species have provided a possible mechanism to explain flash-lag. Investigating the frog retina, Barlow (1953) noted that the greatest "off" response of ganglion cells to a moving fly's retinal image occurs closer to the fly's future position. More recently, Berry, *et al*. (1999) reported similar anticipatory responses to motion in the retinal ganglion cells of the rabbit and the salamander. When retinal ganglion cell responses to moving and flashed stimuli were measured, it was observed that cell responses to moving stimuli were spatially shifted in the direction of motion, analogous to the forward shift of moving stimuli in relation to flashed stimuli shown in psychophysical experiments in humans (Nijhawan, 1994). The proposed retinal mechanism is based on well-known properties of retinal ganglion cells such as spatial-pooling, biphasic responses and contrast gain control (Berry *et al.* 1999). Whether similar mechanisms operate throughout the visual system, and produce perceptual phenomena like the flash-lag effect in humans, remains an open question.

Given that the flash-lag effect has been demonstrated under diverse stimuli conditions and manipulations (for example, it has been demonstrated in experiments employing non-moving, changing stimuli; coined "generalized flash lag"), retinal mechanisms driven by motion alone would appear to be inadequate (Sheth, Nijhawan, and Shimojo, 2000). Furthermore the flash-lag effect was recently demonstrated in an experiment in which subjects moved their hand in the dark and judged the position of a flash relative to the felt position of their moving hand (Nijhawan and Kirschfeld, 2003). Subjects perceived the position of the flash, presented in spatial alignment with their moving hand, as lagging relative to the felt position of their invisible hand. This “motor flash-lag” suggests that non-visual mechanisms of the central nervous system that monitor limb positions can and do contribute to the flash-lag effect. Given that flash-lag is observed both at the “early” retinal level and the motor level, and the presence of generalized flash lag across visual attributes, one may ask if mechanisms contributing to this effect are also present at the intermediate levels, between the retina and motor cortical areas.

To examine this issue, we investigate the visual flash-lag by employing a well-known technique that engages neither retinal nor overt motor processes. Human subjects perceive motion either when moving objects are distinguished from the background in terms of their mean luminance or color (first-order properties), or in terms of texture or binocular disparity in the absence of accompanying luminance variations (second-order properties) (Cavanagh and Mather, 1989). We presented moving and flashed items, isoluminant with the background, defined only by binocular-disparity in correlated random dot patterns, to measure the strength of cortical visual compensation. We compared this with the FLE observed under the same background conditions, but with solid gray, monocularly detectable stimuli. Then, in a ‘mixed’ condition, by rendering the moving stimulus as luminance-defined we were able to gauge the additional contribution, if any, of retinal mechanisms to the flash-lag effect.

Experiments

A motion sensor activated only by luminance differences (Reichardt, 1957) would not detect any net motion for second-order stimuli whose average luminance is the same as the background. Thus, if mechanisms based in the retina are responsible for flash-lag, then purely cyclopean stimuli would not elicit the effect. In these experiments we demonstrate the existence of flash-lag for stimuli that can be detected only via correlation of images from both eyes, and compare the magnitude of the effect to conditions in which either the moving target, the flashed target, or both targets are luminance defined.

Condition 1

Participants: Two authors and two naive observers with normal or corrected to normal vision were used. The same observers were used in all experiments.

Apparatus: All experiments were performed on a Macintosh computer running MATLAB (MathWorks Inc., Natick, MA) and the Psychophysics Toolbox extensions (Brainard, 1997). Stimuli were displayed on a Nokia 445xi CRT screen (60 Hz refresh; 40 cm × 30 cm) which observers viewed through a mirror haploscope. The haploscope was positioned such that the observer’s right eye viewed only the right half of the screen while the left eye viewed only the left half of the screen, and then adjusted so that the two images comfortably fused (figure 3.1). The total viewing distance was 57cm (1 cm on screen = 1° of visual angle).

Stimuli: The displays for the left and right halves of the screen were constructed of correlated dynamic noise consisting of equal numbers of 0.13° × 0.13° black (0.00 cd/m^2) and white (45.9 cd/m^2) squares. Noise frames changed every 35.9 ms. A 2.0° × 2.0° white fixation cross, 0.25° wide with a 0.25° black border was visible at the center of the right and left displays throughout the experiment. The fixation cross was displayed with a crossed disparity of 0.13° (the cross was shifted to the right by 0.063° in the left eye display and to the left by 0.063° in the right eye display; for an observer with inter-ocular distance of 3.5 cm, this crossed disparity corresponds to an object placed 1 cm in front of the display). In each trial, a smoothly moving target (6.68° × 2.67°) would travel horizontally across the display, 4.0° below the fixation cross, at one of two speeds (4.65° or 9.30° sec^{-1}). These moving stimuli were defined solely by binocular disparity (Julesz, 1971; Smith and Scott-Samuel, 1998). While the moving target was at one of seven positions near the center of the display, a flashed target of the same size (6.68° × 2.67°) appeared centered 4.0° above the fixation cross. Both flashed and smoothly moving stimuli were composed of correlated dynamic noise presented at a crossed disparity of 0.13° with respect to the background (at the same apparent depth as the fixation cross, about 1cm in front of the display). All frames of animation lasted for 71.8 milliseconds (2 frames of noise), thus the position of the smoothly moving target updated every 71.8 ms, and the flashed target was present for 71.8 ms. Despite this relatively slow animation rate, observers reported that the motion of the translating stimulus appeared smooth under these conditions. In order to ensure that the stereoscopic stimuli were undetectable monocularly, prior to the start of experimental blocks, observers were asked to view the a few trials of the experimental display with either the left or the right eye alone. Observers reported seeing only the fixation cross.

Task: Observers were asked to judge, in a two-alternative forced-choice task, whether the flashed target appeared to the left or right of the continuously moving target. Each observer was presented both directions of motion in two blocks for a total of 420 trials. The velocity and the position of the flashed bar were randomly selected without replacement for each trial.

Analysis: The observers' responses under each speed-direction combination were fitted with a probit curve to determine the point of subjective equality, measured in milliseconds by which the flash preceded the moving bar's central alignment. A 3-way ANOVA (observer, speed, direction) was performed to determine the relevance of speed and direction in the magnitude of the effect. Both were found to be uncorrelated with the magnitude of the effect in the time domain. Based on this, observer responses across both speeds and both directions were pooled within observer and fitted again with a probit function to determine each observer's average measured effect.

Results: These conditions produced a strong flash-lag effect in all subjects at both velocities and in both directions. The moving bar was perceived as aligned with the flashed bar when the flash was presented on average 163.4 ± 7.5 ms before the two bars were actually aligned (figure 3.2a). Thus, 'cyclopean' stimuli that cannot be detected monocularly produce a reliable flash-lag effect implicating cortical mechanisms located beyond the computation of binocular disparity. A high-level 'feature-tracking' mechanism may be responsible for detection of motion in disparity-defined stimuli (Lu and Sperling, 1995a, 1995b; Cavanagh, 1992). These results show that mechanisms that are qualitatively different from those invoked previously, may contribute significantly to the flash-lag effect in humans (Berry *et al*., 1999).

Condition 2

For comparison with more traditional flash-lag experiments, observers were presented with homogeneous gray (8.2 cd/m^2) moving and flashed targets. All other conditions of the experiment were the same as in the first condition (a background of correlated dynamic noise, viewing through the haploscope). The same four observers participated.

Results: A strong, consistent flash-lag was observed. The moving bar was perceived as aligned with the flashed bar when the flash was presented on average 149.2 ± 13.2 ms before the two bars were actually aligned (figure 3.2b). While the magnitude of the effect measured here is smaller than in the cyclopean case, the difference modest and is not statistically significant ($p = 0.2253$; $df = 3$ in within-subject paired t-test).

Condition 3

In order to ascertain any additional contribution of retinally based mechanisms of motion extrapolation, observers were presented with a homogeneous gray (8.2 cd/m^2), and hence monocularly detectable, moving target, but a cyclopean (disparity defined) flashed target. All other conditions of the experiment were the same as in the first condition. The same four observers participated.

Results: Again, observers exhibited a strong flash-lag effect. The moving bar was perceived as aligned with the flashed bar when the flash was presented on average 165.8 ± 19.1 ms before the two bars were actually aligned (figure 3.2c). This represents an insignificant change from the fully cyclopean condition ($p = 0.693$; $df = 3$ in within-subject paired t-test). Thus it seems, under these circumstances, the contribution of retinally based mechanisms of motion extrapolation is modest at best.

Condition 4

Some theories of flash-lag center around the differential latency associated with perception of the transient event with respect to perception of the continuously moving, and hence predictable, target. In order to facilitate the detection of the flash observers were presented with a homogenous gray (8.2 cd/m^2), and hence monocularly detectable, flashed target, but a cyclopean moving target. All other conditions of the experiment were the same as in the first condition. The same four observers participated.

Results: In this case, the size of the effect was diminished, though still robust. The moving bar was perceived as aligned with the flashed bar when the flash was presented on average 125.0 ± 10.0 ms before the two bars were actually aligned (figure 3.2d). This represents a significant 23% decrease from the fully cyclopean condition ($p = 0.0118$; $df = 3$ in within-subject paired t-test). This is consistent with differential latency explanations for flash-lag.

Discussion

Our results are consistent with previous investigations (Fu, Shen, and Dan, 2001) employing stimuli lacking luminance boundaries, suggesting that the human visual cortex is largely responsible for the flash-lag effect. Furthermore, our results suggest that processes peripheral to layer 4 of the primary visual cortex (V1)—the first possible locus in the nervous system where information from the two eyes could potentially converge on the same neurons—make only a modest contribution to visual flash-lag. Berry *et al*. (1999) invoked spatial pooling in the retinal ganglion cells, the fact that a neuron fires to stimuli presented anywhere over a large portion of the visual field, as one of the factors contributing to the flash-lag effect. By virtue of spatial pooling, a retinal ganglion cell begins firing as soon as a moving stimulus impinges on the periphery of a cell's receptive field, thus effectively shifting the stimulus towards the center of the cell's receptive field. In primates the typically measured size of the receptive fields of typical V1 cortical cells representing the fovea, where a large flash-lag effect has been measured, are too small for spatial pooling to completely account for flash-lag. While it has been shown that stimulus conditions can significantly affect the measured size of receptive fields in primates (Kapadia, Westheimer, and Gilbert, 1999)—specifically, both low-contrast stimuli and textured backgrounds tend to increase the observed size of receptive fields—there are other indications that mechanisms qualitatively different from spatial pooling may need to be invoked.

Sheth, Nijhawan, and Shimojo (2000) demonstrated that FLE is not limited to the spatial domain, but can be applied to a variety of feature changes (color, luminance, etc.). A color spot gradually changing from green to red appears redder than a flashed spot of simultaneously identical color. This feature-change flash-lag cannot be explained by spatial-pooling. Our results suggest that even within the spatial domain, the contribution of spatial-pooling to flash-lag is of only limited significance. Whether our continuously moving target was defined by luminance or the disparity of coordinated random dots had no affect on the size of the measured FLE (condition

3). Whatever mechanisms are responsible for FLE under these conditions, engaging luminance detectors for the moving target does not change the observed phenomenon.

Differential latency accounts of flash-lag suggest that there is some facilitation of perception for the changing target which allows it to occur faster than perception of the flashed target (Purushothaman *et al*., 1998; Whitney and Murakami, 1998). The observed lag is a measure of the time difference between the two percepts. By this account, we would expect that changing the flashed target to be more easily detectable should result in a smaller FLE (less of a difference in the time to process its perception). Purushothaman *et al*. (1998) demonstrated this for luminance detectability differences near threshold and our results are consistent with their findings—changing a disparity-defined flash for a luminance-defined one resulted in a significantly smaller effect (condition 4).

Our results, however, point to more fundamental differences between ‘early’ and ‘late’ mechanisms. It is known that signals originating in higher cortical areas can strongly influence cell activity in a given location of layer 4 of V1 through descending signals, as shown in tests in which afferent input to the given layer 4 location was removed (Mignard and Malpeli, 1991). Researchers have identified two types of motion processes in humans; one based on low-level motion detectors and the other based on signals arising in higher cortical areas (Braddick, 1980). There is a debate concerning which processes underlie motion perception in disparity-defined stimuli, and as to whether stereo-motion is based on specialized motion detectors or on high-level ‘feature-tracking’ (Patterson, 1999). Despite this debate, researchers agree that signals originating in ‘higher’ cortical areas play an important role in the perception of motion in disparity-defined stimuli, with suggestions that movement of voluntary attention may contribute generally to motion perception in stimuli lacking luminance or color boundaries (Cavanagh, 1992; Lu and Sperling, 1995a, 1995b). Cavanagh (1992) made this hypothesis explicit by suggesting that voluntary internal signals that move the focus of attention contribute to motion perception in stimuli lacking luminance boundaries.

While our results do not directly address the role of attention in flash-lag, it should be mentioned that the task in our experiments with disparity-defined stimuli is extremely demanding. Consistently perceiving the position of the moving stimulus amid dynamic random noise requires that attention remains focused upon it. Any attention-generated internal signal that contributes to the perception of motion is likely very strong under these conditions. Also, with attentional resources so committed to the moving target, any delay associated with bringing the flashed target into awareness would likely be increased. If such an attentional delay is ultimately responsible for FLE (Baldo and Klein, 1995), this may explain the unusually large magnitude of the effect we observed, though this is certainly not direct evidence for such a theory.

Cyclopean flash-lag does, however, provide further support for the notion that FLE is likely the result of ‘higher’ cortical processes, and adds to the wealth of phenomenology that must be explained by any existing theory.

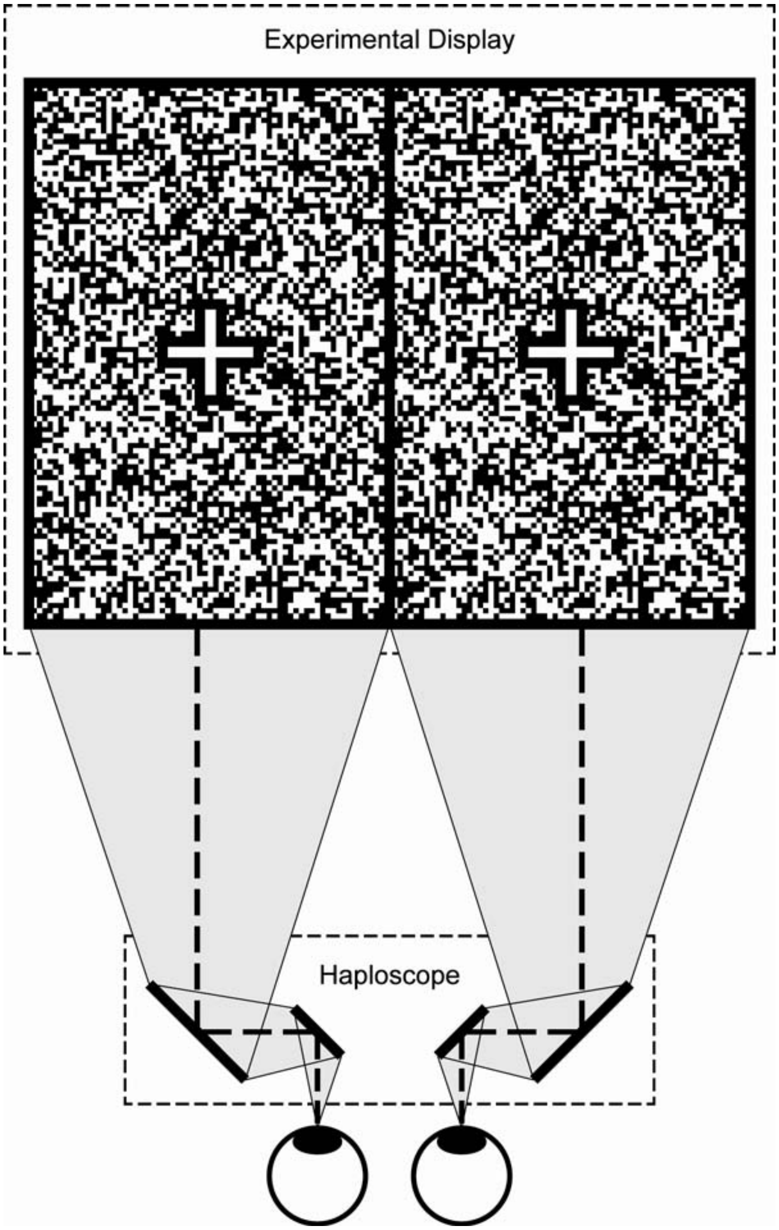


Figure 3.1: The Experimental Apparatus. Observers viewed the computer display through a mirror haploscope, adjusted such that the images received by the left and right eyes appeared comfortably fused.

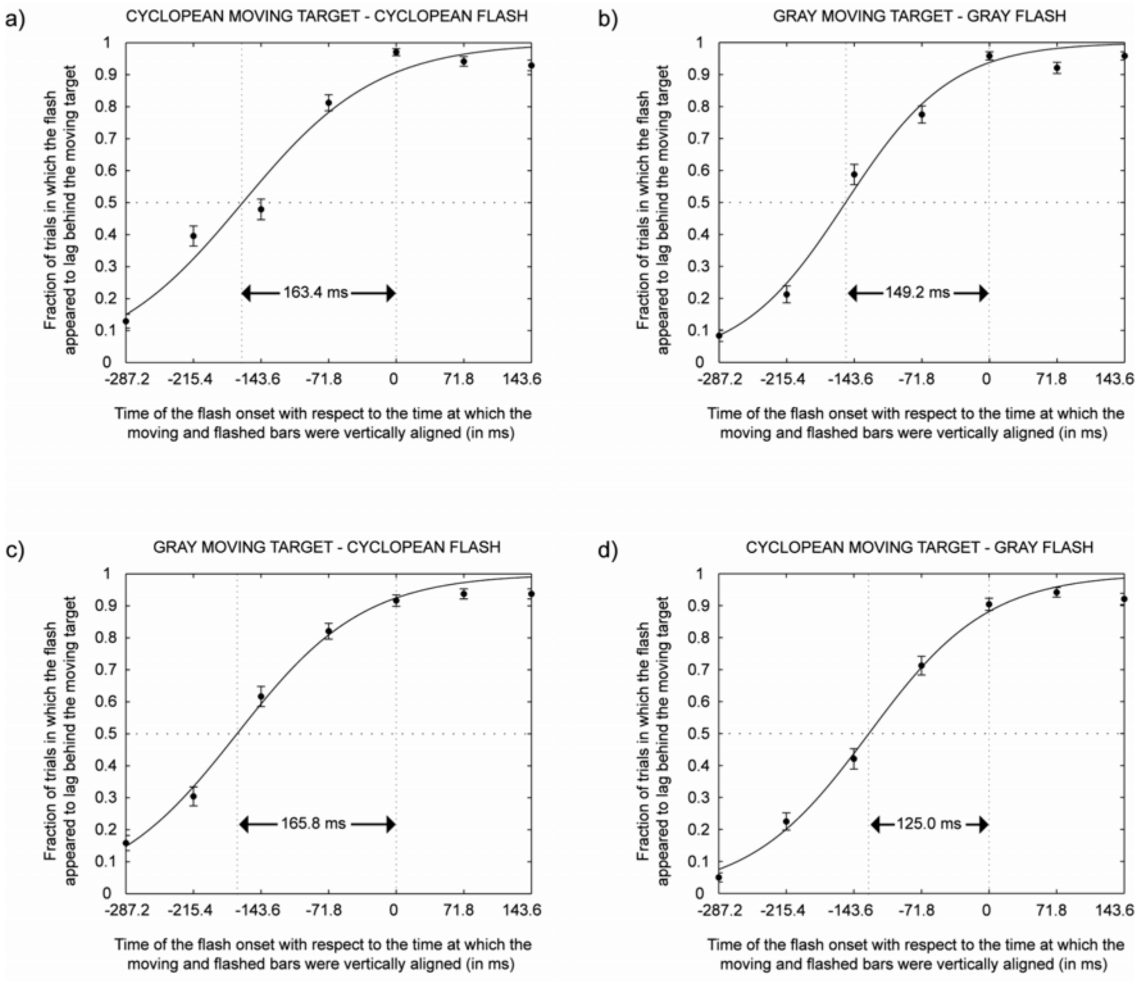


Figure 3.2: <u>The Psychometric Data</u> plots (n = 4) for: (a) completely cyclopean moving and flashed targets, (b) monocularly visible gray moving and flashed targets, (c) monocularly visible gray moving target and cyclopean flashed target, and (d) cyclopean moving target and monocularly visible gray target. The horizontal axis gives the time by which the flashed bar was presented in advance of the moving bar. The value of zero indicates that the flashed bar was spatially aligned with the simultaneous frame of the moving bar. The vertical axis gives the fraction of the total number of responses, at a given spatial lead of the flashed bar, for which the subjects reported the flashed bar as lagging the moving bar. The dashed horizontal line is taken as the point at which the moving and the flashed bars appeared aligned to the subjects.

Chapter 4

Study 3: The Turn-Point Phantom

Abstract

Abrupt changes in a moving object's direction create inconsistencies between the perception of motion and position that may result in positional errors. Previous reports of mislocalization associated with moving objects (Freyd and Finke, 1984; Nijhawan, 1994; Thornton, 2002; Fröhlich, 1923) have generally focused on a single trajectory or direction reversal. The turn-point phantom effect (TPE) is a new form of mislocalization in which the position of an abrupt orthogonal direction change is mislocalized to a point outside of the object's trajectory. Though the direction of the effect is consistent with the idea independent pre- and post-turn percepts, ablating the pre-turn trajectory quells the effect. Attempts to change the temporal perception of the turn with an additional transient event to "capture" the moment of the turn suggest that the perceived timing of the direction change is central to the effect. TPE represents a new form of motion mislocalization that may provide evidence for postdictive visual awareness, and the predominance of the motion over positional information.

Introduction

The visual perception of motion and position are critical to our day-to-day functioning. Catching a ball, driving a car, or simply walking around without continually bumping into every other moving thing in this world requires timely and accurate information about both where things are currently located (position) and where they are going (motion). While imperfect, under normal circumstances the information we receive or construct is sufficient for the activities of everyday life. Importantly, no inconsistencies exist in our minds between our percepts of where an object was moments ago and in what direction and with what speed it might be heading. One possibility may be that we do not actively and routinely keep track of isolated positions in space occupied by objects (Kanai, Sheth, and Shimojo, 2004). However, in cases where demands are made on the awareness of both motion and position, interactions may occur between the two. There exists an assortment of consistent, known errors in position perception under the effect of motion: representational momentum (Freyd and Finke, 1984), flash-lag (Nijhawan, 1994), onset repulsion (Thornton, 2002; Actis-Grosso and Stucchi, 2003), and Fröhlich (Fröhlich, 1923; Müsseler and Aschersleben, 1998). These studies have shown mislocalizations associated with the origination (onset repulsion and Fröhlich) and termination (representational momentum) of moving targets, as well as with localization in time and space with respect to other transient events (flash-lag).

We offer a new type of motion localization error associated with an change in direction—the turn-point phantom effect (TPE)—wherein a moving target that abruptly changes direction is mislocalized backwards along its subsequent motion trajectory. This new effect cannot trivially be explained by existing effects or their proposed mechanisms, though attempts to do so raise questions about the nature of motion continuity in object processing and the online nature of motion and position processing.

The overall motion trajectory of the target in this effect can be thought of in two distinct ways: 1) as a continuous, elbow-shaped path that experiences a transient change of feature (direction), and 2) as the juxtaposition of two distinct, orthogonal trajectories, the first terminating where the second originates. If the target is represented in the first way, we might imagine something similar to the flash-lag effect (FLE), with the temporal perception of the turn lagging behind the position of the target (Whitney and Murakami, 1998), and the turn-point perceived forward, localized somewhere along the post-turn trajectory.

Some flash-lag results (Moore and Enns, 2004) suggest that abrupt changes in object features disrupt and potentially eliminate FLE. Moore and Enns argue that large feature changes cause the generation of a second object representation, and if the abrupt direction change is sufficient in this regard, we may expect our trajectory to behave as two distinct objects that only happen to coincide in the termination of the first and the origination of the second at the position of the turn. This is equivalent to thinking of the target in the second manner delineated above. In this case, with two distinct object representations, it is unclear which object assumes ownership of the turn-point. If the turn-point is represented as the termination of the first trajectory, we expect a contribution of representational momentum (repmo), whereby the representation of the target's position is carried forward along its path, past the actual turn-point (figure 4.1). In our case, for a target that moves down and right and then turns 90 degrees and heads up and to the right, representational momentum would push the perception of the end of the first part of the trajectory down and to the right.

If the turn-point is represented as the origination of the second trajectory, we might expect contributions from either of two onset effects that predict opposite directions: the Fröhlich effect and the onset repulsion effect (ORE). The Fröhlich effect involves the mislocalization of a fast moving target's origin forward in the direction of motion and could push the perceived position of the turn-point up and to the right (the same direction predicted for flash-lag); ORE involves the mislocalization of a moving target's origin backwards along its path of motion and could push our perceived turn-point down and to the left. Thus each effect that considers the two trajectories separately (ORE, repmo, and Fröhlich) predicts a different direction of bias in the perception of the turn-point.

The abrupt change in direction creates a problem of consistency between the perceptions of motion and position. There is a consistent bias in the perceived position of the turn-point, but in contrast to this, the motion trajectory does not appear distorted in any way. Thus, any shift in the perceived location of a moving target that is discontinuous between the pre- and post-turn paths would result in an inconsistency, not merely an inaccuracy, in the perception of the position or motion of the target.

There is ongoing debate as to the relative independence in the perception of position and motion (Nakayama, 1985; Whitney and Cavanagh, 2000), but consistency in both is critical to our functional visual experience. Position is perhaps the most important fundamental aspect of vision and motion “pervades the visual world, a circumstance that has not failed to influence the processes of evolution” (Marr, 1982).

With the turn-point mislocalization effect we seek to establish the essential inconsistency between our internal representations of motion and position that arises with the abrupt transition between two orthogonal trajectories, and explore how it may be resolved. Experiment 1 demonstrates the basic effect with either one or two targets. Experiment 2 tests the similarities to ORE under these conditions. Experiment 3 introduces another transient event (visual flash or auditory beep) to examine similarities to FLE.

Experiments

Experiment 1a

Single target. Observers watched a small, circular, diagonally moving target on a computer screen make an abrupt 90 degree turn. On each trial the target appeared in the upper left quadrant and moved down and to the right towards the center of the screen until it reached the midline when it would abruptly change direction and move up and to the right, towards the upper right corner of the screen. Observers were asked to fixate on a small dot in the center of the screen (below the turn-point of the target) and judge the position of the turn in either the horizontal direction with respect to the fixation point, or in the vertical direction with respect to nearby hash marks.

Participants. Six naive, unpaid observers with normal or corrected to normal vision were used.

Apparatus. All experiments were performed on a Windows computer connected to 19” monitor with a refresh rate of 60Hz and a resolution of 800 x 600. Viewing distance was 57cm, such that 20.7 pixels subtended 1° of visual angle. Software was scripted in Matlab (MathWorks Inc., Natick, MA) utilizing the psychophysics toolbox (Brainard, 1997). The experimental room was dimly lit (0.10 cd/m^2) and the monitor frame and experimental surroundings were visible. The apparatus was the same in all experiments.

Stimuli. On each trial, a circular target 0.87° in diameter appeared to the upper left of fixation and moved diagonally down and to the right at a 45° angle and a speed of 12.3 deg/sec for 1090 ms and then changed direction and moved up and to the right at a 45° angle and the same speed for another 1090 ms (figure 4.2a). A central fixation point 0.10 x 0.10° was constantly visible. In trial blocks where the horizontal position of the turn-point was tested, the target appeared at one of seven horizontally spaced locations such that the turn-point occurred 3.4° above the fixation point either directly aligned or 0.3, 0.6, or 0.9° to the right or left. In trial blocks in which the vertical position of the turning point was tested, small horizontal hash marks 0.10 x 3.9° were visible 3.4° above and 4.8° to the right and left of the fixation point. In these trials, the target appeared at one of seven vertically spaced locations such that the turn-point occurred horizontally centered above the fixation point and either directly aligned with or 0.3, 0.6, or 0.9° above or below the hash marks. The fixation point and all targets were presented in white (83.5 cd/m^2) against a black background (0.215 cd/m^2).

Task. Observers were familiarized with the stimulus display and response keys and then asked to judge, in a 2 AFC task, whether the turn-point of the target appeared to the left or right of fixation in horizontal test blocks, and above or below the visible hash marks in vertical test blocks. They were instructed to hold their gaze on the central fixation point throughout the experiment. Position of the target was randomized across trials in blocks of 140 (20 trials each position).

Analysis. Horizontal and vertical displacements of the perceived turn-point were probed and analyzed separately. For each experimental block, responses were fitted with a psychometric curve:

$$F(x) = 0.5 + \frac{(a+bx)}{2\sqrt{1+(a+bx)^2}}$$

by minimizing the square error. Free parameters a and b were estimated by a least-squares criterion and the point of subjective equality (PSE) was obtained as ($-a/b$). Thus each PSE represented the displacement necessary for the perceived turn-point to be horizontally aligned with the fixation point or vertically aligned with the hash marks. Positive values correspond to rightward and upward displacements (i.e. forward along the post-turn trajectory). The PSEs of individual subjects were then averaged and analyzed using a 2-tailed, unpaired t-test to determine significance.

Results. In the horizontal direction, the average perceived turn-point of the target was 17.3 ± 4.8 arcmin to the left of (behind) the actual turn-point ($t(5) = 3.63$; $p = 1.5 \times 10^{-2}$). In the vertical direction the average perceived turn-point of the target was comparable: 13.6 ± 4.3 arcmin below that actual turn-point ($t(5) = 3.20$;$p = 2.4 \times 10^{-2}$). While the size of the effect is not dramatic, it is significant, and the effect was remarkably consistent across observers (fig 4.2b,e).

Discussion. The direction of the effect observed in this case is inconsistent with FLE. If we consider the abrupt change in direction akin to a visual flash, as in the flash lag effect, we would expect the horizontal position of the target to be perceived forward along its trajectory at the moment of the turn, or rightward. This is not the case. In this dimension, the turn-point is being localized back to the left. In the vertical direction the turn-point appears shifted downward, into a region that the target never actually traverses.

If we were to consider the target as having two separate trajectories, an initial down-right trajectory which terminates abruptly at some position, followed by a subsequent up-right trajectory originating from that same position, we instead expect the interplay of a variety of onset and offset effects: repmo at the termination of the first trajectory, and some combination of ORE and Fröhlich at the origination of the second. We believe this approach is the first case where these effects are placed in such close juxtaposition. This result is consistent with the idea that the two parts of the target's motion trajectory are represented as independent objects, that the turn-point itself is represented as the origin of the second part of the trajectory, and that under these conditions the onset of the second trajectory is subject to ORE.

Experiment 1b

Two targets. In this case, observers were presented with two targets, one above the fixation point, exactly as in experiment 1a, the other similar but rotated 180° about the fixation point (thus below, fixation, moving from down right towards fixation and then turning abruptly towards down-left). Subjects were instructed to attend both targets, and compare the positions of the two turn-points.

Participants. The same six observers from experiment 1a were used.

Stimuli. The display was the same as in the previous experiment, with the addition of a second target, also 0.87° in diameter (appearing to the lower right of fixation moving diagonally up and to the left at a 45° angle and a speed of 12.3 deg/sec for 1090 ms and then changed direction and moved down and to the left at a 45° angle and the same speed for another 1090 ms. At all times, the position of the second target was a point reflection of the first about the fixation point. As a result, both targets changed directions synchronously.

Task. Observers were asked to judge whether the turn-point of the top target appeared to the left or right of the bottom target in horizontal test blocks. Otherwise the task was the same as in experiment 1a.

Results. In the horizontal direction, the average perceived turn-point of the top target was 16.0 ± 5.8 arcmin to the left of the bottom target, which was significantly different from zero, or veridical perception ($t(5) = 2.76$; $p = 4.0 \times 10^{-2}$), but not significantly different from the single target case ($t(5) = 0.50$; $p = 0.64$; paired t-test). In the vertical direction the misperception of the top target was 15.1 ± 5.3 arcmin in the downward direction, a significant effect ($t(5) = 2.85$; $p = 3.6 \times 10^{-2}$), but again not significantly different from the single target case ($t(5) = 0.56$; $p = 0.60$; paired t-test).

Discussion. The effect here is the same for one or two targets. Attending to a second target below the fixation point and attempting to localize both turn-points for comparison purposes does little to modulate the effect. This may suggest that the role of attention in this effect is limited.

Experiment 2

ORE vs. TPE. The direction of the turn-point phantom effect is consistent with the idea that a second object representation is initiated at the turn-point and that the new object representation is subject to onset repulsion. To explore whether or not this explanation is sufficient to explain TPE, we used the same parameters from experiment 1a, ablating the pre-turn portion of the target's path. In each trial, the target would appear near the midline and move through only the second part of the trajectory.

Participants. Six naive, unpaid observers with normal or corrected to normal vision were used.

Stimuli. On each trial, a circular target 0.87° in diameter appeared 3.4° above the fixation point either directly aligned or 0.3, 0.6, or 0.9° to the right or left and moved diagonally up and to the right at a 45° angle at one of five speeds (4.1, 8.2, 12.3, 16.4, and 20.5 deg/sec). A central fixation point 0.10 x 0.10° was constantly visible. In all trial blocks, the horizontal position of the turn-point was tested. The fixation point and all targets were presented in white (83.5 cd/m^2) against a black background (0.215 cd/m^2). In a separate experimental block, the TPE trajectory from experiment 1a was tested at these five velocities for comparison.

Task. Observers were asked to judge the position of the appearance point of the target with respect to the fixation point.

Results. At the tested velocities, under these conditions, observers failed to consistently demonstrate significant ORE. Conversely, at the velocities tested, under these conditions, observers demonstrated significant TPE (figure 4.3).

Discussion. While the direction of TPE is consistent with an ORE explanation, the conditions of the experiment do not generate ORE. Another explanation is necessary.

Experiment 3a

Visual Flash. The motivation for this experiment is borrowed from the study of flash-lag. The differential latency theory of flash-lag suggests that neural processing time associated with detection of a flash is longer than that associated with the monitoring of continued motion. Because of this, the position of the flash is compared with a later position of the moving object and the flash appears to lag, or so the theory goes. As we showed before, constructing a similar explanation for TPE—the turn-point is detected late and then compared to a position further along the trajectory—predicts an effect forward along the post-turn path. This is not what we observe. However, if issues related to the timing of transient stimulus detection are common to both effects, then perhaps we can modulate the effect by capturing the perceived time of the turn-point with another visual transient.

Participants. Ten naive, unpaid observers with normal or corrected to normal vision were used.

Stimuli. The stimuli in the this experiment were the same as in experiment 1a, except that here a second identical target, horizontally aligned with the fixation point was flashed for 17ms at the time of the turn, or 50ms before the turn. In different blocks the second target appeared either 3.4° below the fixation point, or 3.4° above the turn-point. Results were the same in both cases.

Task. Observers were asked to judge the horizontal position of the turn-point of the target with respect to the fixation point. They were instructed to ignore the flashed visual target.

Results. In the case of the visual flash at the time of the turn, the position of the turn-point was perceived 14.0 ± 2.2 arcmin to the left of the actual turn-point, about the same magnitude of TPE observed in the original experiment($t(9) = 1.45$; $p = 0.18$, compared to no flash condition). In the case of the visual flash 50 ms prior to the turn, the position of the turn was perceived 8.5 ± 2.7 arcmin to the left of the actual turn-point. This effect is considerably smaller than previously observed. The flash presented prior to the turn significantly modulated TPE (39% change in magnitude; $t(9) = 3.04$; $p = 1.4 \times 10^{-2}$, paired t-test within subject).

Discussion. The presence of a second target had no real effect on the TPE observed, and again, the addition of a visual transient temporally coincident with the turn, has no effect. Presenting the flash prior to the turn does indeed modulate the effect, perhaps by capturing the moment of flash perception, and giving the flash a head-start for its slower detection process.

Experiment 3b

Cross-modal transient. Here we used an auditory transient, instead of a visual flash, in our attempts to capture the perceived moment of the turn.

Participants. The same ten observers from 3a were used.

Stimuli. The stimuli in the this experiment were the same as in experiment 1a, except that here observers wore headphones and a 400 Hz pure tone beep of 50 ms in duration would begin either at the time of the turn, or 50ms before the turn.

Task. Observers were asked to judge the position of the turn-point of the target with respect to the fixation point. They were instructed to ignore the auditory transient.

Results. In the case of the auditory transient at the time of the turn, the position of the turn-point was perceived 23.4 ± 2.2 arcmin to the left of the actual turn-point, a much greater magnitude TPE than in the original experiment. In the case of the auditory transient 50 ms prior to the turn, the position of the turn was perceived 17.9 ± 2.5 arcmin to the left of the actual turn-point. This effect, while larger than the original within-modality TPE, still represents a significant modulation from the temporally coincident case (24% change in magnitude; $t(9) = 5.31$; $p = 4.9 \times 10^{-4}$, paired t-test within subject).

Discussion. With the addition of an auditory transient, we still see the relative modulation with respect to transient timing, but we also see an overall increase in the effect. If, as has been postulated for FLE, the turn-point phantom is seated in the latency of transient detection, then perhaps what we are seeing here is the added latency of integrating cross-modal cues. Further discussion of possible explanations for these results follows.

General Discussion

In these experiments we demonstrate and explore a new example of motion-based mislocalization. The turn-point phantom effect (TPE) refers to the tendency of observers to localize the perceived point of a moving target's abrupt orthogonal direction change backwards along its eventual trajectory. We show the consistent presence of this effect both in the coordinate of continuous motion, horizontal in this case, and the coordinate of discontinuous motion, vertical in this case (experiment 1a), and that the effect is unchanged in the symmetrical case of two targets (experiment 1b). We show that while the direction of the effect is consistent with the onset repulsion effect, ORE alone is not a sufficient explanation under these conditions (experiment 2). Attempts to "capture" the moment of the turn and draw it to an earlier time modulate the effect, suggesting the timing of the perception of the turn is a relevant factor in generating this effect (experiments 3a and 3b).

Depending on how an abrupt change of direction is interpreted, the turn-point phantom can be thought of as a special case of other motion illusions. If we conceive of the turn as an abrupt transient event in an otherwise consistent motion display, the stimulus is yet another peculiar flash-lag condition. In this case, the "flash" is not a separate target, but a change of direction in a single target. The task here is to localize the position of the transient with respect to the reference frame (the fixation point) instead of the moving target. We will now consider the dominant explanations for FLE, and their ability to explain the findings here.

Nijhawan (1994) proposed active extrapolation of the continuously moving stimulus as a neural mechanism to recover the lost time associated with neural delays in perception. In such a theory, the moving stimulus is projected forward along its path of motion so that at the time visual information reaches the brain, the perception of the moving object matches its true position. There are numerous flash-lag results that strain this explanation, most notably the flash-terminated condition in which a temporally and spatially consistent flash accompanying the disappearance of a moving target does not produce FLE. The turn-point phantom provides another potential flash-lag finding that disputes motion extrapolation. Such an account would predict that the perception of the target would be actively pushed forward in the horizontal direction, which it is not. The opposite effect is seen.

Whitney and Murakami (1998) suggested that latency differences in the perception of a dynamic flash and a predictable, constantly moving target are responsible for flash lag. Since the neural delay associated with detecting the abrupt change involves a longer latency than the continued updating of the moving target's position, at the moment in which the two positions are compared, the moving target's position information is "newer" than the flash's. Consequently, the position of the flash is compared to a later, post-flash position of the moving object, and the flash appears to lag behind. In the case of TPE, the transient is a change in the motion vector and we expect a slowdown in target position updating on par with the latency of transient detection. Because the change in motion is the "flash," there should be no obvious latency differential in the processing of the two stimuli, unless we consider that it is the perception of motion and position that have different latencies. If the processing of the turn-point's position is separate from the processing of the target's motion, we may be seeing a faster latency for motion processing and a slower latency associated with the localization of the turn's position. The results of experiment 3—in which attempts to capture the turn point with an earlier transient and hence give it a head start result in a diminished effect—suggest that the relative timing of the turn-point is the issue and but a differential latency account still fails to predict the direction of TPE. The turn-point is not localized to a position later along the motion trajectory, but a position that the target has never actually occupied.

Eagleman and Sejnowski (2000) provide some insight with their postdiction theory of flash-lag. They argue that the issue is not one of differential latencies, so much as a matter of the way in which the percept is retrospectively constructed. If the flash effectively resets motion integration and the motion is later calculated as a weighted spatial average in some time window between the occurrence of the flash and the conscious perception of it, then the perceived position of the moving target is pushed forward relative to the flash. In our case, however, the "flash" has no separate and concrete positional information for the moving target to be compared with. The transient change is just a moment within the motion trajectory.

An important distinction between postdiction and previous accounts is that the integration window for localizing the moving object begins after the flash, and it is visual information subsequent to the transient event that is responsible for the percept. This is a compelling point in interpreting our effect. If we assume that the position of the turn-point is not perceived separately from the motion of the target but as a part of it, then the transient event is marked more by the moment in which it happens than any positional information. Postdiction gives us a framework for thinking about how that positional information is then constructed. We postulate that if the abrupt change in direction marks a reset in motion integration, and the position of the moving target over some subsequent time window is taken as a weighted spatial average, then the position of the turn-point must be constructed as a backwards projection in time of the motion. This explains why the turn-point phantom is perceived outside of the target's true trajectory. If it were only a matter of referencing the position at an earlier time (as in a differential latency account) then we might have expected the horizontal displacement observed in TPE, but it is only by backwards projection that we can explain both the horizontal and vertical components of the displacement. This also explains how capturing or shifting the temporal perception of the turn (experiment 3) results in a relative diminishing of the effect.

Implicit to this idea is the assumption that the position of the turn-point, and in general the instantaneous position of a moving object, is not maintained separately from the perception of the object's motion (Kanai, Sheth, and Shimojo, 2004). Since the perception of the overall motion trajectory is veridical—the turn is distinctly perceived as a right angle—and the perceived position of the turn-point is offset, we assume there is no separate positional information regarding the pre-turn trajectory. If there were, maintaining a consistent percept of the overall trajectory would require a shift in the perception of the entire first half of the target's path. This systematic shift in the perception of the first leg would occur even before the abrupt change in direction. If this shift were to occur retrospectively, subsequent to the turn, there would arise a moment of positional inconsistency when the perceived location of the object shifts from along its initial trajectory (or even past it), to a place backwards along its final trajectory. Yet all of this must occur without deforming the perceived shape of the overall path, which is perceived accurately, and without a conscious perception of the shift.

In single direction trajectories, such as those used in representational momentum, Fröhlich, and onset repulsion experiments, there are no inconsistencies between motion perception and position perception. In those cases there is no necessary stress created by the mapping of position onto the perceived motion or motion onto the perceived position because there is no second path which must be reconciled with the perception of the first. The abrupt direction change used in our experiments juxtaposes competing motion and position percepts and highlights the inconsistency between the two.

How does this inconsistency arise? Perhaps it is due to the dissociation between brain areas primarily processing motion (MT or MST) and areas processing position only (V1). It is widely accepted that modularity within the brain gives rise to local functional units solving different perceptual tasks. Within the visual system, the ventral stream deals primarily with perception of form, while the dorsal stream deals with analysis of position and motion. These streams are further divided into areas for the processing of luminance, color, texture, depth, complex motion, etc. The brain integrates perceptual information from different areas to construct a consistent percept of the world. When the information from two areas is in conflict, the brain must resolve the inconsistency. The turn-point phantom may be evidence that the visual system resolves discrepancies between inputs from regions involved in the perception of motion and regions involved in the perception of position by ignoring position.

With the turn-point phantom we demonstrate a new form of object mislocalization that may provide further evidence that 1) visual awareness is not strictly predictive or online, but relies on reconstruction of past events, and 2) in the case of moving objects, positional information is not maintained separately from the percept of motion, or that it is disregarded when inconsistencies arise.

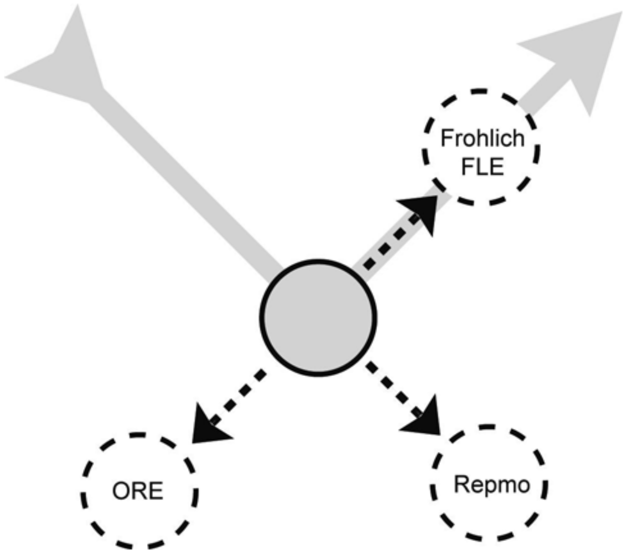


Figure 4.1: <u>Predicted contributions of known mislocalization effects on the perception of the turn-point</u>. Representational momentum (repmo) is projected forward from the end of the first leg of the trajectory. The Fröhlich effect is projected forward along the direction of motion from the start of the second leg of the trajectory, and the onset repulsion effect (ORE) is projected backwards along the direction of motion from the start of the second leg of the trajectory. Different latency theories of flash-lag predict that the target will be perceived further along its trajectory at the moment of the transient event.

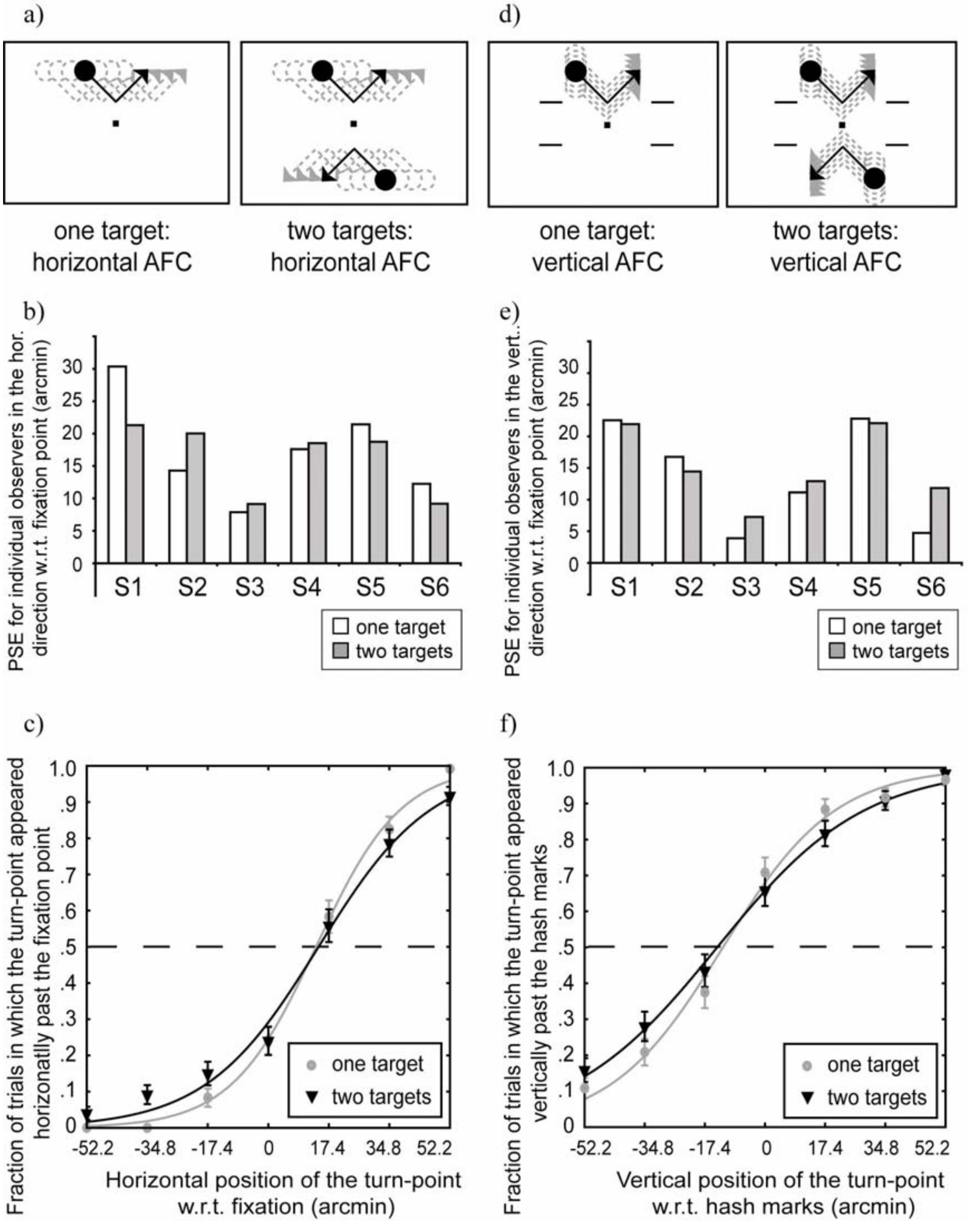


Figure 4.2: The Turn-point phantom for one and two targets. (a) and (d) Schematics of the experimental displays. (b) and (e) Points of Subjective Equality for each observer. (c) and (f) Psychometric curves for the average of all subject data under each condition.

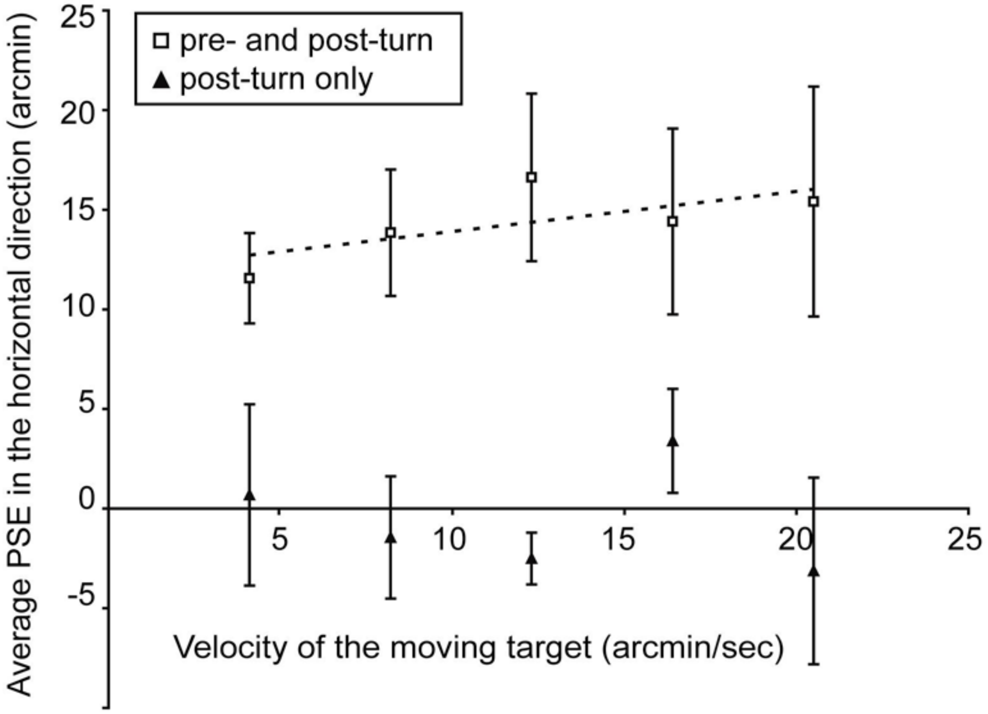


Figure 4.3: <u>TPE or ORE: Average shift in the perceived horizontal position of the turn-point at five different target velocities with and without the pre-turn part of the trajectory</u>. The "pre- and post-turn" condition contains the abrupt orthogonal turn. The "post-turn only" condition contains no turn. Observers were asked to judge the onset position in this case.

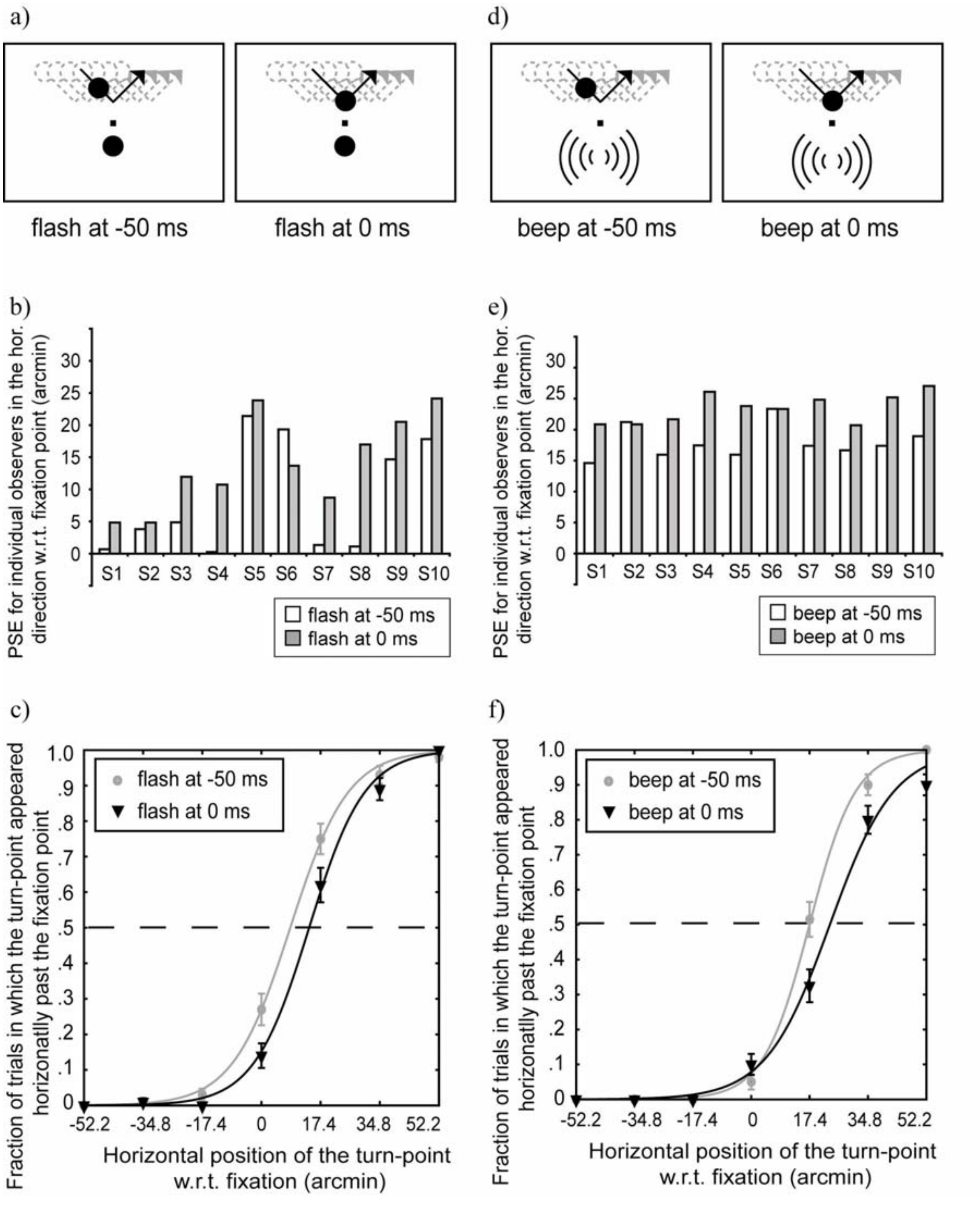


Figure 4.4: Modulation by an pre-turn transient event. (a) and (d) Schematics of the experimental displays. (b) and (e) Points of subjective equality for each observer. (c) and (f) Psychometric curves for the average of all subject data under each condition.

Chapter 5

Study 4: Vestibular Influence on Position Encoding of Visual and Auditory Stimuli

Abstract

A transient target presented prior to a passive rotation is perceived as shifted in the direction of the rotation. Previous studies (Blouin, Gauthier, and Vercher 1995, 1997; Blouin *et al.* 1995, 1998) have employed saccade and pointing response paradigms, and been limited to visual targets and whole-body rotations. Here, classical methods of constant stimuli were used to demonstrate the generality of this effect across visual and auditory stimuli, and across head-only and whole-body rotations. The effect was shown to diminish with added time delay between target presentation and turn initiation in the case of both head-only and whole-body turns. This suggests that the perception and encoding of target position is susceptible to vestibular influence over some time window well beyond the presentation of the target.

Introduction

Successful interaction with objects in the environment requires accurate information about their position relative to the body. Visual information, arriving in a retinal coordinate system, and auditory information, arriving in a head-centric coordinate system, must be synthesized to form a continuous, egocentric representation of the surrounding environment. In order to do this, position sense in the auditory domain must take into account head position with respect to the body, and position sense in the visual domain must take into account both head position with respect to the body and gaze direction of the eyes within the orbit. On top of all this, the brain must take into account the position of the body within the environment.

Normally, under conditions of active, intended movement, information about the state of the body is available from visual information, proprioceptive stretch receptors (Roll, Velay, and Roll, 1991), efferent copies of programmed motor movements, and vestibular signals of head acceleration. In this way, eye and head position during active orienting behaviors are precisely monitored (Howard, 1982). Under conditions of passive movement in darkness, only proprioceptive and vestibular information is available and positioning systems are considerably less accurate (Blouin *et al.*, 1998). It has been previously reported (Blouin, Gauthier, and Vercher 1995, 1997; Blouin *et al.* 1995, 1998) that updating the position of a visual cue following passive whole-body rotation about the yaw axis is subject to considerable error. This was shown using verbal estimates of target position with respect to body position, using a laser pointer to indicate perceived target location, pointing in darkness with an unseen finger tip, and making saccadic eye movements to the believed target position. In all cases, it was found that passive whole-body rotations produced a consistent mislocalization of the visual target in the direction of the rotation.

It has also been shown that the magnitudes of passive rotations in both the horizontal and vertical axes are systematically underestimated (Israel, Fetter, and Koenig, 1993). Such an underestimation of rotation magnitude, applied as a position transformation to visual space, would produce a shift in targets in the direction observed. Previous studies (Blouin, et. al., 1995b), however, simultaneously measured the perceived magnitude of rotation and perceived position of the target, and found that the size of vestibular underestimation is not sufficiently large to account for the perceptual shift in target position.

We took another approach to the issue of vestibular mislocalization. First, we chose not to use a motor or visual-motor pointing task as the response paradigm. By employing standard psychophysical methods of constant stimuli, we sought to isolate the perceptual effect from any adaptation in the spatial-motor system. Second, we sought to generalize the effect in a number of ways. We attempted to demonstrate vestibular mislocalization for passive head-only rotations, as opposed to whole-body rotations. We also attempted to demonstrate the effect in the auditory domain, without visual cues. We then tested for crossover of the effect between visual and auditory domains. Finally, in efforts to rule out a simple vestibular-underestimation explanation, we tested the time dependence of the effect.

Experiments

Experiment 1

Here we demonstrate the effect of passive head and trunk rotations on the localization of the position of a flashed target, and test whether the effect generalizes to head-only turns.

Participants: One author and four naive observers with normal or corrected to normal vision were used.

Apparatus: All experiments were performed on a Windows computer running Matlab (MathWorks Inc., Natick, MA) and the Psychophysics Toolbox extensions (Brainard, 1997).Observers sat in a completely dark room on an Accutronics motorized chair. The head was immobilized within the chair's helmet. The chair was capable of executing precise, passive head-only rotations, in which the helmet turned relative to the seat position, or head and trunk rotations, in which the helmet maintained a fixed position relative to the seat and both rotated together. The Accutronics chair was controlled from within Matlab via a serial connection with the experimental computer. Stimuli were presented on an arced hoop (radius = 87 cm) with presentation module ports spaced in five degree increments ranging 50 degrees to the left or right of center. The hoop was fixed to the main body of the chair, and thus yoked to the position of the observer's body. Each audio-visual presentation module consisted of a three-color red/green/amber 5 mm LED mounted in a cutout plate directly in front of the center of a 2 inch speaker. Modules affixed to the inside of the hoop and positioned LEDs at a viewing distance 80.5 cm from the observer. The presentation hoop was operated via a custom-built control box which offered simultaneous independent control of LEDs and speakers with 100 microsecond timing. The control box was programmed prior to each trial via a serial connection with the experimental computer. Observers responded by pressing pre-assigned keys on a keyboard that sat on their lap throughout the experiment. Prior to the start of experimental blocks, observers were familiarized with the chair, and its safety mechanisms.

Head-turn stimuli: In these experimental blocks, the seat position (and therefore trunk position) was fixed throughout, and only the head was rotated. Each trial began with the observer's head turned 20 degrees with respect to their body and a green fixation light (24.8 cd/m^2) visible for 750 ms at a position 0 degrees with respect to the head. After 300 ms of complete darkness, a red target (25.1 cd/m^2) was flashed for 80 ms at a position 0 degrees with respect to the body (20 degrees off center with respect to the head and visual field). After an observer-specific delay, adjusted before the start of experimental blocks to the smallest time that avoided any apparent motion of the flashed red target, observers experienced a passive 40 degree head turn towards center over a period of 1000 ms, such that the final head position was oriented 20 degrees off the trunk axis. A second identical red target was flashed for 80 ms at one of five positions (-10, -5, 0, 5, 10 degrees with respect to the first target) shortly after the completion of the turn (again, adjusted for individual subjects to avoid apparent motion of the target). Right-turn trials were followed by left-turn trials, and vice-versa. See figure 5.1 for a schematic of the spatial-temporal organization of head turn trials.

Whole-body turn stimuli: In these experimental blocks, the helmet position was fixed relative to the seat and the head and trunk were always rotated as a unit. Because the presentation hoop rotated with the chair (and the body), position of the presentation-modules within the hoop was adjusted to maintain the consistent physical location modules within the room. In addition, head and trunk-turns were executed at a slightly slower speed, taking 1300 ms to reach their final position. In this case, right and left turn directions were tested in separate experimental blocks to allow adjustment of the module positions. See figure 5.2 for a schematic of the spatial-temporal organization of whole-body turn trials.

Task: Observers were asked to judge the real position within the environment of the second red target with respect to the first in a two-alternative forced-choice task. They were instructed not to shift their gaze to the flashed targets, but to try and keep their gaze fixed directly in front of them throughout the experiment. Position of the second target was randomized across trials in blocks of 100 (20 trials each position).

Analysis: Displacements of the perceived position of the first target were probed and analyzed separately for left and right turns. For each observer and each turn-direction, responses were fitted with a psychometric curve

$$F(x) = 0.5 + \frac{(a + bx)}{2\sqrt{1 + (a + bx)^2}}$$

by minimizing the square error. Free parameters a and b were estimated by a least-squares criterion and the point of subjective equality (PSE) was obtained as ($-a/b$). Thus each PSE represented the displacement in the direction of turn of the second flashed target with respect to the first necessary for the two targets to appear at the same location. Positive values correspond to perceived displacement of the first target in the direction of the turn. A 2-way ANOVA (observer, direction) was performed to determine the relevance of turn direction in the magnitude of the effect. Direction was found to be uncorrelated with the magnitude of the effect. The PSEs of individual subjects were then averaged across direction and analyzed using a 2-tailed, unpaired t-test to determine significance.

Results: Both head-only turns and head-and-body turns were found to produce significant mislocalizations in the direction of the turn in all subjects, for both left and right turns (figure 5.3). Passive head turns of 40 degrees on average produced a 4.04 ± 0.59 degree shift in the stimulus flashed before the turn ($t(4) = 16.6$; $p = 7.7 \times 10^{-5}$). Passive whole body turns of 40 degrees on average produced a 4.32 ± 0.59 degree shift in the stimulus flashed before the turn ($t(4) = 19.2$; $p = 4.3 \times 10^{-5}$). The magnitude of the shift in the two cases was not significantly different ($t(4) = 0.78$; $p = 0.48$; within subject paired t-test).

Discussion: Vestibular mislocalization indeed does generalize to passive head turns. Despite the added information regarding the position of the head that is available from neck proprioception in the case of fixed-body, head-only turns, the visual system still experiences an error in updating position subsequent to the turn. Previous studies (Blouin, Gauthier, and Vercher, 1997), demonstrated that active head turns produced accurate post-turn localizations. The results here suggest that efferent copies of the programmed motor activity, and not neck proprioception, is largely responsible for the accurate updating of visual position. In this case, where there is no planned motor activity and only vestibular sense and proprioceptive information is available, the mislocalization is not significantly different from what is observed with whole-body rotation, where only vestibular information is available. The important distinguishing feature seems to be the passive nature of the turns.

Experiment 2

The vestibular mislocalization of visual targets generalizes to both head and body turns and is not significantly different between the two. Here we test whether the perceptual shift in target position generalizes to auditory targets as well.

Participants: The same observers from experiment 1 were used.

Stimuli: In these experimental blocks, the seat position was fixed throughout, and only the head was rotated. Each trial began with the observer's head turned 20 degrees with respect to their body and a green fixation light (24.8 cd/m^2) visible for 750 ms at a position 0 degrees with respect to the head. While the fixation light was visible a 1320 Hz fixation tone (80 ms in duration, 62.5 dB) was played twice (300 ms in between; 37.9 dB background noise level) at the same location (0 degrees with respect to the head). After 300 ms of complete darkness and silence, an 80 ms burst of white noise (full spectrum, 67.7 dB) was played at a position 0 degrees with respect to the body (20 degrees off center with respect to the head). After the same observer-specific delay from experiment 1, observers experienced a passive 40 degree head turn towards center over a period of one second, such that the final head position was oriented 20 degrees off the trunk axis. A second, identical white noise burst played for 80 ms at one of five positions (-10, -5, 0, 5, 10 degrees with respect to the first target) shortly after the completion of the turn (again, using the same individual subject delays from experiment 1). Right-turn trials were followed by left-turn trials, and vice-versa.

Task: Observers were asked to judge the position of the second white noise burst with respect to the first in a 2 AFC task. As before, they were instructed not to try to shift their gaze towards the location of the sound, but to try to keep their gaze fixed directly in front of them throughout the experiment. Position of the second sound was randomized across trials in blocks of 100 (20 trials each position).

Results: Significant vestibular mislocalization was observed for auditory stimuli in all subjects for both left and right turns (figure 5.4). Passive head turns of 40 degrees on average produced a 4.92 ± 0.79 degree shift in the white noise stimuli played before the turn ($t(4) = 14.0$; $p = 1.5 \times 10^{-4}$). While slightly greater than the effect observed for visual cues, the magnitude of the effect in both cases is not significantly different ($t(4) = 1.19$; $p = 0.30$; within subject paired t-test).

Discussion: Vestibular mislocalization indeed generalizes to the updating of auditory space as well. The mislocalization observed with auditory cues was indistinguishable from that observed with visual cues. This suggests that whatever effect the vestibular sense has on the updating of position, it either occurs in a similar fashion in both the visual and auditory domains, or it occurs in a common coordinate system. If the latter, it is not clear what this common coordinate system is likely to be. If these experiments employed an active spatial-motor response paradigm, it might be convenient to infer that the position of the pre-turn cue is preemptively coded in the coordinate system of the response paradigm—gaze direction for saccadic responses, reach coordinates for pointing tasks. In this case however, observers were making passive perceptual judgments as to the relative positions of cues before and after the turn. They were instructed, as well as they could, to maintain their gaze position midline with respect to the head. There is no clear choice of common coordinate frame.

Experiment 3

The vestibular mislocalization of targets occurs within both vision and audition. Here we explore how the effect operates across modalities.

Participants: The same observers from the previous two experiments were used.

V-A stimuli: The stimuli were the same as in experiment 2, except that the first white noise burst was replaced with a flashed red target (same parameters as experiment 1) with a white noise burst still played after the turn.

A-V stimuli: The stimuli were the same as in experiment 2, except that second first white noise burst was replaced with a flashed red target (same parameters as experiment 1) with a white noise burst still played before the turn.

Task: Observers were again asked to judge the position of the second stimulus with respect to the first in a two AFC task. As before, they were instructed not to try to shift their gaze towards the location of the stimuli, but to try to keep their gaze fixed directly in front of them throughout the experiment. Position of the second stimulus was randomized across trials in blocks of 100 (20 trials each position). V-A and A-V experiments were run in separate blocks.

Results: Cross-modal vestibular mislocalization was observed with head-only turns in all subjects for both left and right turns under both A-V and V-A conditions (figure 5.5). The magnitude of the observed effects was very different in the two cases. For an auditory cue played before the turn compared to a visual cue flashed after the turn, on average the pre-turn stimuli was perceived 4.11 ± 0.97 degrees shifted in the direction of the turn ($t(4) = 6.99$; $p = 2.2 \times 10^{-3}$). This is not significantly different from the effect observed with two visual cues ($t(4) = 0.08$; $p = 0.94$; within subject paired t-test) or two auditory cues ($t(4) = 1.22$; $p = 0.29$; within subject paired t-test). For a visual cue flashed before the turn compared to an auditory cue played after the turn, on average the pre-turn stimuli was perceived only 1.84 ± 0.44 degrees shifted in the direction of the turn ($t(4) = 6.76$; $p = 2.5 \times 10^{-3}$). This is a considerably smaller effect than any of the others ($t(4) = 6.17$; $p = 3.5 \times 10^{-3}$ w.r.t visual-visual; $t(4) = 5.87$; $p = 4.2 \times 10^{-3}$ w.r.t auditory-auditory; $t(4) = 3.59$; $p = 2.3 \times 10^{-2}$ w.r.t auditory-visual; within subject paired t-tests).

Discussion: A pre-turn auditory cue localized with a post-turn visual cue behaves no differently from either of the within-modality conditions. A pre-turn visual cue localized with a post-turn auditory cue exhibits a considerably diminished effect. This could be the result of either a diminished spatial error with regard to the position of the pre-turn visual cue, or a spatial error in the opposite direction with regard to the position of the post-turn auditory cue. We believe it to be the latter, and suggest that the spatial visual attention directed at localizing the pre-turn cue attracts the perceived position of the subsequent auditory cue.

Experiment 4

These results have shown that the perceived positions of both auditory and visual stimuli are shifted in the direction of immediately subsequent passive head or head and body turns. It is not yet clear whether this perceptual shift is the result of errors in the vestibular sense regarding estimation of the turn, errors in the visual or auditory localization of the target stimulus, or both. If the shift is due to underestimation of the magnitude of the turn, then the timing of the turn relative to the target stimuli should be largely irrelevant. Here, we test how additional time delays between target presentation and the execution of passive head turns affect the perceptual shift in visual-target position.

Participants: One author and four naive observers with normal or corrected to normal vision were used.

Stimuli: In these experimental blocks, the seat position was fixed throughout, and only the head was rotated. Experimental parameters were similar to experiment 1, except that in each trial, one of four additional delays was included between the presentation of the first red target and the initiation of the head turn (0, 250, 500, or 1000 ms beyond the subject-specific delay necessary to avoid apparent motion).

Task: The task and instructions were identical to those used in previous experiments. Timing of the turn (additional pre-turn delay) and position of the second stimulus were randomized across trials in blocks of 800 (alternating left and right turn trials: 20 trials at each of 5 positions, and each or 4 delays, in each turn direction).

Results: As the delay between the presentation of the first flashed stimulus and initiation of the head turn increases, we see the effect diminish (figure 5.6). While the average shift in initial target position with no added delay (3.95 ± 0.35 degrees) was comparable to what was previously observed ($t(4) = 0.15$; $p = 0.89$; unpaired t-test) with an added delay of 250 ms, the effect significantly diminished to 3.14 ± 0.34 degrees (20% decrease; $t(4) = 4.48$; $p = 1.1 \times 10^{-2}$ w.r.t no delay, within subject paired t-test). With a delay of 500ms, the observed shift diminished further to 2.32 ± 0.46 degrees (26% decrease; $t(4) = 4.01$; $p = 1.6 \times 10^{-2}$ w.r.t 250ms delay). And with a 1000 ms delay, the observed shift diminished still further to 1.13 ± 0.32 degrees (51% decrease; $t(4) = 5.87$; $p = 4.2 \times 10^{-3}$ w.r.t 500ms delay). Correlating the added delay time to the magnitude of the perceived effect yields a slope of -2.8 degrees of shift per second of added delay ($r = 0.82$).

Discussion: The addition of a delay between the flashed visual stimulus and the turn results in a diminished perceptual error. This suggests that there is some period just subsequent to the presentation of the target in which the observer’s representation of the target’s position is highly susceptible to vestibular influence. Perhaps this time window corresponds to the encoding of the position of the object, and the vestibular-prompted shift in perceptual coordinate frame that occurs during this encoding time (just subsequent to the target presentation) affects a shift in the encoded position of the pre-turn target. It should be noted that even after a one-second, pre-turn delay, there is still some small perceptual shift in target. Perhaps this residual error is due to the underestimation of the magnitude of the turn.

Experiment 5

Similar to experiment 4, here we further test the time dependence of the vestibular influence on visual target localization, in this case for yoked head and body turns.

Participants: One author and five naive observers with normal or corrected to normal vision were used.

Apparatus: The apparatus was the same as in previous experiments, except that the presentation module used to display the second red target was a visual-only, higher-spatial-resolution module with multiple LEDs spaced at 1.67° intervals. The positions of the modules were adjusted to account for the turning of the hoop and keep the pre- and post-turn positions of the flashed targets in the same real-world space.

Stimuli: In these experimental blocks, the helmet position was fixed relative to the seat and the head and trunk were always rotated as a unit. Six positions were used for the second flashed red target, spaced at -5, -3.33, -1.67, 0, 1.67, and 3.33 degrees relative to the first. In the one experimental block, additional delays of 0, 125, 250, 500, 1000, and 2000 ms (randomized within block) were added between the first flashed target and the initiation of the turn. In another experimental block, the same set of six delays was added between the conclusion of the turn, and the flashing of the second target. In a control block, the same delays were added, but no turns were executed.

Task: The task and instructions were identical to those used in previous experiments. Timing of the turn (additional delay) and position of the second stimulus were randomized across trials in blocks of 720 (20 trials at each of 6 positions, and each or 6 delays). Observers completed three different blocks, one in which the additional delay occurred just prior to the turn, one in which the additional delay occurred just subsequent to the turn, and one in which no turns occurred. Only rightward turns were used, as previous experiments had shown no relevant effect of turn direction.

Results: Again, we find that as the delay between the initial flashed target and the initiation of the turn increases, the perceived shift in the flashed target diminishes (figure 5.7). While the average shift in initial target position with no added delay (4.15 ± 0.58 degrees) was comparable to what was previously observed (t(5) = 0.20; p = 0.85; unpaired t-test) with each added delay, the effect diminished, reaching 1.34 ± 0.51 degrees after 2000 ms (68% overall decrease; t(5) = 4.89; p = 4.5 x 10^{-3} w.r.t no delay, within subject paired t-test). Correlating the added delay time to the magnitude of the perceived effect yields a slope of -1.3 degrees of shift per second of added delay (r = 0.96).

An added delay between the termination of the body turn and the occurrence of the second flashed stimulus did not produce the same results (figure 5.8). While the average shift in initial target position with no added delay (3.46 ± 0.11 degrees) was somewhat smaller than was previously observed, this difference was not significant (t(9) = 1.45; p = 0.18; unpaired t-test). More importantly, with added delays, the effect did not diminish. With the addition of a 2000 ms delay, the magnitude of the perceptual shift was 3.30 ± 0.25 degrees (5% overall decrease; t(5) = 0.67; p = 0.53 w.r.t no delay, within subject paired t-test). Correlating the added delay time to the magnitude of the perceived effect yields an essentially flat slope (0.014 degree change in shift per second) and an insignificant correlation (r = 0.036).

In the no-turn control blocks, no significant shift was observed with any delay (figure 5.9). Correlating the added delay time to the magnitude of the perceived effect yields an essentially flat slope (0.089 degree change in shift per second) and an insignificant correlation (r = 0.37).

Discussion: Again, for whole-body rotations we see that the addition of a delay between the flashed visual stimulus and the turn results in a diminished perceptual error. This is further evidence that the effect observed for passive head-only rotations is analogous to what has been previously observed for passive whole-body rotations. It is also worth noting, that in cases where the time delay followed the turn, before the presentation of the second stimulus, no time dependence of the perceptual shift was observed. Only an added delay between the initial target and the turn caused diminishment of the effect. This further supports the notion that spatial perception is susceptible to subsequent vestibular influence in only a brief time window.

Experiment 6

Despite instructing subjects to try and maintain a head-locked, centered gaze direction throughout the experiments, the vestibular ocular reflex is very strong and involuntary eye movements were reported by observers after experimental runs. In order to examine what relation, if any, these reflexive eye movements have to the perceptual shift in the pre-turn flashed target, we recorded observer eye movements during runs of the no-delay head- and trunk-turn experiment (head and trunk-turn condition of experiment 1).

Participants: One author and 3 naive observers who had participated in all of the previous experiments were used.

Apparatus: The apparatus was the same as in experiment 1, with the addition of a head mounted Eyelink II eye-tracker.

Stimuli: The stimulus was the same as in the head and trunk-turn condition of experiment 1, with the following exception: prior to the start of each trial, the green fixation target appeared for purposes of eye-tracker drift correction. The fixation target remained lit until observers reported that their gaze position was fixed upon it by pressing a pre-assigned key.

Task: The task and instructions were identical to those used in previous experiments. Position of the second stimulus was randomized across trials in blocks of 100 (20 trials at each of 5 positions). Only rightward turns were used, as previous experiments had shown no relevant effect of turn-direction.

Analysis: Eye position was recorded at 2 ms sample intervals throughout each trial. Data for each trial were then analyzed in MATLAB. Shift in eye position around the time of the turn was determined for each trial by taking the mean horizontal position of the eye a small time window before the turn and comparing it to the mean horizontal eye position in a small time window after the turn. At each target presentation position (5 total), the mean eye shift for trials in which an observer responded "left" was then compared to the mean eye shift for trials in which an observer responded "right" using a non-paired, equal variance t-test. This was done to gauge whether the size of the VOR on a given trial would correlate with the observer's response. This analysis was repeated using time windows of 10, 20, 50, and 100 ms, and shifting the time point used as the before-after threshold over a range from 0 to 300 ms after the actual turn initiation.

Results: There was no observed correlation between the size of the VOR on a given trial, and the observer's response. While there were very specific combinations of parameters (time window, perceived start time of the turn) that showed marginal a correlation between VOR and response for one target position or another, these were inconsistent, and likely spurious. No consistent trend was observed.

Discussion: The lack of correlation between magnitude of the vestibular ocular reflex (as measured by the average shift in eye position associated with the onset of the turn) and perceived shift in the target (as measured by observer response on a given trial) suggests that the vestibular mislocalization observed in these experiments is not likely to be a byproduct of the VOR.

General Discussion

We have shown, in these experiments, that there is a robust shift in perceived real-world location of targets presented just prior to passive turns. This effect generalizes across head-only and whole-body turns (experiment 1), and across visual and auditory domains (experiments 2 and 3). This effect is also time dependent and degrades as the time between target presentation and onset of the turn increases (experiments 4 and 5). Furthermore, this effect cannot be easily explained as a byproduct of reflexive eye movements (experiment 6).

Prior studies have described a shift in the reported position of a visual target flashed prior to a passive whole-body turn using spatial-motor response paradigms. In these studies, observers made active orienting movements, either with the eye or with the hand. In a sense, this forced observers to translate the perceived position of the target into a particular egocentric coordinate frame—body centered in the case of hand movements, head centered in the case of eye movements. Furthermore, spatial-motor response paradigms invite the possibility that the reported error is not a perceptual one so much as a vestibular-influenced motor programming error. In our experiments, we have instead used a perceptual paradigm in which observers passively witness both the target stimulus and a later probe, and then compare the perceived positions of the two. In this way, we have avoided encouraging a particular coordinate representation. It still remains a possibility that visual and auditory positions are perceived and encoded in gaze coordinates as a matter of saccade orienting efficiency. But, in the absence of an impending spatial-motor response, observers are free to passively perceive the locations of the target and probe stimuli without needing to form a spatial-motor plan for responding. This suggests that the reported shift in position is not merely an error in motor programming, but an error in the internal representation of the target's position.

The swift degradation of this effect with added delay between the target and turn suggests that this is not merely a failure to correctly update the position of the target with respect to the turn. If this were only a failure to correctly update the position, it would be unlikely that added time before the turn would have any effect. If anything, it seems more likely that added time after the turn, before the presentation of the probe stimulus, would allow observers to more accurately perform the correct coordinate shift necessitated by the turn. This is not what we observe. This strongly argues that the moments after the presentation of the initial target are crucial for the accurate perception of its location.

We propose that the position of the flashed target is perceived and encoded over some time window subsequent to the target's onset and that this window extends past the moment in which the target has been extinguished and through the initiation of the turn. As such, the localization of the target is influenced by the vestibular signal despite the fact that there is no perceived apparent motion. Under this framework, the results of experiments 4 and 5 suggest that this time window is not fixed for visual perception, but is variable as a function of the degree of uncertainty in spatial perception. While the observed shifts for head-only and whole-body turns were remarkably similar throughout our experiments, there was a distinction between the two that we observed in the time dependence of the effects. In the case of head-only turns, the effect diminished almost twice as fast with added delay. This may have been the contribution of neck proprioception interfering with the vestibular influence on position encoding in a time dependent manner, but we believe it more likely that it was a function of the underlying uncertainty of environmental coordinate frames in the two experiments. In experiment 4, observers experienced passive head turns, but throughout the experiment the position of their bodies was stable within the room. In experiment 5, observers experienced whole-body turns in which the degree of uncertainty with regard to the position of their body within the room was much greater. We believe this served to compound their degree of uncertainty with regard to target position, and consequently lead to a longer time window for target position encoding.

We believe that these findings argue strongly for an explanation in which the localization of the target is a process that begins at the time of target presentation, but continues over a time period that stretches through the initiation of the turn. We assert that this is not merely an error in the updating of the object position following subsequent vestibular input, but an error in the primary perception and encoding of position. It is a passive turn occurring after the target has been extinguished—after the retinal input involved in the localization task is complete—that influences the target's perceived location. As such, we feel that this effect provides further evidence of the postdictive nature of spatial perception.

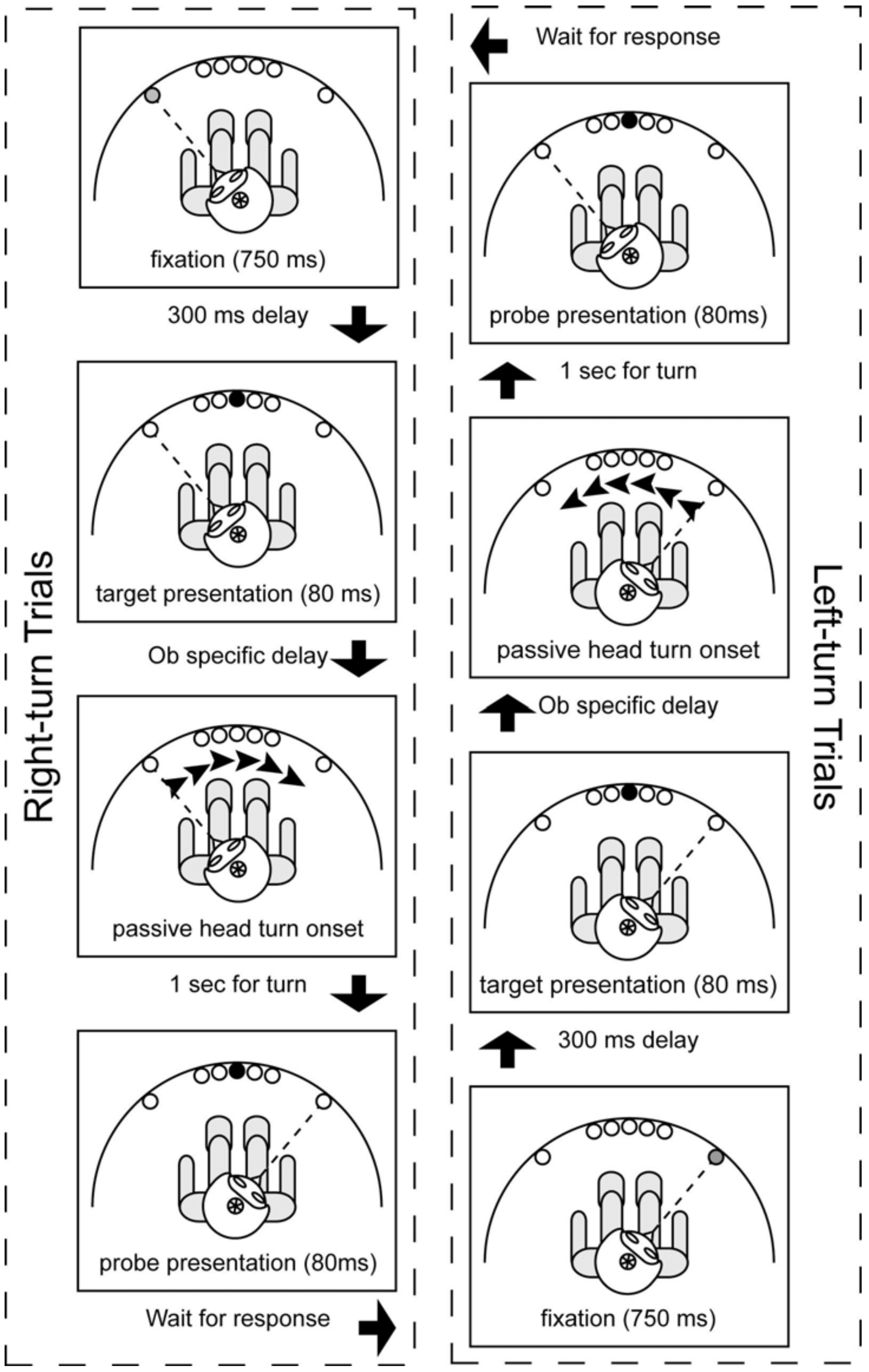


Figure 5.1: Schematic representation of experimental display for head-turn experiments.

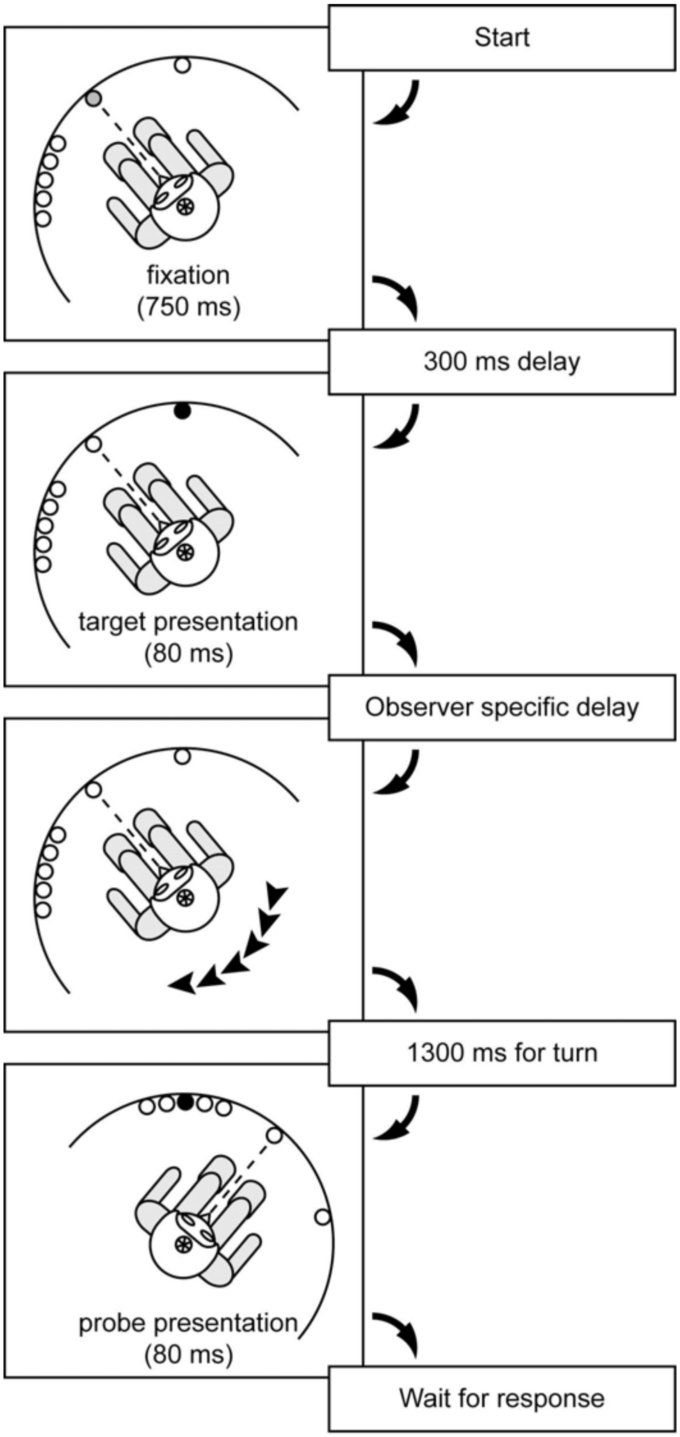


Figure 5.2: Schematic representation of experimental display for whole-body turn experiments.

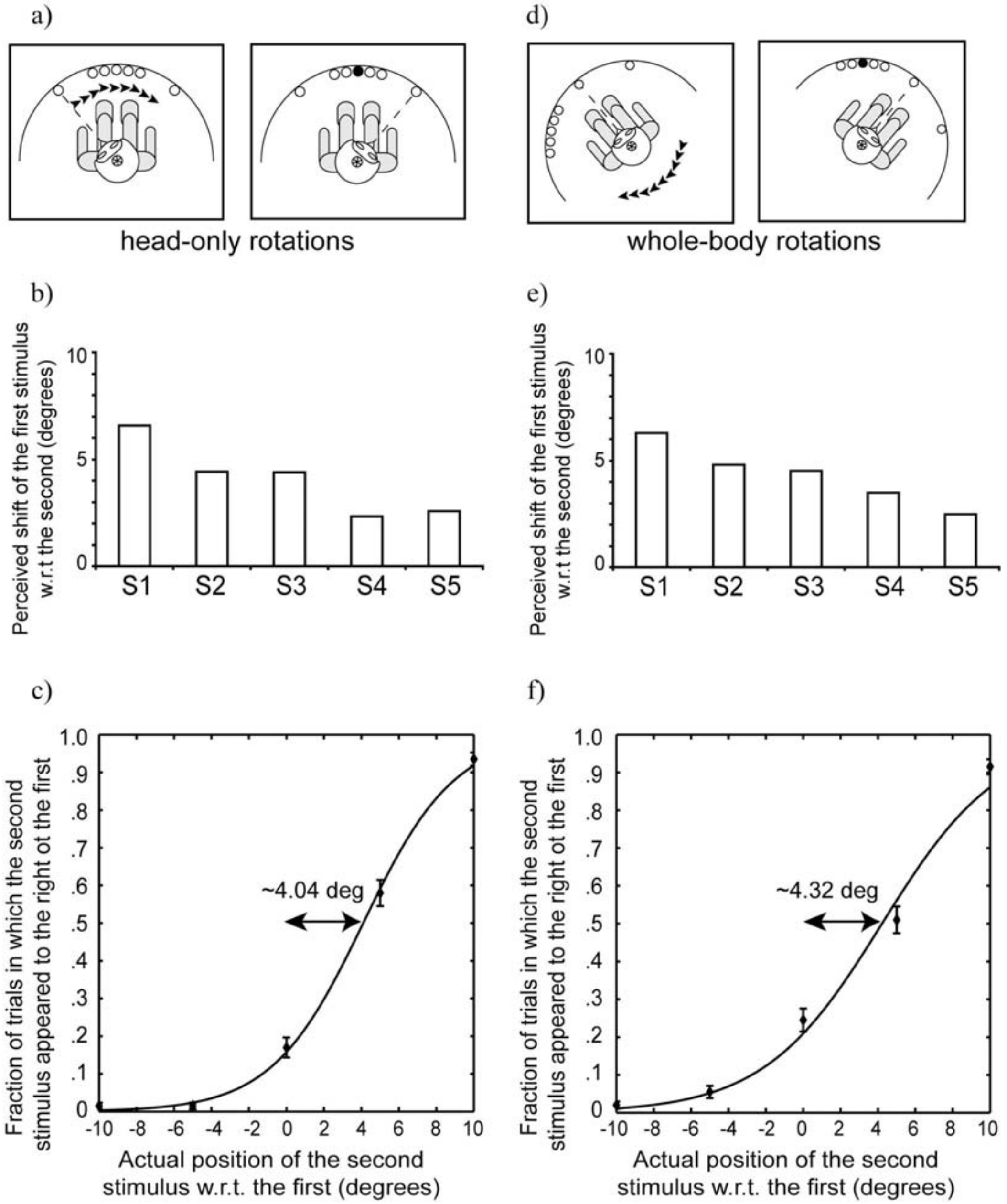


Figure 5.3: Localization errors for visual targets after head-only (a–c) or whole-body (d–f) rotations: (a) and (d) Abbreviated schematic of the experimental condition. (b) and (e) Points of subjective equality for each observer. (c) and (f) Psychometric curves for the average of all subject data under each condition.

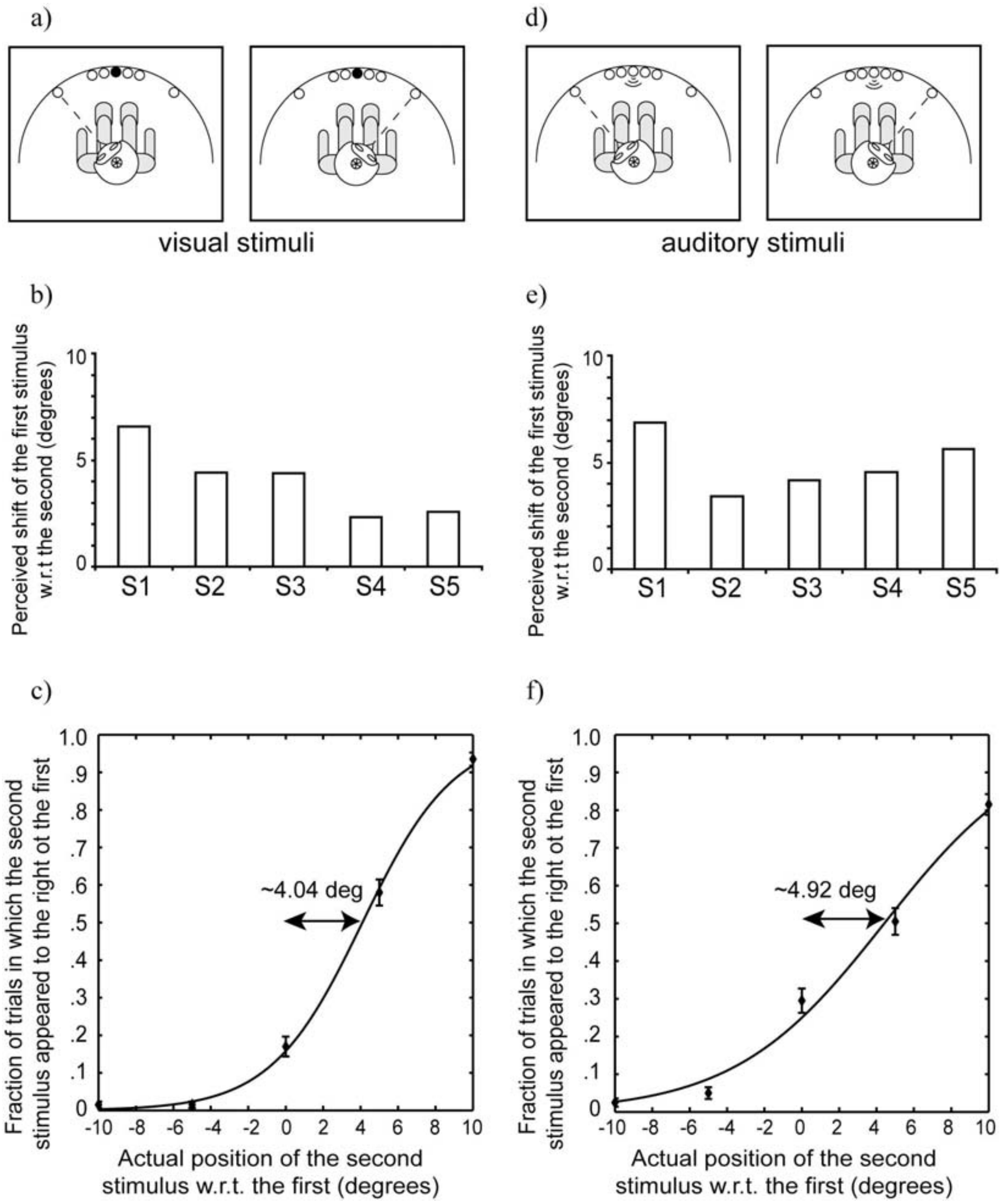


Figure 5.4: Localization errors after head-only rotations for visual (a–c, same as in Figure 2, provided for comparison) and auditory (d–f) stimuli: (a) and (d) Abbreviated schematic of the experimental condition. (b) and (e) Points of subjective equality for each observer. (c) and (f) Psychometric curves for the average of all subject data under each condition.

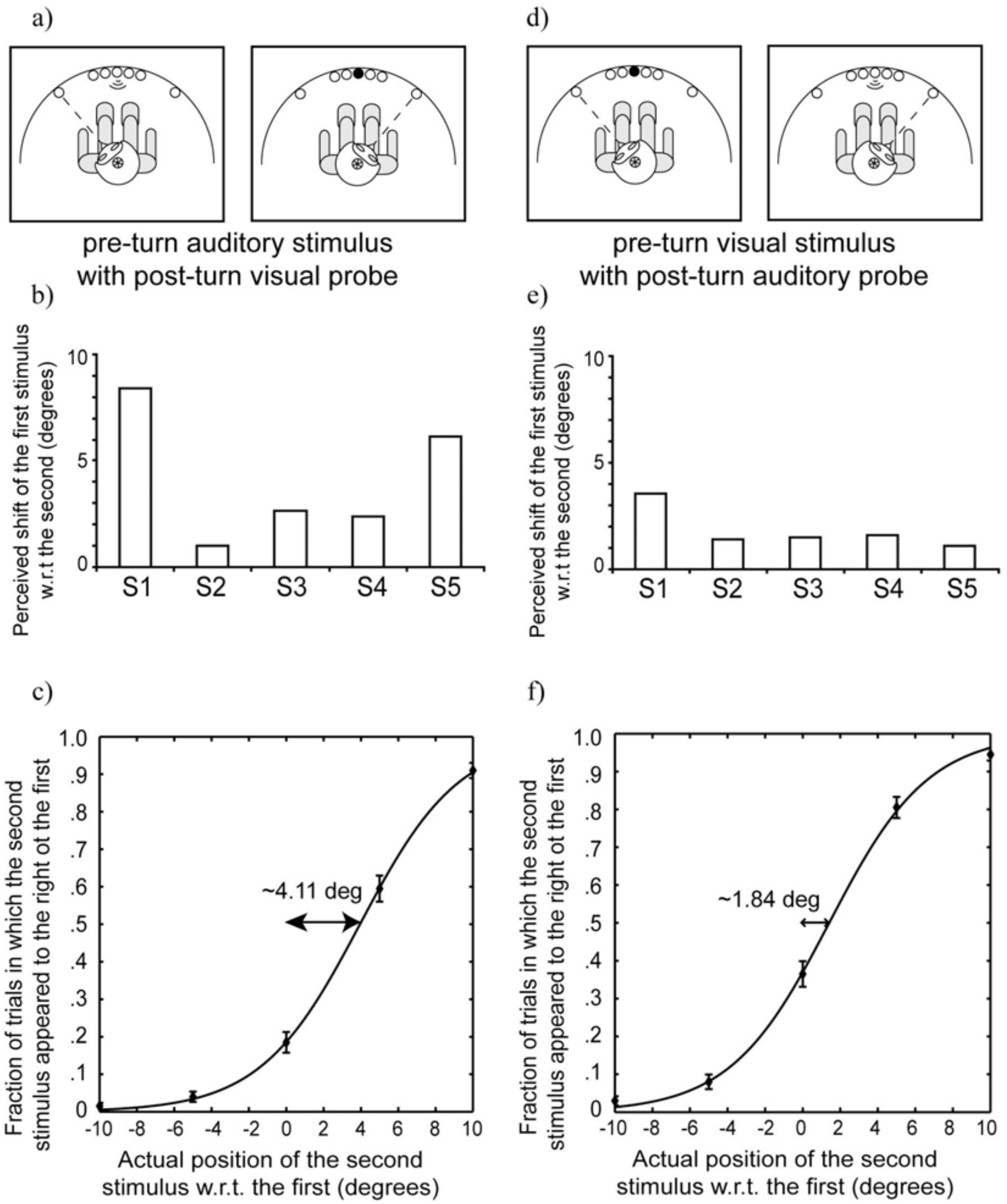


Figure 5.5: Cross-modal localization errors after head-only rotations for a pre-turn visual stimulus localized with a post-turn auditory probe (a–c) and for a pre-turn auditory stimulus localized with a post-turn visual probe (d–f): (a) and (d) Abbreviated schematic of the experimental condition. (b) and (e) Points of subjective equality for each observer. (c) and (f) Psychometric curves for the average of all subject data under each condition.

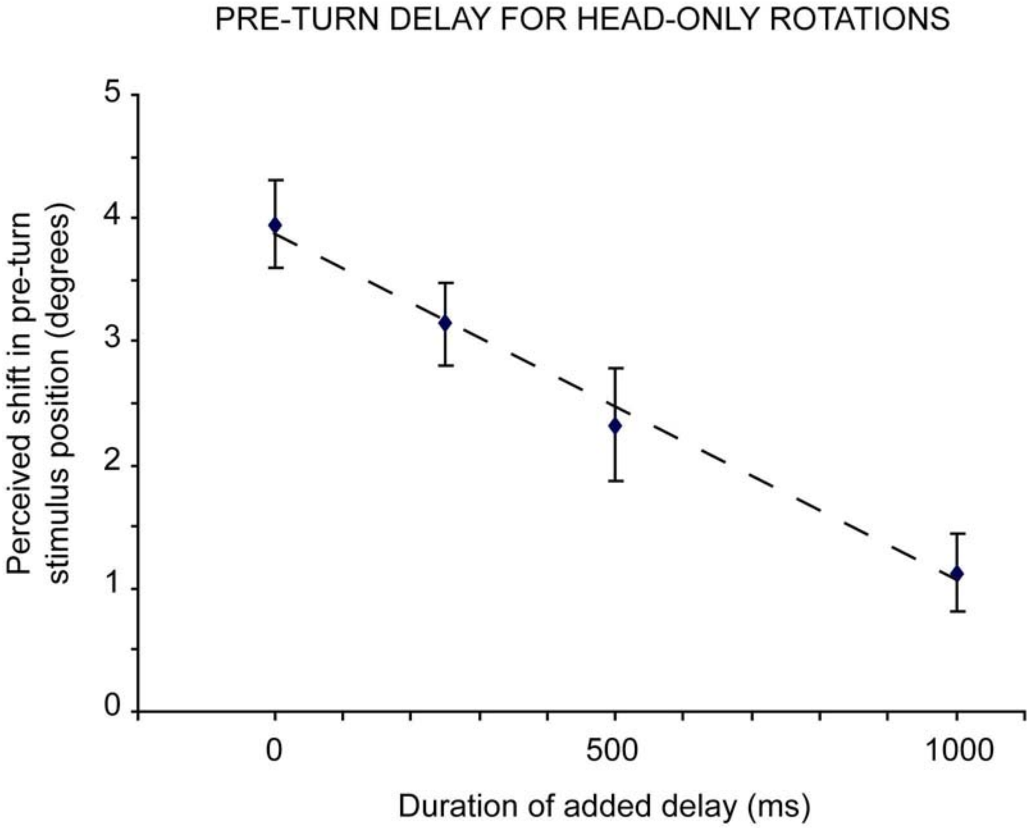


Figure 5.6: The localization error diminishes with added delay between the target presentation and the initiation of a passive head turn.

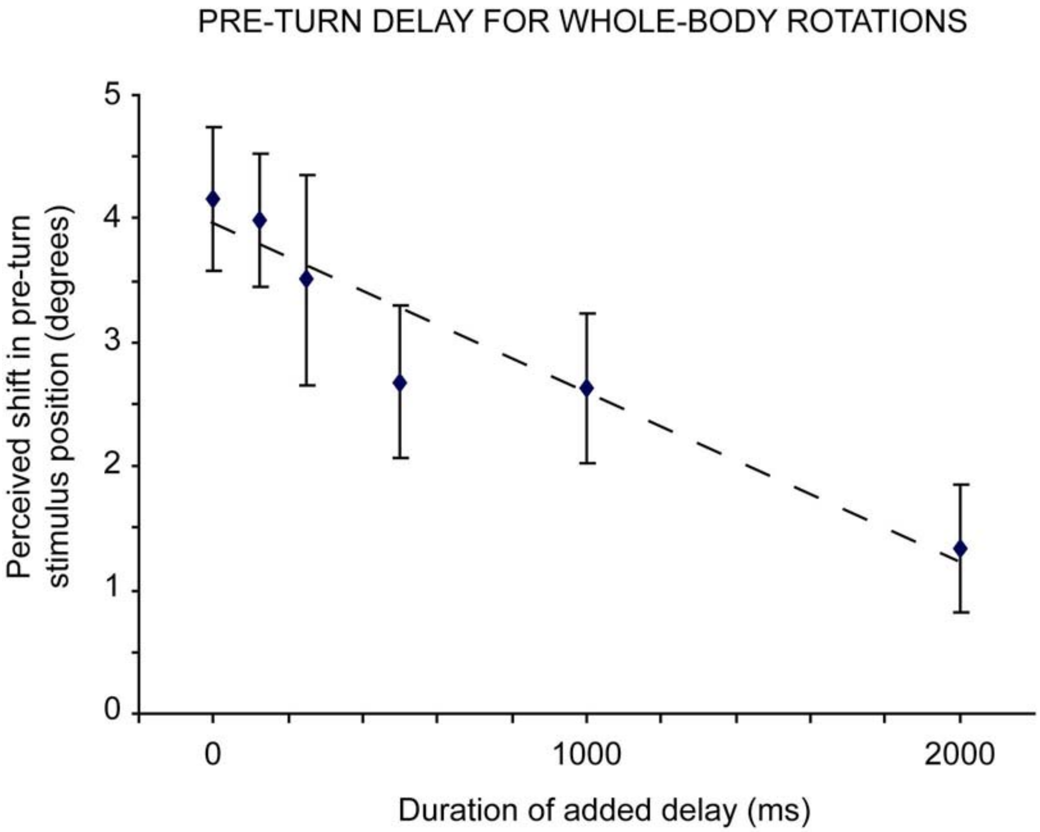


Figure 5.7: The localization error diminishes with added delay between the target presentation and the initiation of a passive whole-body turn.

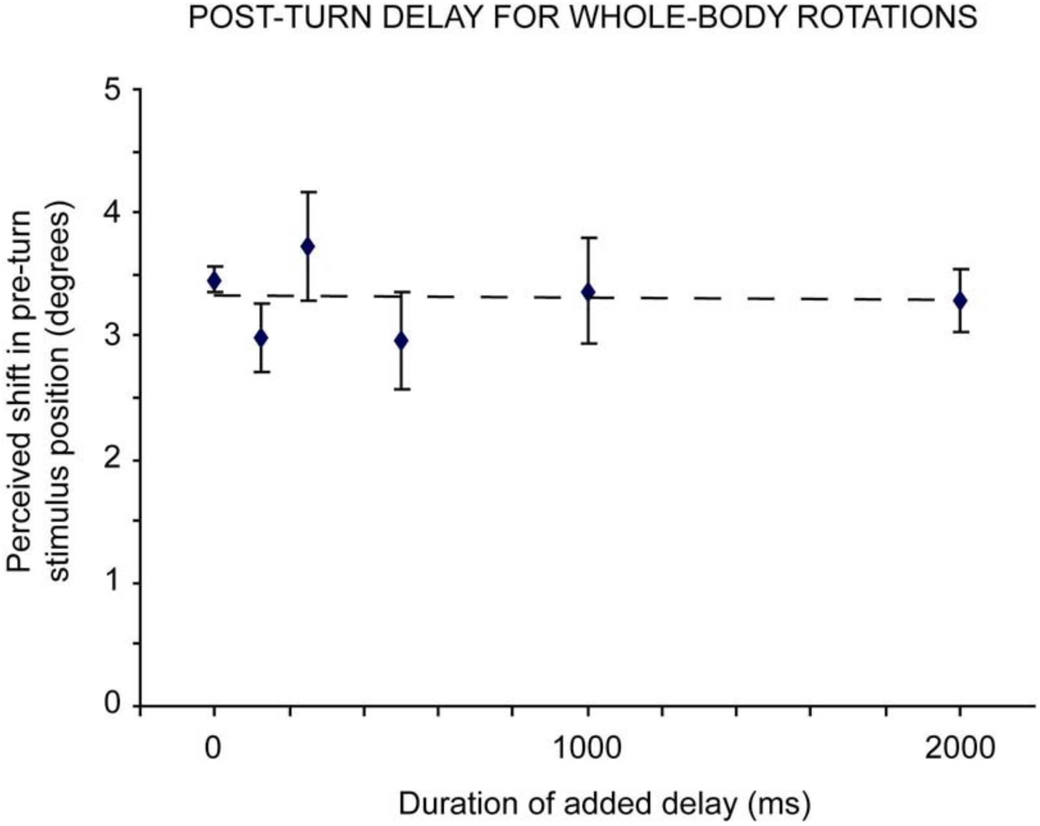


Figure 5.8: The localization is unaffected by added delays between the completion of the passive whole-body rotation and the presentation of the probe stimulus.

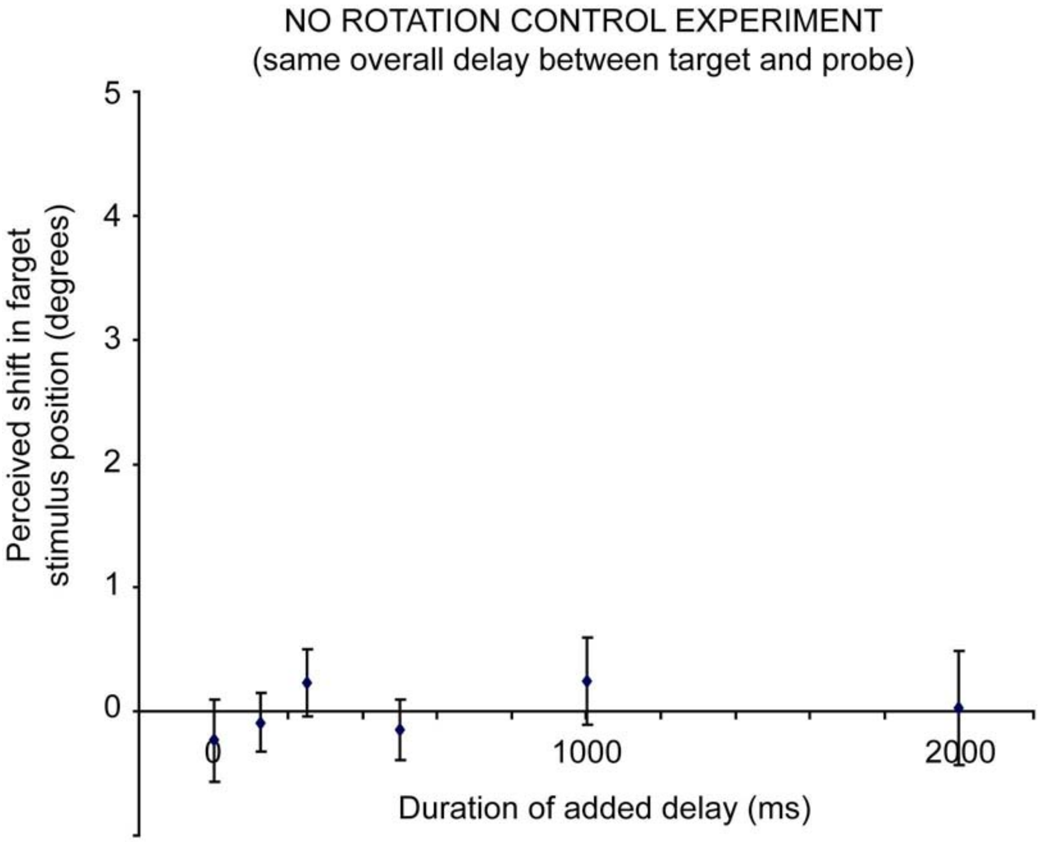


Figure 5.9: There is no significant localization error without the execution of a passive rotation.

Chapter 6

Study 5: Synchrony in Opposite Hemifield and Peripheral Ventriloquism

Abstract

The saccadic localization of auditory targets has been shown to be attracted to synchronously presented visual distracters. In this study, we demonstrated that this “ventriloquism” effect was greatest for visual distracters closer to the midline than their auditory targets and that in this situation saccades exhibited a “center of gravity” effect similar to that described for visual distracters accompanying visual targets. The effect was consistent in the perceptual domain as well as the visual-motor, though the relative magnitude was diminished in the former. Results also suggest that auditory and visual stimuli in opposite hemifields exhibit a qualitatively different kind of interaction in which one location or the other appears to win out. Timing of the visual distracter was found to be irrelevant for early presentations, but visual distracters presented after auditory target onset exhibited gradually diminished effects. These results suggest that auditory localization is susceptible to influence from irrelevant visual cues for some time after auditory target onset, and as such, auditory spatial awareness is postdictive in nature.

Introduction

The visual system plans and executes many fast, voluntary eye movements that serve to bring areas of interest into the high acuity fovea. While more often than not these saccades are instigated by visual stimulation, with its high spatial resolution, saccades can be directed at or elicited by auditory (Zahn, Abel, and Dell’Osso, 1978; Zahn *et al*., 1979; Zambarbieri *et al*., 1982) or somatosensory stimuli as well (Amlot *et al*., 2003). The superior colliculus is known to be responsible for the planning and control of these saccadic eye movements (Wurtz and Albano, 1980; Sparks and Mays, 1980) and has been implicated as a center for multi-sensory integration (Sparks and Nelson, 1987; Sparks, 1988). Neurophysiological evidence suggests that sensory modalities are each represented topographically in the superior colliculus and that despite the different reference frames associated with each modality—eye-centered reference frame for vision, head-centered reference frame for hearing, body-centered reference frame for somatosensory information—all maps exist in spatial register (Jay and Sparks, 1984, 1987a, 1987b). This is important so that information from multiple modalities can be quickly and efficiently synthesized to form a relevant motor plan for eye movements.

It has been shown that the human visual system's response to a visual target is influenced by the concurrent presentation of additional visual targets (Coren and Hoenig, 1972; Findlay, 1982), suggesting that saccade programming is subject to modulation by extraneous visual stimuli and that it demonstrates an overall "center of gravity" effect. This effect is observed in cases in which the distracters are in the same visual hemifield, temporally consistent, and spatially close together (Ottes, van Ginsberg, and Eggermont, 1984, 1985). It has also been shown that saccades to auditory targets are influenced by the concurrent presentation of visual distracters in a similar fashion (Lueck *et al.*, 1990). In the case where visual distracters are presented in the same hemifield as the auditory target, at the same time, the distracter appears to attract the saccade destination proportionally to its distance from the auditory target. This "ventriloquism" effect is much stronger for distracters nearer to the midline than for those at eccentricities greater than the auditory target.

Previous studies (Lueck *et al.*, 1990), focused on the role of spatial compatibility by varying the spatial disparity between the auditory and visual stimuli, and were limited to only one eccentricity of auditory target (15 degrees). We investigated the effect of visual distracters on auditory targets at this eccentricity and closer to the midline. We reexamined capture across opposite hemifields for auditory targets closer to the midline, and compared it to the degree of capture observed with same-side peripheral stimuli with distracters both nearer and farther than the auditory target. The degree of perceptual capture was also tested and compared, in order to gauge how much of this effect is limited to the visual-motor system. Also, we explored the role of timing in the presentation of visual distracters under these conditions, testing the relative effects of both early and late distracters.

Experiments

Experiment 1: Visual-motor ventriloquism

Here we examine the influence of a synchronized visual distracter on saccades to an auditory target. This experiment was performed in two separate blocks, one for auditory and visual distracters at different eccentricities in the periphery, and another for auditory and visual cues in opposite visual hemifields.

Participants: One author and five naive observers with normal or corrected to normal vision were used. The same observers were used in all experiments.

Apparatus: All experiments were performed on a Windows computer running Matlab (MathWorks Inc., Natick, MA) and the Psychophysics Toolbox extensions (Brainard, 1997). Observers sat in a completely dark room at the center of an arced hoop (radius = 87 cm) with presentation module ports spaced in five degree increments ranging 50 degrees to the left or right of center. Each audio-visual presentation module consisted of a three-color red/green/amber 5 mm LED mounted in a cutout plate directly in front of the center of a 2 inch speaker. Modules affixed to the inside of the hoop and positioned LEDs at a viewing distance 80.5 cm from the observer. The presentation hoop was operated via a custom built control box which offered simultaneous independent control of LEDs and speakers with 100 microsecond timing. The control box was programmed prior to each trial via a serial connection with the experimental computer. Observers responded by executing saccadic eye movements. Eye-tracking was performed using a head-mounted Eyelink II system which communicated with the experimental computer via a direct internet connection. Prior to the start of experimental blocks, observers were familiarized with the eye-tracking system and the hoop.

Peripheral Stimuli: Each trial began with the observer fixating a green LED (24.3 cd/m^2) at a position 0 degrees with respect to the head. Observers pressed a button on a handheld controller to in order to recalibrate the eye-tracking system and signal their readiness. After a delay of 500 ms the fixation LED extinguished, and after another 500 ms in complete darkness stimulus presentation began. In each trial a full spectrum white noise stimulus (67.7 dB) was played for 200 ms from a speaker at a position 5 or 15 degrees to the right of center, while a red LED (25.1 cd/m^2) was visible for the same 200 ms either at the same position as the sound, or 10 degrees offset (either 5 or 15 degrees to the right of fixation). Observers then made a saccade to the source of the sound, after which they pressed a button on the controller to complete the trial. See figure 6.1 for a schematic representation of the experimental procedure.

Opposite hemifield stimuli: All parameters were the same, except for the positions of the stimuli. All stimuli were presented at positions 5 degrees to the left or right of fixation.

Task: Observers were instructed to keep their gaze fixed straight ahead at the fixation target until the presentation of the sound stimulus, and then to look at the source of the sound as quickly and as accurately as they could. To ensure that observers began each trial with a neutral gaze position, observers pressed a button while fixating the green LED at the start of the trial. Trials only progressed if the eye-tracking system reported the observer's eye position consistent with the fixation target. The positions of the visual and auditory stimuli were randomized across trials in blocks of 80 (20 trials each condition, 2 visual target positions, 2 auditory target positions). Separate blocks of trials were performed for peripheral and opposite hemifield stimuli.

Analysis: Prior to the start of each experimental block, the eye-tracker was calibrated for both eyes, and the eye which provided the most accurate calibration results was tracked throughout the experiment. Eye position was recorded at 2 ms sample intervals throughout each trial. Data for each trial were then analyzed in MATLAB. A saccade event was defined as any change in eye position in which the velocity exceeded 0.3 degrees/second or the acceleration exceeded 4000 degrees/second^2. Trials in which observers executed a saccade before the presentation of the stimulus were eliminated from the analysis. For each of the remaining trials, the horizontal displacement and onset time of the first saccade subsequent to stimulus presentation was considered.

Results: In all cases, the average displacement of the initial saccade towards an auditory target accompanied by a spatially coincident visual cue was significantly different from the average displacement of the initial saccade towards the same auditory target with a spatially displaced visual cue.

In experimental blocks testing peripheral ventriloquism, the near auditory target (5 degrees) presented with a coincident visual target elicited an average initial saccade displacement of 3.38 ± 0.22 degrees toward the auditory target, while the same auditory target accompanied by a visual distracter 10 degrees further in the periphery elicited an average initial saccade displacement of 6.53 ± 0.63 degrees ($t(10) = 4.69$; $p = 8.57 \times 10^{-4}$; non-paired, two-tailed t-test). The far auditory target (15 degrees) presented with a coincident visual target elicited an average initial saccade displacement of 9.43 ± 0.55 degrees, while the same auditory target accompanied by a visual target 10 degrees nearer to fixation elicited an average initial saccade displacement of 3.89 ± 0.20 degrees ($t(10) = 9.54$; $p = 2.45 \times 10^{-6}$; non-paired, two-tailed t-test).

Average reaction times for each condition were not significantly different in any of the peripheral stimulus conditions. For the near auditory target with coincident and disparate visual cues, the average onset of the first saccade occurred 269 ± 26 ms and 307 ± 33 ms after the onset of the auditory stimulus, respectively ($t(10) = 0.93$; $p = 0.38$; non-paired, two-tailed t-test). For the far auditory target with coincident and disparate visual cues, the average onset of the first saccade occurred 260 ± 19 ms and 246 ± 19 ms after the onset of the auditory stimulus respectively ($t(5) = 0.50$; $p = 0.63$; non-paired, two-tailed t-test). For the near auditory target, a disparate visual distracter further in the periphery increased saccade latency while for the far auditory target, a disparate visual distracter closer to the fovea, slightly reduced saccade latency. Neither of these differences was significant.

In experimental blocks testing opposite hemifield ventriloquism, an auditory target (5 degrees from fixation) presented with a coincident visual cue elicited an average initial saccade displacement of 3.51 ± 0.18 degrees toward the auditory target, while an auditory target accompanied by a visual distracter in the opposite hemifield (5 degrees from fixation) elicited an average initial saccade displacement of 0.63 ± 0.37 degrees ($t(22) = 6.99$; $p = 5.16 \times 10^{-7}$; non-paired, two-tailed t-test). If we look at the amplitude of the initial saccade, without regard for its direction, we find instead that for the auditory target accompanied by a visual distracter in the opposite hemifield, the average initial saccade amplitude was 3.03 ± 0.18 degrees ($t(22) = 2.04$; $p = 0.053$; non-paired, two-tailed t-test). Thus the amplitude of the initial saccade was smaller, but not significantly so, for the disparate visual-cue conditions, however the average displacement of the initial saccade was quite different.

Average reaction times for each condition were not significantly different for coincident vs. disparate visual cues in the opposite hemifield experimental blocks. For the coincident and disparate visual cues, the average onset of the first saccade occurred 275 ± 15 ms and 314 ± 16 ms after the onset of the auditory stimulus respectively ($t(22) = 1.83$; $p = 0.081$; non-paired, two-tailed t-test). The disparate stimuli elicited longer saccade latencies than the coincident stimuli, though this difference was not significant.

Discussion: We observed some degree of visual-motor “ventriloquism” in all conditions, however the measured effect was stronger for distracters closer to the midline than for those farther in the periphery. This agrees with what Lueck *et al*. (1990) found for a 15 degree auditory target. In our experiments, capture across the midline was much stronger than what Lueck observed, perhaps owing to the smaller disparity between our target and distracters (10 degrees in our case, as opposed to ≥18.75 degrees).

Closer examination of capture across the midline revealed that average amplitude of the initial saccades was modulated (14%; 3.03 vs. 3.51 degrees), but the greater effect was due to many of those saccades being executed in entirely the wrong direction. If we conceive of the "center of gravity" effect as the synthesis of multiple competing motor programs which can occur either through weighted averaging of the alternatives or competition between them, the opposite hemifield results strongly argue for competition. How this sort of competition affects the spatial perception of targets in the absence of eye-movements is unclear.

Experiment 2: Perceptual Ventriloquism

In order to compare the measured degree of perceptual capture with what we observed in experiment 1 for visual-motor capture, the ventriloquism effect was quantified using methods of constant stimuli. This experiment tested only for a 15 degree auditory target.

Stimuli: The presentation of auditory and visual targets was similar to experiment 1. Each trial began with the observer fixating a green LED (24.3 cd/m^2) at a position 0 degrees with respect to the head. After a delay of 500 ms the fixation LED extinguished, and then after a randomly selected delay between 500 and 1500 ms (uniform distribution) in complete darkness, stimulus presentation began. In each trial a full spectrum white noise stimulus (67.7 dB) was played for 200 ms from a speaker at a position 15 degrees to the right of center, while a red LED (25.1 cd/m^2) was visible for the same 200 ms either at the same position as the sound, or 10 degrees offset (5 degrees to the right of fixation). After 1000 ms in total darkness, an amber LED (24.8 cd/m^2) appeared for 200 ms at one of 5 positions (0, 5, 10, 15, or 20 degrees to the right of fixation).

Task: Observers were instructed to keep their gaze fixed straight ahead at the fixation target throughout the experiment, and report, using pre-assigned buttons on a hand-held controller, whether the location of the amber probe stimulus was to the left or right of the origin of the sound.

Analysis: For each observer and each combination of auditory and visual stimulus presentation, responses were fitted with a psychometric curve

$$F(x) = 0.5 + \frac{(a + bx)}{2\sqrt{1 + (a + bx)^2}}$$

by minimizing the square error. Free parameters *a* and *b* were estimated by a least-squares criterion and the point of subjective equality (PSE) was obtained as ($-a/b$). Thus each PSE represented the perceived position of the origin of the auditory stimulus. The PSEs of individual subjects were then averaged across observer and analyzed using 2-tailed, within-subject paired t-tests to determine significance.

Results: The perceived position of the auditory cue with coincident visual cue was essentially veridical, while the perceived position of the auditory cue with disparate visual cue was significantly different (figure 6.2). The auditory target with coincident visual cue (15 degrees) was perceived to be at the same location as an amber light at 14.75 ± 0.31 degrees. The auditory target with disparate visual cue (5 degrees) was perceived to be at the same location as an amber light at 7.96 ± 0.56 degrees (t(5) = 8.45; p = 3.8 x 10^{-4}; within subject, paired t-test).

In order to compare these results with the visual-motor results from experiment 1, in both cases we scaled the perceptual displacement between the coincident and disparate conditions as a function of the perceptual distance between the auditory targets in the two coincident positions. Thus here, for a 15 degree auditory target, the 5 degree visual distracter demonstrated a 72% degree of capture in the perceptual response paradigm and a 91.5% degree of capture for the saccade response paradigm.

Discussion: It appears that the effect is present in the perceptual domain as well as the visual-motor domain, as we would except from the common experience of ventriloquism. The measured perceptual shift in auditory target position was smaller than that observed for saccades. Thus it seems as though there is a greater degree of capture in the visual-motor system than in the perceptual domain.

Experiment 3: Visual-motor ventriloquism for early visual distracters

In order to investigate the role of synchrony in audio-visual capture, we varied the relative timing of auditory and visual cues, presenting the visual distracter prior to the onset of the white noise burst.

Stimuli: The presentation of stimuli was the same as in experiment 1 except that in each trial, the onset of the auditory stimulus was delayed by one of five time delays (0, 50, 100, 150, or 200 MS).

Task: The task was the same as in experiment 1. Observers were instructed to keep their gaze fixed straight ahead at the fixation target until the presentation of the sound stimulus, and then to look at the source of the sound as quickly and as accurately as they could. The positions of the visual and auditory stimuli as well as the added delay in the auditory stimulus were randomized across trials in blocks of 400 (20 trials each condition, 2 visual target positions, 2 auditory target positions, 5 time delays). Separate blocks of trials were performed for peripheral and opposite hemifield stimuli.

Results: As in experiment 1, in all cases, the average displacement of the initial saccade towards an auditory target accompanied by a spatially coincident visual cue was significantly different than the average displacement of the initial saccade towards the same auditory target with a spatially displaced visual cue. Also, the difference in average initial saccade displacement between coincident and disparate stimulus conditions remained stable across the timing disparities tested.

In experimental blocks testing peripheral ventriloquism, the differences in average initial saccade displacement between the coincident and disparate conditions for the 15 degree auditory target (figure 6.3) were 5.74, 5.20, 5.24, 5.02, and 5.54 degrees when the visual cue preceded the auditory target by 200, 150, 100, 50, and 0 ms, respectively. Timing of the visual distracter prior to the auditory target showed no significant correlation with the difference in average initial saccade displacement (R = 0.32). For the 5 degree auditory target (figure 6.4), displacements were 3.87, 3.93, 4.05, 2.91, and 3.14 degrees when the visual cue preceded the auditory target by 200, 150, 100, 50, and 0 ms, respectively. Timing of the visual distracter prior to the auditory target showed no significant correlation with the difference in average initial saccade displacement (R = 0.69).

In experimental blocks testing opposite hemifield ventriloquism (figure 6.5), the differences in average initial saccade displacement between the coincident and disparate conditions were 2.78, 2.71, 2.68, 2.69, and 2.88 degrees when the visual cue preceded the auditory target by 200, 150, 100, 50, and 0 ms, respectively. Timing of the visual distracter prior to the auditory target showed no significant correlation with the difference in average initial saccade displacement (R = 0.37).

Reaction time data showed a general decrease in latency with early visual distracter presentation, though no significant difference in the latency between coincident and disparate conditions as a function of visual distracter timing. In experimental blocks testing peripheral ventriloquism, the differences in average initial saccade latency between the coincident and disparate conditions for the 15 degree auditory target (figure 6.6) were 50, 9, 28, 13, and 13 ms when the visual cue preceded the auditory target by 200, 150, 100, 50, and 0 ms, respectively. Timing of the visual distracter prior to the auditory target showed no significant correlation with the difference in average initial saccade latency (R = 0.64). For the 5 degree auditory target (figure 6.7), differences in latency were -2, 6, 19, 1, and 39ms when the visual cue preceded the auditory target by 200, 150, 100, 50, and 0 ms, respectively. Timing of the visual distracter prior to the auditory target showed no significant correlation with the difference in average initial saccade latency (R = 0.57).

In experimental blocks testing opposite hemifield ventriloquism (figure 6.8), the differences in average initial saccade latency between the coincident and disparate conditions were 7, 39, 59, 62, and 39 ms when the visual cue preceded the auditory target by 200, 150, 100, 50, and 0 ms, respectively. Timing of the visual distracter prior to the auditory target showed no significant correlation with the difference in average initial saccade latency (R = 0.63).

Discussion: As the visual distracter preceded the auditory target, the differences in saccade magnitudes between coincident and disparate conditions were stable. While the opposite hemifield conditions showed an overall diminishment in saccade magnitude with early visual cue presentation, this was regardless of position and the magnitude of the "capture" was roughly constant throughout. This suggests that the resolution of whatever competition occurs between visual and auditory stimulus is largely unaffected by early presentation of the visual stimulus. At the same time, the reaction time in early visual presentation cases decreased, indicating that while the relative outcome of the competition remained unaffected, the resolution was reached with greater speed.

Experiment 4: Visual-motor ventriloquism for late visual distracters

To further investigate the role of synchrony in audio-visual capture, we presented the visual distracter subsequent to the onset of the white noise burst.

Stimuli: The presentation of stimuli was the same as in experiment 1 except that in each trial, the onset of the visual stimulus was delayed by one of five time delays (0, 50, 100, 150, or 200 MS).

Task: The task was the same as in experiment 1. Observers were instructed to keep their gaze fixed straight ahead at the fixation target until the presentation of the sound stimulus, and then to look at the source of the sound as quickly and as accurately as they could. The positions of the visual and auditory stimuli as well as the added delay in the visual stimulus were randomized across trials in blocks of 400 (20 trials each condition, 2 visual target positions, 2 auditory target positions, 5 time delays). Separate blocks of trials were performed for peripheral and opposite hemifield stimuli.

Results: As in experiment 1, in all cases, the average displacement of the initial saccade towards an auditory target accompanied by a spatially coincident visual cue was significantly different from the average displacement of the initial saccade towards the same auditory target with a spatially displaced visual cue. For late visual cues, however the difference in average initial saccade displacement between coincident and disparate stimulus conditions significantly decreased as a function of the visual delay.

In experimental blocks testing peripheral ventriloquism, the differences in average initial saccade displacement between the coincident and disparate conditions for the 15 degree auditory target (figure 6.9) were 5.54, 4.71, 4.98, 3.65, and 2.21 degrees when the auditory target preceded the visual cue by 0, 50, 100, 150, and 200 ms, respectively. Timing of the visual distracter subsequent to the auditory target showed a significant negative correlation with the difference in average initial saccade displacement (R = 0.93; slope = -0.015 degrees of capture per ms). For the 5 degree auditory target (figure 6.10), displacements were 3.14, 3.49, 3.28, 2.47, and 1.77 degrees when the auditory target preceded the visual cue by 0, 50, 100, 150, and 200 ms, respectively. Timing of the visual distracter subsequent to the auditory target showed a significant negative correlation with the difference in average initial saccade displacement (R = 0.95; slope = -0.0075 degrees of capture per ms).

In experimental blocks testing opposite hemifield ventriloquism (figure 6.11), the differences in average initial saccade displacement between the coincident and disparate conditions were 2.87, 2.71, 2.37, 2.30, and 1.25 degrees when the auditory target preceded the visual cue by 0, 50, 100, 150, and 200 ms, respectively. Timing of the visual distracter subsequent to the auditory target showed a significant negative correlation with the difference in average initial saccade displacement (R = 0.91; slope = -0.0073 degrees of capture per ms).

Reaction-time data showed no correlation between time of the visual delay and latency differential between coincident and disparate cases. In experimental blocks testing peripheral ventriloquism, the differences in average initial saccade latency between the coincident and disparate conditions for the 15 degree auditory target (figure 6.12) were 13, 27, 43, 13, and 15 ms when the auditory target preceded the visual cue by 0, 50, 100, 150, and 200 ms, respectively. Timing of the visual distracter subsequent to the auditory target showed no significant correlation with the difference in average initial saccade latency (R = 0.14). For the 5 degree auditory target (figure 6.13), differences in latency were 39, 16, 23, 12, and 30 ms when the auditory target preceded the visual cue by 0, 50, 100, 150, and 200 ms, respectively. Timing of the visual distracter subsequent to the auditory target showed no significant correlation with the difference in average initial saccade latency (R = 0.32).

In experimental blocks testing opposite hemifield ventriloquism (figure 6.14), the differences in average initial saccade latency between the coincident and disparate conditions were 39, 95, 63, 77, and 68 ms when the auditory target preceded the visual cue by 0, 50, 100, 150, and 200 ms, respectively. Timing of the visual distracter subsequent to the auditory target showed no significant correlation with the difference in average initial saccade latency (R = 0.31).

While in no experiments did the difference in latency between coincident and disparate conditions correlate with added visual delay, it was the case, in all experiments, that delaying the visual cue increased the latency to the auditory target. In peripheral experiments with the 5 degree auditory stimulus, the latency increased 0.27 ms per ms of visual delay for the coincident visual cue (R = 0.91) and 0.30 ms per ms of visual delay for the disparate visual cue (R = 0.93). In peripheral experiments with the 15 degree auditory stimulus, the latency increased 0.57 ms per ms of visual delay for the coincident visual cue (R = 0.93) and 0.53 ms per ms of visual delay for the disparate visual cue (0.95). In opposite hemifield experiments, the latency increased 0.29 ms per ms of visual delay for the coincident visual cue (R = 0.94) and 0.36 ms per ms of visual delay for the disparate visual cue (R = 0.89).

Discussion: As the visual distracter onset followed the auditory target onset, the differences in saccade magnitudes between coincident and disparate conditions diminished, thus the relative amount of “capture” decreases with late distracter presentation. This is different from the early visual presentation, in which the relative amount of capture remained unchanged, suggesting that the late presentation of a visual distracter helps favor the auditory program, whereas the converse was not the case. We might expect that the late presentation of the visual cue would completely ablate the capture effect, particularly for the 200ms case in which the visual cue does not begin until the auditory cue has come to its end. Overall, reaction times continue to increase as the visual cue is presented later and later, reinforcing the notion that even a late and irrelevant visual cue affects some competition with the auditory target.

General Discussion

At the time of auditory target presentation, the visual motor system must make a choice. Provided with a sound and light in close temporal alignment, it must reconcile competing spatial representations in order to make a best likelihood estimation of the sound's position. In cases where the spatial information is consistent, there is little problem accomplishing this task. In cases where the spatial information between modalities is in conflict, some resolution must be reached.

As previously pointed out, topographic maps of sensory space in the superior colliculus are kept in alignment, despite their different reference frames. This is an important consideration if those sensory maps are used for quickly computing eye-movement motor programs. The position corresponding to an object in the eye-centered visual field should correspond to its position in the head-centered auditory sensory space, regardless of the direction of gaze. It has been shown that the auditory receptive fields of multi-modal cells in the superior colliculus of monkeys do, in fact, shift with eye movement so that these spatial maps can remain aligned (Jay and Sparks, 1984).

We demonstrated, in experiment 2, that the influence of a near visual distracter on an auditory target in the periphery is a perceptual one, not just a visual-motor error. This strongly suggests that, at least in this case, auditory space is shifted under the influence of the visual distracter. In a completely dark room, with only proprioceptive information from stretch receptors in the extraocular muscles, without corroborating visual evidence of the position of the eyes within the head, it is not hard to believe that the relative alignment of the two perceptual spaces is labile. If, under normal conditions, the alignment is enforced by the regular coincidence in time and space of real world multi-modal objects and events, perhaps under these impoverished perceptual conditions, the lone high resolution spatial stimuli becomes, by default, perceptually aligned with the singular auditory stimuli. We suggest that what we are seeing with the peripheral visual capture of auditory targets is a shift in the perceived auditory space under the influence of a visual stimulus.

We found that this shift in auditory space can occur, to a degree, with any close temporal alignment of the auditory and visual stimuli, though the size of the shift diminishes as the visual cue is presented later than the auditory target. Synchrony of onset and offset of the two stimuli does not seem to be that relevant, as the degree of capture is roughly stable so long as the visual stimulus precedes the auditory one. Instead, we suggest that that the presence of the visual cue during some time window after the onset of the auditory target is sufficient to produce the effect. In the case where the visual presentation begins early, the visual distracter is still lit for some time following audio onset. Only in the case of the full 200ms sound delay is the visual distracter extinguished at the time of audio onset. Owing to the shorter processing latency in the auditory domain, it is possible that the neural signal of the visual cue, in this case, is still arriving at the brain when the neural activity corresponding with auditory stimulus onset occurs.

At the time of auditory stimulus onset, the saccadic decision making process begins. The fact that the capture effect occurs at all in the case where the visual distracter follows the auditory cue suggests that localization occurs over a broad time window subsequent to auditory target onset, and that the relative alignment of visual and auditory space is susceptible to the influence of visual distracters throughout that period. The smooth degradation of the effect with increased delay in the visual target presentation suggests that the time window does not have a fixed or rigid end point, but continues to affect the localization decision with diminished relevance until the time of visual-motor response.

For visual and auditory stimuli in opposite hemifields, we found something different. Whereas the degree of capture was considerably larger than what was previously found for stimuli placed further in the periphery, this capture was not a matter of "center of gravity" averaging saccades. Indeed, the average amplitude of saccades in the opposite hemifield case was only slightly diminished (14%), and instead the saccade displacements fell into a bimodal distribution (figure 6.15), with some fraction of saccades centered on the auditory target, and another centered on the visual distracter. In this case, it seems more likely that there is a competition between two representations that cannot be easily reconciled with a shift in auditory space, and that competition is resolved with one motor program or the other winning out. Indeed, a closer look at the saccade distributions in all of the conditions (figure 6.16) suggests that visual distracters placed farther in the periphery exhibit this kind of competition as well.

This study has demonstrated the relative degrees of visual-motor capture of auditory targets by visual distracters nearer to midline, farther in the periphery, and in the opposite hemifield. We have shown the effect is larger in the visual-motor domain than in perceptual space, but that the effect is also a perceptual one. As well, we have demonstrated the relevance of distracter timing in the computation of auditory target localization.

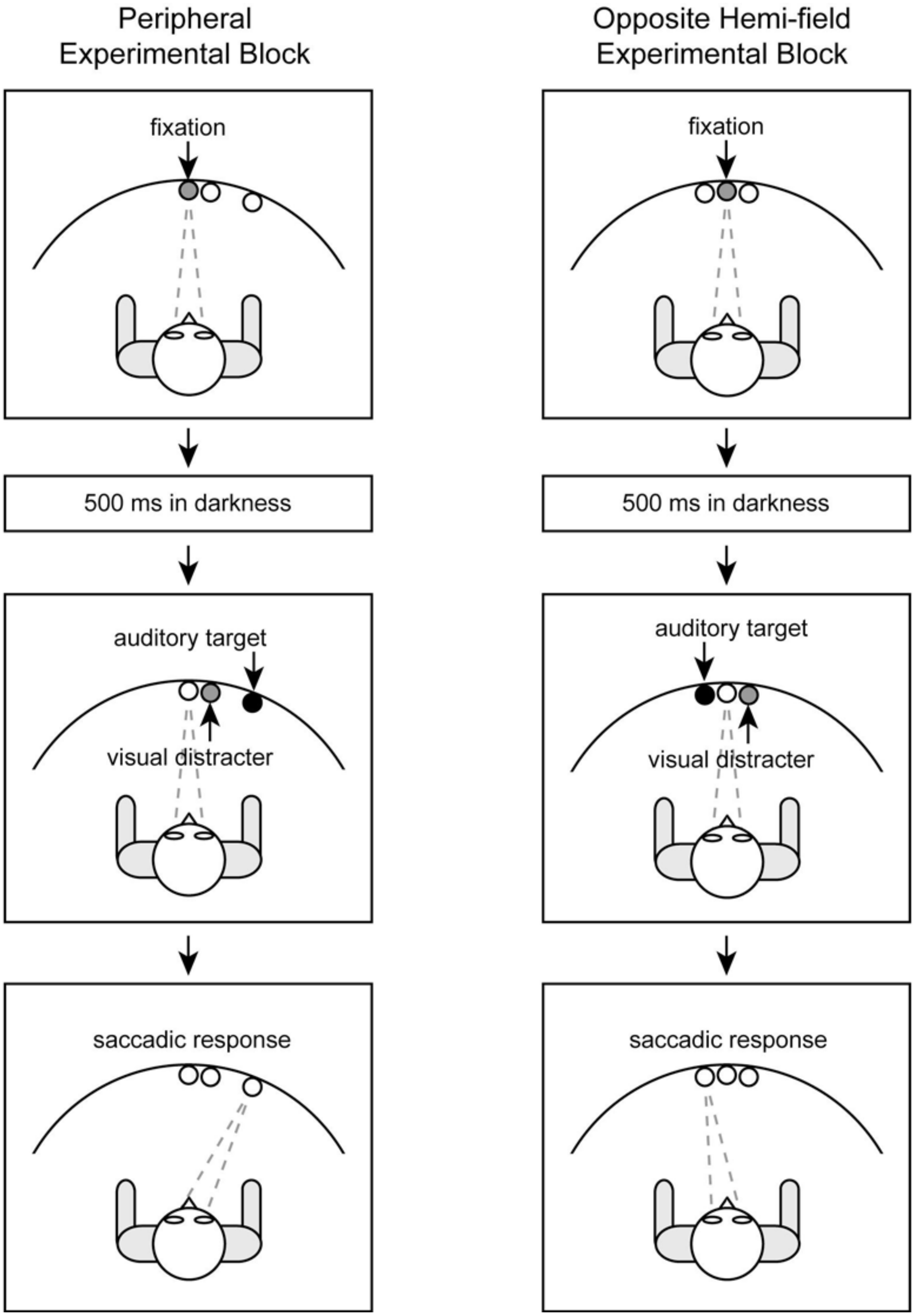


Figure 6.1: Schematic representation of experimental display peripheral and opposite hemifield experimental blocks.

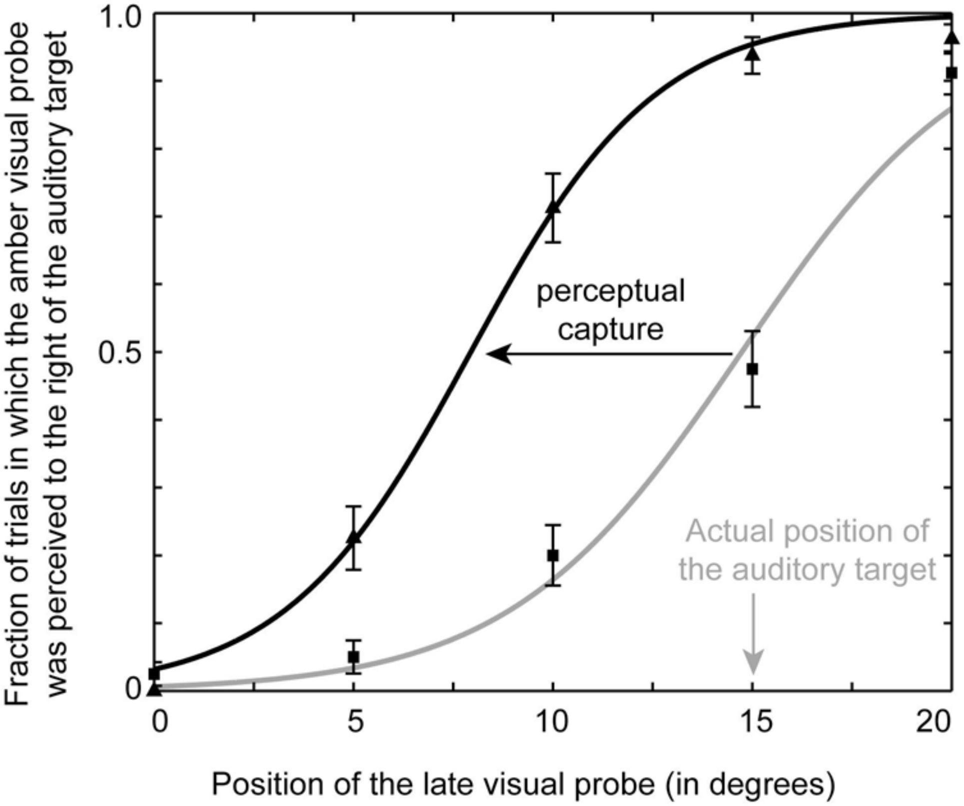


Figure 6.2: Perceptual localization of auditory targets influenced by visual distracters. Psychometric curves for the average all subject data for auditory an auditory target at 15 degrees with visual cues at 15 degrees (gray) and 5 degrees (black).

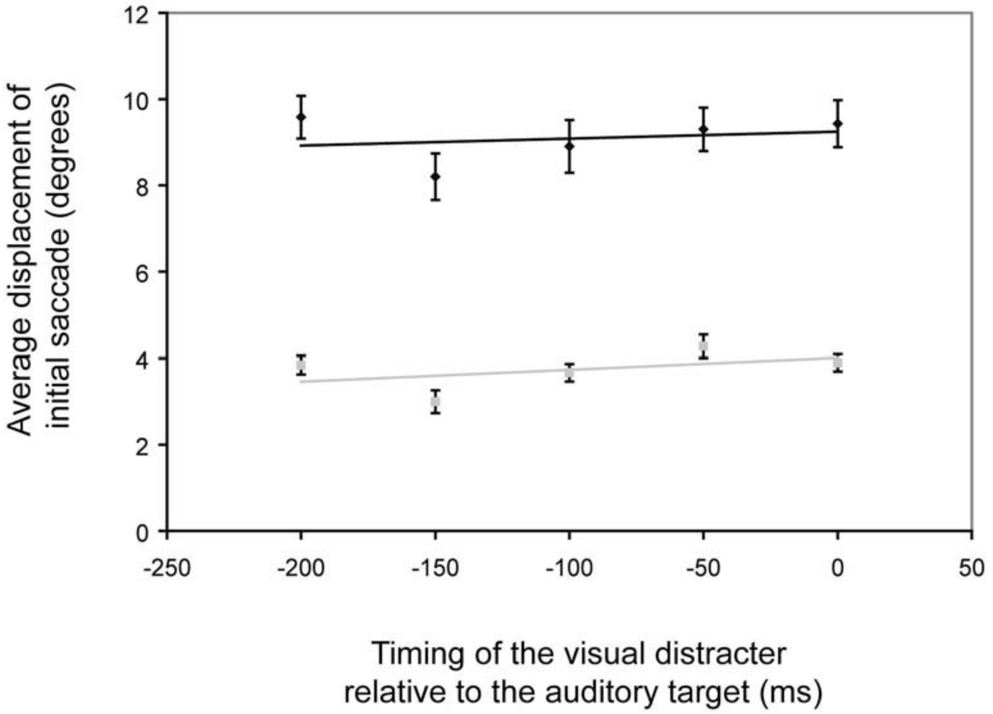


Figure 6.3: <u>Average initial saccade displacement towards an auditory target at 15 degrees accompanied by a 15 degree visual cue (black) or a 5 degree visual cue (gray) when the visual cue precedes the auditory target</u>.

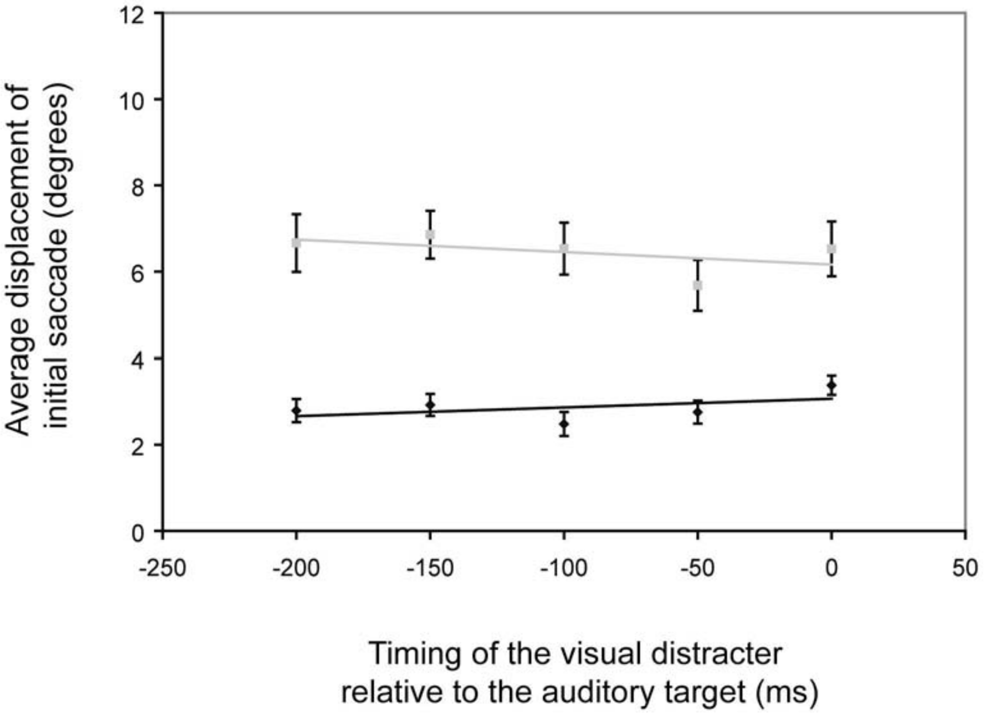


Figure 6.4: Average initial saccade displacement towards an auditory target at 5 degrees accompanied by a 5 degree visual cue (black) or a 15 degree visual cue (gray) when the visual cue precedes the auditory target.

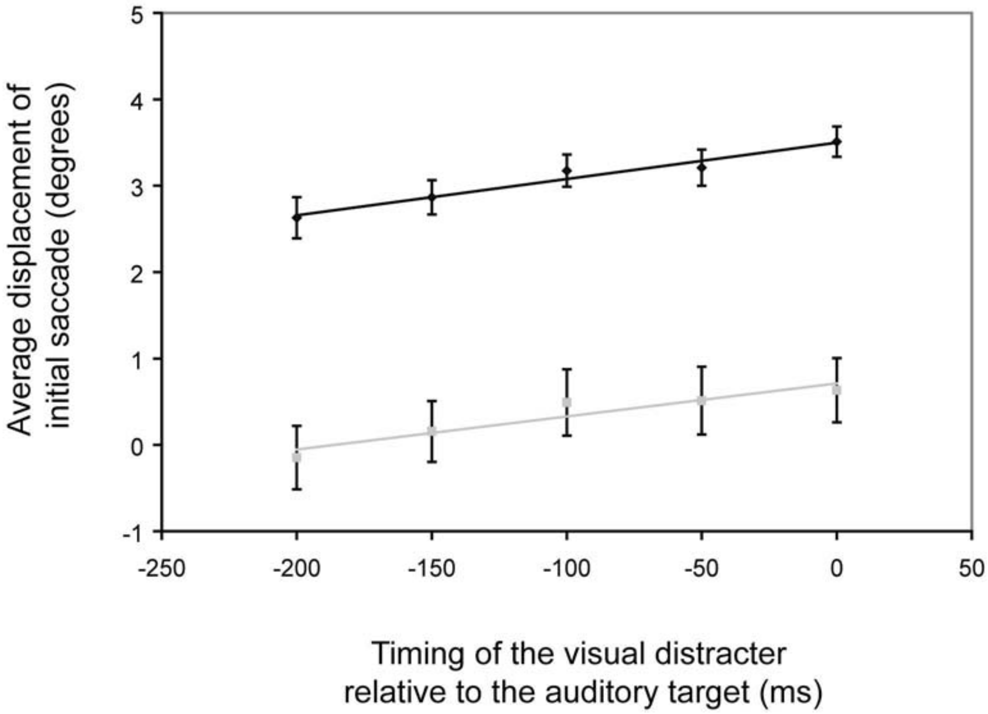


Figure 6.5: Average initial saccade displacement towards an auditory target at 5 degrees accompanied by a 5 degree visual cue in the same hemifield (black) or a 5 degree visual cue in the opposite hemifield (gray) when the visual cue precedes the auditory target.

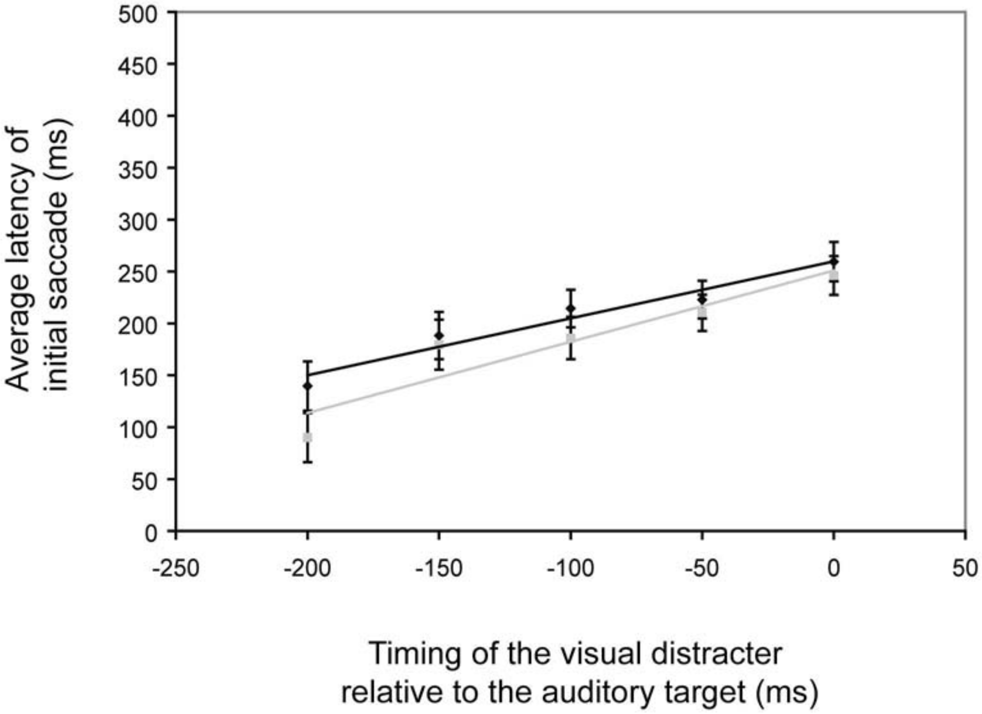


Figure 6.6: <u>Average latency of initial saccades towards an auditory target at 15 degrees accompanied by a 15 degree visual cue (black) or a 5 degree visual cue (gray) when the visual cue precedes the auditory target</u>.

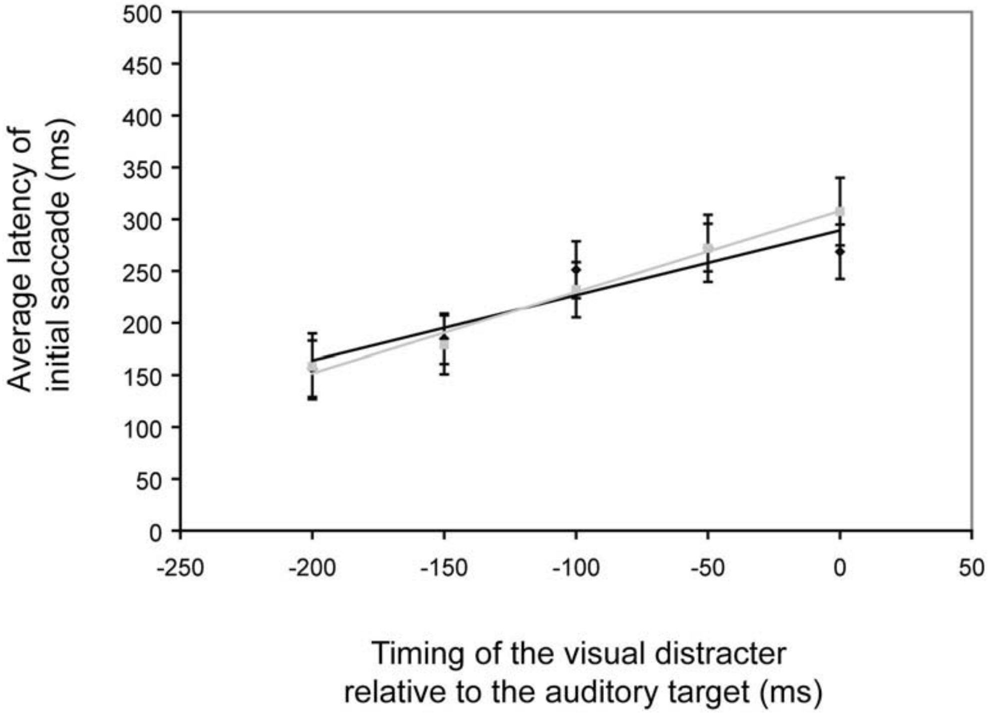


Figure 6.7: <u>Average latency of initial saccades towards an auditory target at 5 degrees accompanied by a 5 degree visual cue (black) or a 15 degree visual cue (gray) when the visual cue precedes the auditory target</u>.

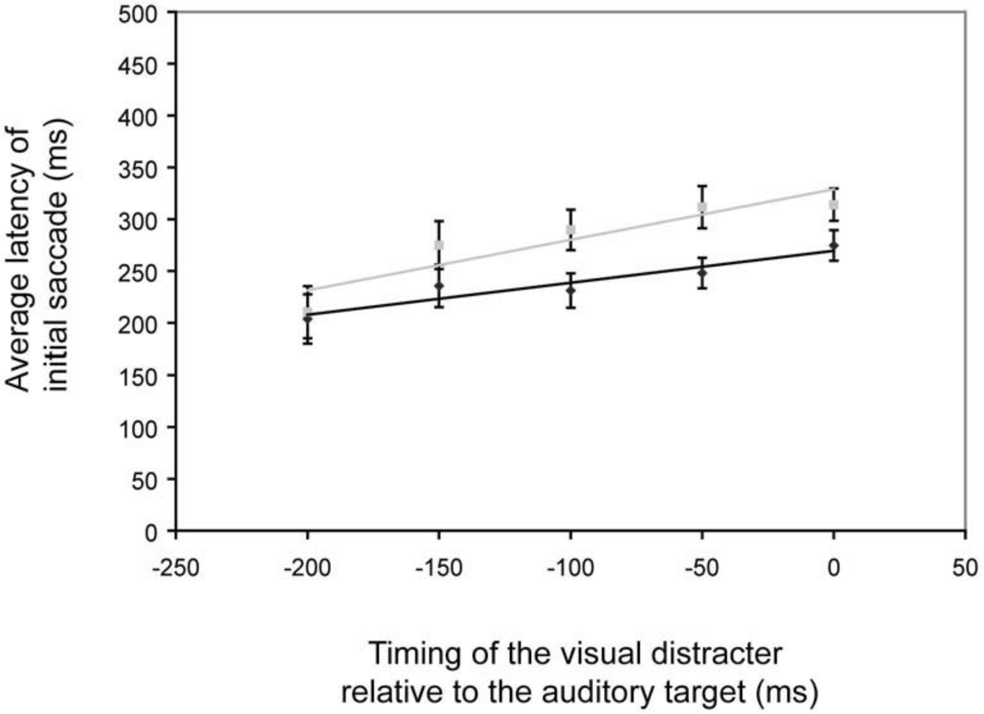


Figure 6.8: Average latency of initial saccades towards an auditory target at 5 degrees accompanied by a 5 degree visual cue in the same hemifield (black) or a 5 degree visual cue in the opposite hemifield (gray) when the visual cue precedes the auditory target.

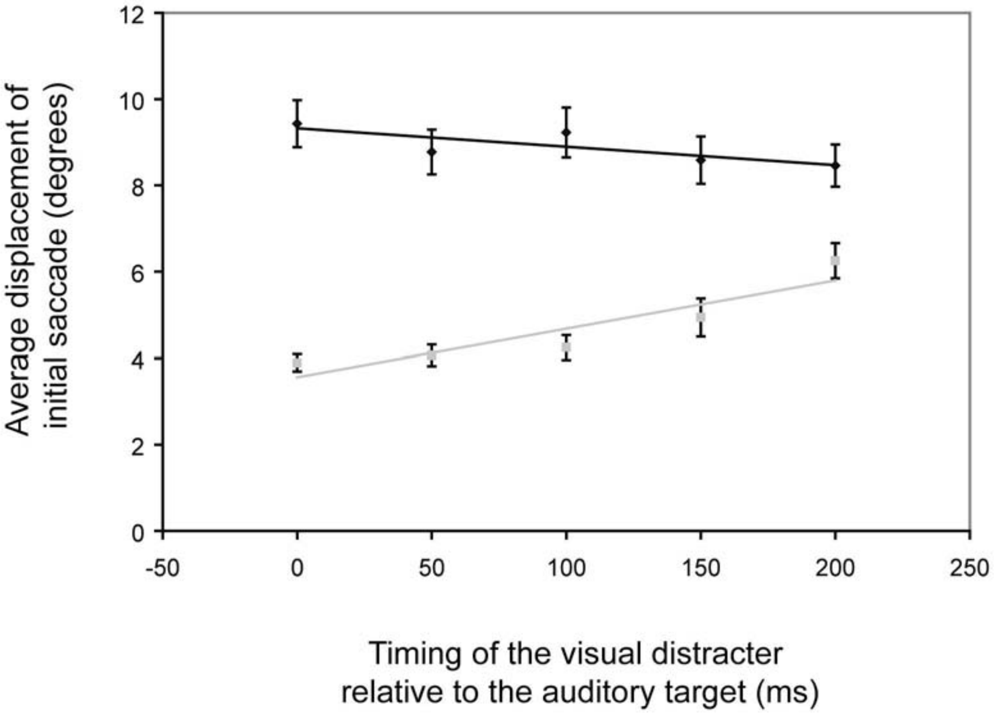


Figure 6.9: Average initial saccade displacement towards an auditory target at 15 degrees accompanied by a 15 degree visual cue (black) or a 5 degree visual cue (gray) when the visual cue follows the auditory target.

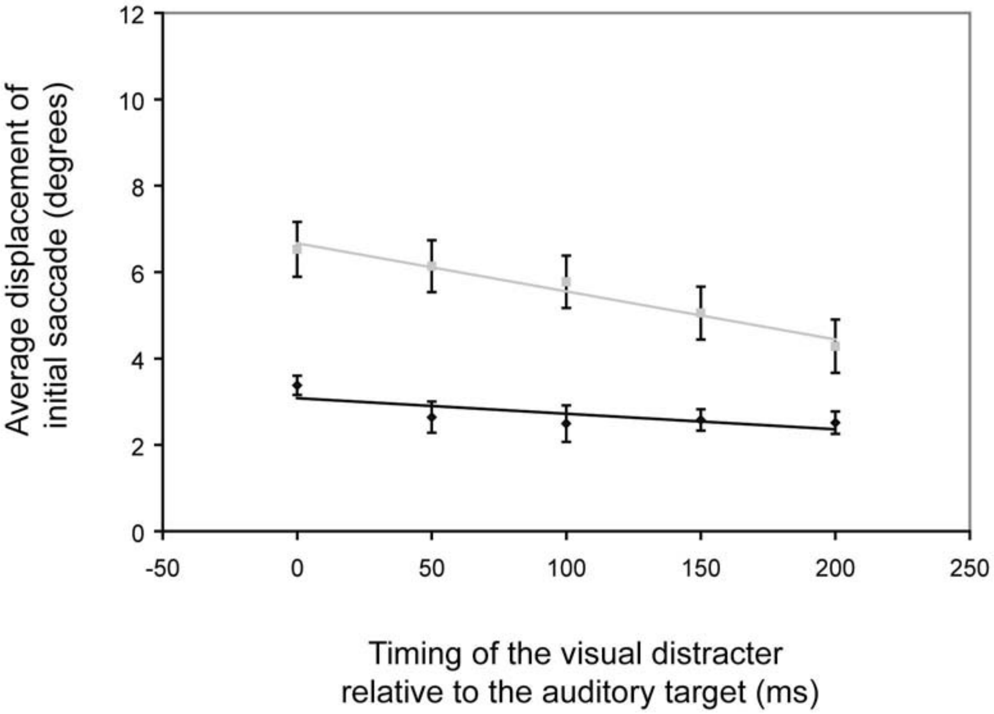


Figure 6.10: Average initial saccade displacement towards an auditory target at 5 degrees accompanied by a 5 degree visual cue (black) or a 15 degree visual cue (gray) when the visual cue follows the auditory target.

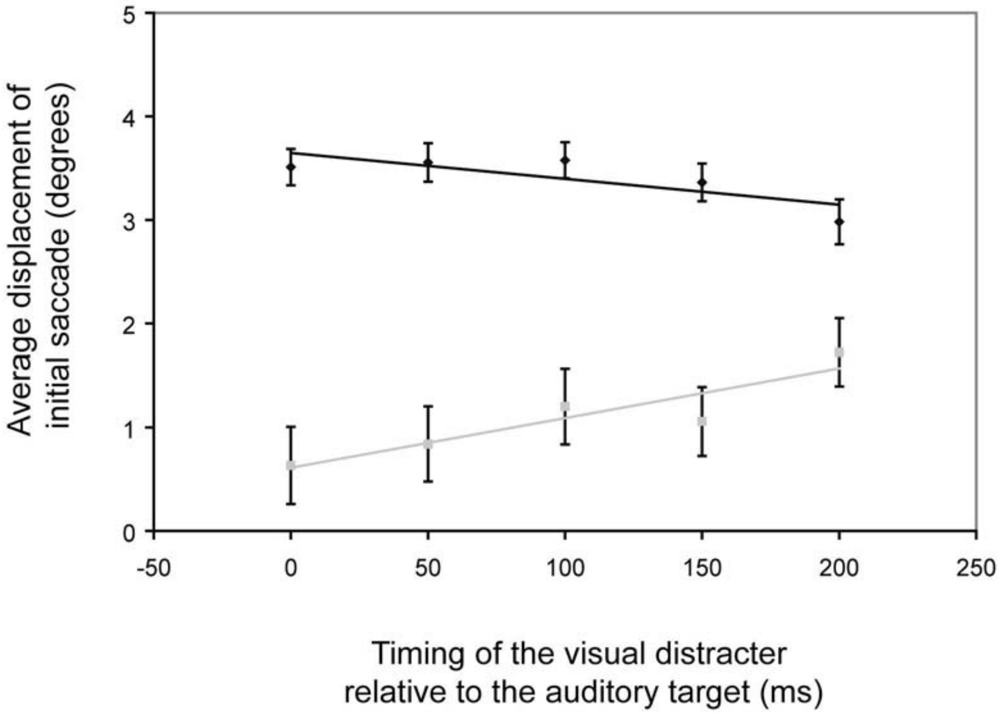


Figure 6.11: Average initial saccade displacement towards an auditory target at 5 degrees accompanied by a 5 degree visual cue in the same hemifield (black) or a 5 degree visual cue in the opposite hemifield (gray) when the visual cue follows the auditory target.

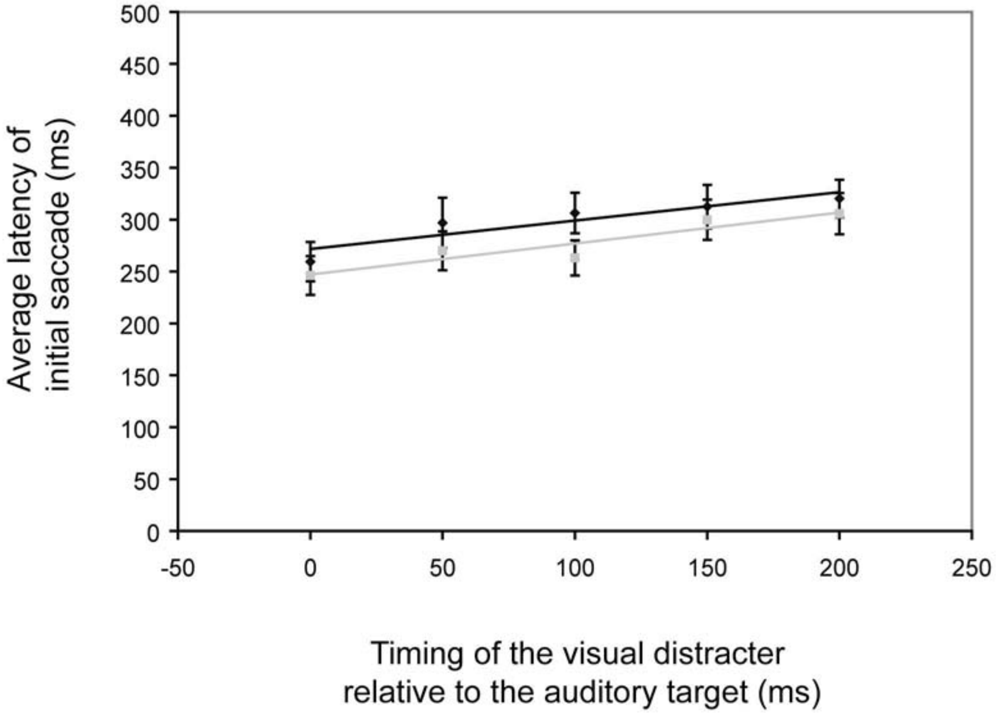


Figure 6.12: Average latency of initial saccades towards an auditory target at 15 degrees accompanied by a 15 degree visual cue (black) or a 5 degree visual cue (gray) when the visual cue follows the auditory target.

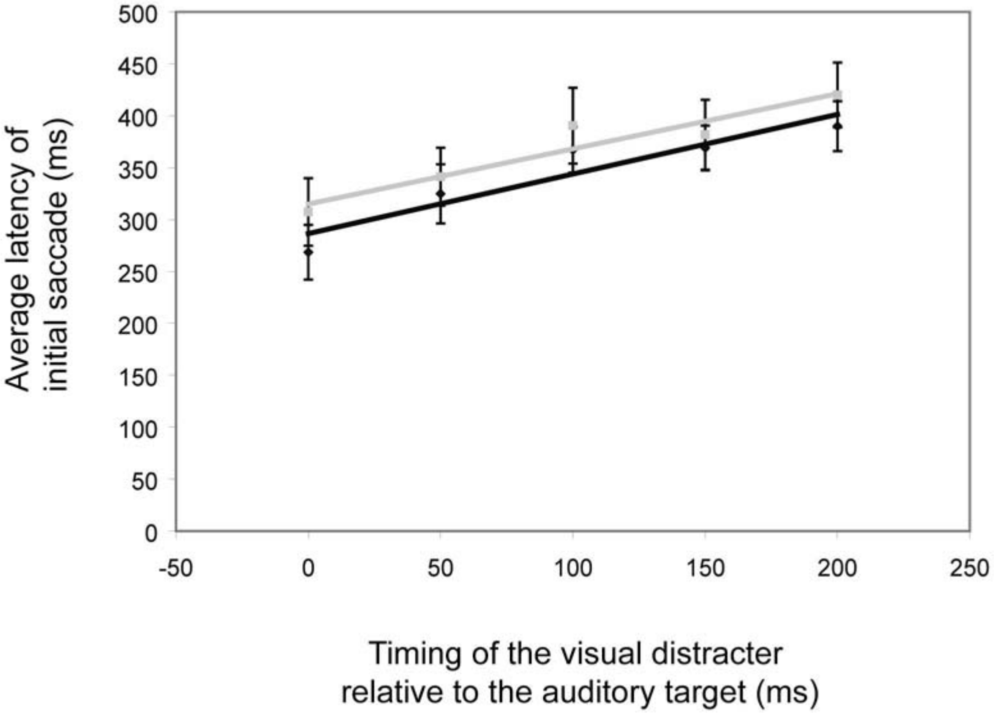


Figure 6.13: Average latency of initial saccades towards an auditory target at 5 degrees accompanied by a 5 degree visual cue (black) or a 15 degree visual cue (gray) when the visual cue follows the auditory target.

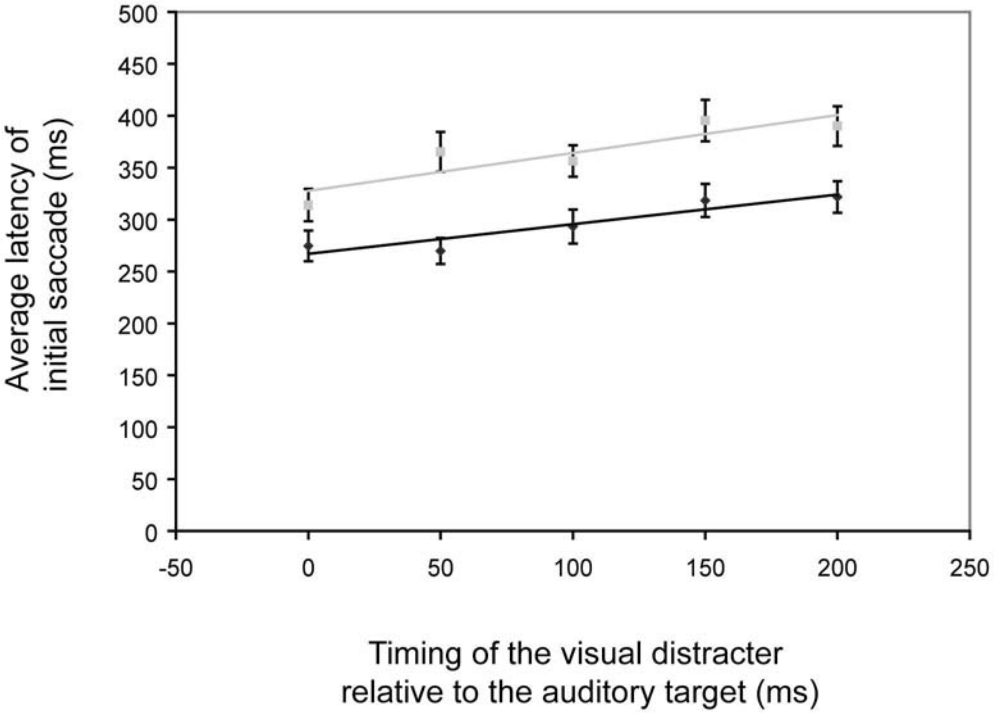


Figure 6.14: Average latency of initial saccades towards an auditory target at 5 degrees accompanied by a 5 degree visual cue in the same hemifield (black) or a 5 degree visual cue in the opposite hemifield (gray) when the visual cue follows the auditory target.

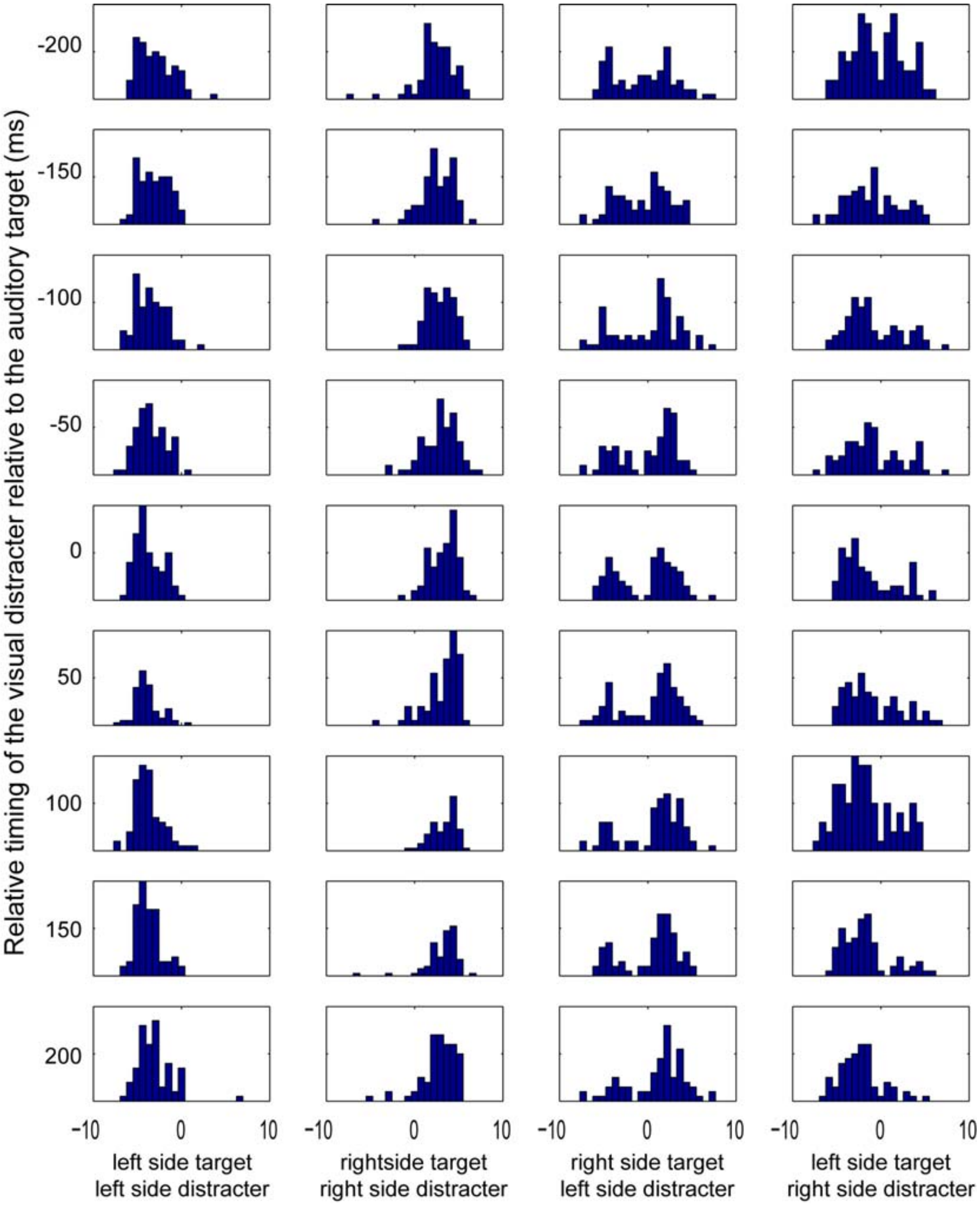


Figure 6.15: Histograms of initial saccade displacement for opposite hemifield experimental blocks.

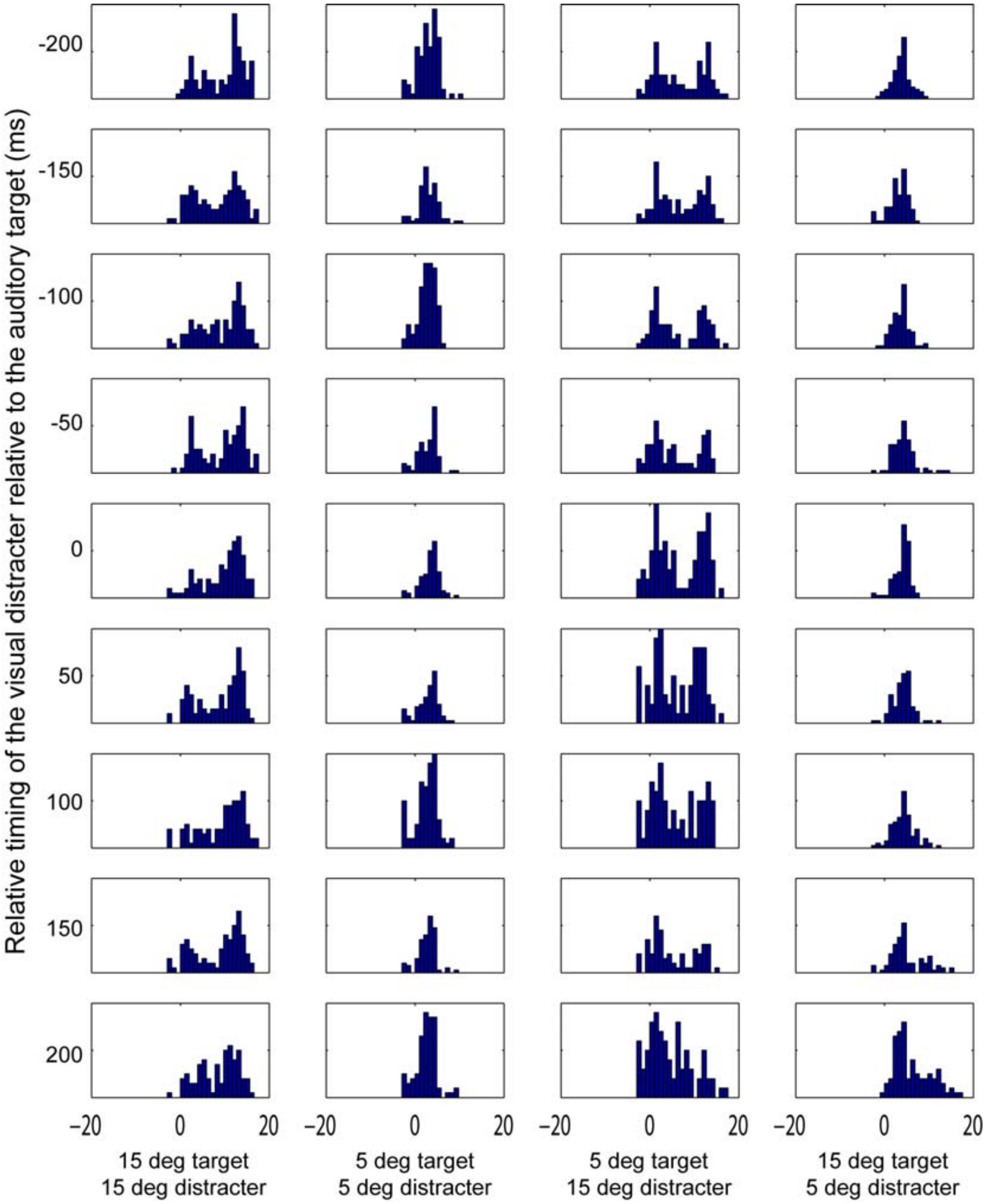


Figure 6.16: Histograms of initial saccade displacement for peripheral experimental blocks.

Chapter

Study 6: Repetition Suppression

Abstract

Saccadic response times (SRT) for eye movements towards visual or auditory target stimuli with subsequent colocalized distracters were measured and compared with single target SRTs. Experimental context was varied to include within-modality vs. cross-modal distracters, identical-duration vs. long-duration distracters, and identical-feature vs. color-distinct distracters. Detection times, reported via button press, were also measured for within-modality vs. cross-modality distracters. There was a strong context dependence of the SRT suppression and facilitation observed. When experimental blocks included cross-modal integration tasks, late visual distracters facilitated saccades to auditory targets while late auditory targets suppressed saccades to visual targets. Late within-modality distracters exhibited significant suppression on saccadic responses. We have termed this effect "repetition suppression." This effect disappeared when experimental context included only within-modality, identical-feature stimuli, but was observed for color-distinct distracters. These results suggest a general time-window-of-integration model involving both stimulus-specific as well as context-dependent parameters of integration and provide another example of a postdictive phenomenon in which late exposure to cognitively irrelevant stimuli affects the experience of prior stimuli.

Introduction

Research on multi-sensory perception has shown that saccadic response time (SRT) to a visual target is facilitated by the synchronous presentation of an irrelevant auditory stimulus at the same location (Harrington and Peck, 1998). This facilitation is time dependent, and only occurs for stimuli in close temporal alignment: with a delay in the onset of the auditory stimulus, latency was found to increase (Frens, van Opstal, and Van der Willigen, 1995). Neurophysiological studies of the superior colliculus (Meredith, Nemitz, and Stein, 1987) have provided evidence for enhancement of multi-modal neural activity with high temporal overlap in the peak discharge periods evoked by individual modalities, as well as decay of this enhancement with increased temporal disparity, and even inhibition of multi-modal neurons associated with large temporal disparities between those peak discharge periods. As such, it has been shown that despite its confirmatory positional information, the late presentation of an auditory distracter can increase the latency to a visual target.

Models have been proposed for this cross-modal interaction (Colonius and Arndt, 2001; Colonius and Diederich, 2004), in which stimuli are individually detected within modality, and are then integrated in a separate stage over some extended time window whose length is dependent on the any number of factors related to the modalities involved, the quality of the stimuli, and the perceptual context. These models of multi-modal integration rely on initial segregation of within-modality stimulus perception, and keep stimulus representation distinct, prior to some integration process. They fail to address the role of delayed, colocalized distracter stimuli within modalites. What is the effect of a spatially identical late visual distracter on saccades to a visual target?

Studies of the “gap effect” (Fischer and Ramsperger, 1984; Saslow, 1967), in which saccadic latencies are facilitated when a visually fixated stimulus is extinguished simultaneous to or just prior to the appearance of the visual target, suggest that fixation involves saccadic inhibition. This inhibition has been demonstrated in the rostral pole of the superior colliculus (Munoz and Wurts, 1993a, 1993b) and has been shown to correlate with saccade latency (Doris and Munoz, 1995; Doris, Pare, and Munoz, 1997). It has been postulated that the release of this saccadic inhibition results in the facilitation observed in the gap effect. Taken in another way, the gap effect is an example of saccadic inhibition by an existing visual target.

Prior presentation of a spatially uninformative visual cue has been shown to exhibit facilitation for visually directed saccades when presented from the same position or in the same visual hemifield, when compared to presentation in the opposite hemifield (Jonides 1981; Posner and Cohen 1984; Yantis and Jonides 1984), though this effect is short lived (200–300 ms). When stimulus onset asynchrony (SOA) is increased beyond 300 ms, the opposite effect is observed: increased latency for ipsilateral targets (Posner and Cohen, 1984). This later inhibitory effect can last for several seconds (Tassinari *et al*., 1994) and does not require the execution of eye movements. Posner and Cohen (1984) suggested that there is a brief attentional shift toward the distracter, and that when attention moves on, it leaves an inhibitory tag behind. So, attentional shifts towards a premature distracter can produce positional facilitation if the SOA is small, but if sufficient time allows for the distracter to be disregarded, suppression is observed instead.

For a target presented prior to its spatially identical distracter, there is no reason to predict such suppression. If target presentation evokes a shift in attention to the location of the target, and observers are attempting to execute a saccade to that location, then subsequent presentation of a distracter at the same location should evoke no further shift in positional attention. By this reasoning, saccade latency should be unaffected by late distracter position.

Experiments

Experiment 1

We began by measuring the saccade latency of visual and auditory saccades to a single cue, a repeated cue, and a cue in one modality followed by a second cue in another.

Participants: One author and four naive observers with normal or corrected to normal vision were used. The same observers were used in all experiments.

Apparatus: All experiments were performed on a Windows computer running Matlab (MathWorks Inc., Natick, MA) and the Psychophysics Toolbox extensions (Brainard, 1997). Observers sat in a completely dark room at the center of an arced hoop (radius = 87 cm) with presentation module ports spaced in five degree increments ranging 50 degrees to the left or right of center. Each audio-visual presentation module consisted of a three-color red/green/amber 5 mm LED mounted in a cutout plate directly in front of the center of a 2 inch speaker. Modules affixed to the inside of the hoop and positioned LEDs at a viewing distance 80.5 cm from the observer. The presentation hoop was operated via a custom built control box which offered simultaneous independent control of LEDs and speakers with 100 µs timing. The control box was programmed prior to each trial via a serial connection with the experimental computer. Observers responded by executing saccadic eye movements. Eye-tracking was performed using a head-mounted Eyelink II system which communicated with the experimental computer via a direct internet connection. Prior to the start of experimental blocks, observers were familiarized with the eye-tracking system and the hoop.

Stimuli: All experimental stimuli were 50 ms in duration and consisted of either a visual cue (red LED, 25.1 cd/m^2) or an auditory cue (full spectrum white noise, 67.7 dB). Each experimental trial began with the observer fixating a green LED (24.3 cd/m^2) at a position 0 degrees with respect to the head. Observers pressed a button on a hand-held controller in order to recalibrate the eye-tracking system and signal their readiness. After a delay of 500 ms the fixation LED extinguished, and after another 500 ms in complete darkness stimulus presentation began. In all trials an initial (target) stimulus, either auditory or visual, was then presented at a position 5 degrees to the left or right of fixation, followed by a brief inter-stimulus interval (ISI) of 50, 100, or 150 ms. In one quarter of trials, observers were then presented with an identical second (distracter) stimulus at the same location. In one-quarter of trials, observers were then presented with the alternate (distracter) stimulus (visual for auditory, auditory for visual) at the same location.

Task: Observers were instructed to keep their gaze fixed straight ahead at the fixation target until the presentation of the initial stimulus, and then to look at the source of that stimulus as quickly and as accurately as they could. To ensure that observers began each trial with a neutral gaze position, observers pressed a button while fixating on the green LED at the start of the trial. Trials only progressed if the eye-tracking system reported the observer’s eye position consistent with the fixation target. Experimental trials were blocked by inter-stimulus interval and by modality of the initial stimulus, and before each block observers were informed of the modality of the initial stimulus to expect. Observers were also informed that there would sometimes be a second stimulus following the first and they were instructed to ignore it. The position of the target, the presence of absence of the distracter, and modality of the distracter were randomized across trials in blocks of 160 (20 trials each: 2 target positions, 2 distracter modalities; 40 trials each: 2 target positions, no distracter).

Analysis: Prior to the start of each experimental block, the eye-tracker was calibrated for both eyes, and the eye which provided the most accurate calibration results was tracked throughout the experiment. Eye position was recorded at 2 ms sample intervals throughout each trial. Data for each trial was then analyzed in MATLAB. A saccade event was defined as any change in eye position in which the velocity exceeded 0.3 degrees/second or the acceleration exceeded 4000 degrees/second^2. Trials in which observers executed a saccade before the presentation of the target were eliminated from the analysis. For each of the remaining trials, the onset time of the first saccade subsequent to target presentation was considered.

Results: For visual targets, there was, in every case, a subtle increase in the average latency with the addition of a distracter (figure 7.1 and figure 7.2). While this increase failed to reach significance within most individual blocks, across all ISIs, the average latency to the visual target increased significantly from 190.5 ± 3.0 ms for the target alone to 205.1 ± 5.7 ms with a late visual distracter ($t(1768) = 2.49$, $p = 0.013$; unpaired t-test) and 204.9 ± 5.3 ms with a late auditory distracter ($t(1765) = 2.55$, $p = 0.011$; unpaired t-test).

For auditory targets, there was, in every case, an increase in the average latency with the addition of a late auditory distracter (figure 7.3), but a decrease in average latency with the addition of a late visual distracter (figure 7.4). While these differences again failed to reach significance within half of the individual blocks, across all ISIs, the average latency to the auditory target increased significantly from 253.4 ± 3.5 ms for the target alone to 251.2 ± 7.0 ms with a late auditory distracter (t(1748) = 3.98, $p = 7.27 \times 10^{-5}$; unpaired t-test) and decreased significantly to 241.5 ± 4.7 ms with a late visual distracter (t(1751) = 2.01, p = 0.045; unpaired t-test).

In taking the average saccade latency across all trials in the double stimulus conditions, we have included in our analysis many instances in which the observer responded with a targeting saccade before the presentation of the late distracter even began. As such, the power of this comparison is diluted by the inclusion of what are essentially target-only trials in the double stimulus groups. In order to avoid this, we performed this analysis again, only considering trials in which the observer's first saccade occurred late enough to allow for the onset of the distracter presentation. Thus, in blocks with an ISI of 50 ms, only trials in which the observer's first saccade came after 100 ms were considered. In this way, we aimed to compare corresponding "later" regions of the saccade latency distributions across conditions.

Looking only at "late" saccades, the average latency to the visual target was 219.0 ± 4.5 ms. This increased to 244.3 ± 8.5 ms with a late visual distracter (t(1088) = 2.88, p = 0.004; unpaired t-test) and 245.5 ± 8.0 ms with a late auditory distracter (t(1076) = 3.11, p = 0.0019; unpaired t-test). For "late" saccades to the auditory target, the average latency was 270.5 ± 3.7 ms for the target alone. This increased to 302.0 ± 7.6 ms with a late auditory distracter (t(1520) = 4.23, $p = 2.49 \times 10^{-5}$; unpaired t-test) and decreased to 256.5 ± 5.0 ms with a late visual distracter (t(1524) = 2.23, p = 0.026; unpaired t-test).

Discussion: Presentation of a late auditory distracter increased latency for visual-target saccades, similar to what some previous studies have reported (Kirchner and Colonius, 2004, 2005). This agrees with the physiological data as well (Meredith, Nemitz, and Stein, 1987). Presentation of a late visual distracter decreased latency for auditory-target saccades. If we accept a two-stage model of multi-modal processing (Colonius and Arndt, 2001; Colonius and Diederich, 2004), this could easily be explained by a relatively longer localization time for the auditory stimulus as compared with the later visual stimulus.

Presentation of a late within-modality distracter resulted in saccadic suppression for both visual and auditory targets. Attempts to explain this unexpected phenomenon will be reserved for the general discussion.

Experiment 2

The latency differences we observed in experiment 1 may have been associated with the computation of position in the saccade localization task observers were asked to perform. Conversely, the latency differences could merely be a function of the detection time necessary under the different stimulus conditions. To examine this, we asked observers to perform the same set of experiments again, but instead of responding with a saccade towards the target, observers were instructed to maintain fixation throughout, and respond with a button press as soon as they detected the first stimulus.

Stimuli: All experimental stimuli and block design were identical to the first experiment.

Task: Observers were instructed to hold their gaze fixed straight ahead and press a button on the handheld controller as soon as they detected the first stimulus.

Analysis: The timing of button presses relative to stimulus presentation was recorded. Trials in which observers responded before the presentation of the target were eliminated from the analysis. For each of the remaining trials, the time of button press was considered.

Results: The button-press data were not consistent with the effects observed for saccades. Across all ISI conditions, the average response time for detection of the visual target was 311.6 ± 4.4 ms. This decreased, though insignificantly, to 302.3 ± 3.5 ms with a late visual distracter (t(1438) = 1.39, p = 0.16; unpaired t-test; figure 7.5) and increased, though not significantly, to 317.6 ± 9.8 ms with a late auditory distracter (t(1438) = 0.65, p = 0.52; unpaired t-test; figure 7.6). These results changed only marginally if we considered only "late" saccades. The average "late" response time for detection of the visual target was 314.6 ± 4.4 ms. This decreased to 307.4 ± 3.1 ms with a late visual distracter (t(1416) = 1.10, p = 0.27; unpaired t-test) and increased to 319.6 ± 9.8 ms with a late auditory distracter (t(1422) = 0.53, p = 0.59; unpaired t-test).

Average response time for detection of the auditory target was 270.0 ± 5.4 ms for the target alone. This increased, though not significantly, to 281.1 ± 6.5 ms with a late auditory distracter (t(1438) = 1.27, p = 0.20; unpaired t-test; figure 7.7) and decreased negligibly to 267.67 ± 6.5 ms with a late visual distracter (t(1438) = 0.25, p = 0.80; unpaired t-test; figure 7.8). Again, these results changed only marginally if we considered only "late" saccades. The average "late" response time for detection of the auditory target was 281.0 ± 5.8 ms. This increased to 293.4 ± 7.0 ms with a late visual distracter (t(1293) = 1.29, p = 0.20; unpaired t-test) and remained essential unchanged at 281.1 ± 6.5 ms with a late auditory distracter (t(1286) = 0.01, p = 0.99; unpaired t-test).

Discussion: Though response time to an auditory target with an auditory distracter was increased in this experiment, the pattern of results was inconsistent with the visual-motor task in the first experiment. The suppression effect we observed for repeated stimuli in the first experiment is likely not related to stimulus detection, but arises somewhere in the process of localization for orienting.

Experiment 3

Within both audition and vision, the repetition of the stimulus increased saccade latency. With such short inter-stimulus intervals there may be the potential for the repeated target to be processed as a single target that spans a larger time window. In order to investigate this possibility, we measured saccade latency for longer visual and auditory targets.

Stimuli: In experimental blocks using a visual target, presentation conditions were as follows: In half of the trials, a single visual target appeared exactly as in the first experiment. In one quarter of the trials, observers were presented with the same visual stimulus twice, exactly and in the first experiment. In the remaining quarter of the trials, observers were presented with a single “long” visual stimulus that began at the same time as the first of the repeated stimuli and ended at the same time as the second. Essentially, this condition kept the target LED through what would have been the inter-stimulus interval. Experimental blocks using auditory targets paralleled this, with the white noise stimulus continuing to play from the onset of the first of the repeated stimuli until the offset of the second.

Results: No combination of stimulus conditions showed any significant difference in saccade latency. The average latency to a single 50 ms visual target was 164.7 ± 1.8 ms. This was virtually unchanged at 164.4 ± 2.6 ms with a late visual distracter (t(1781) = 0.08, p = 0.94; unpaired t-test; figure 7.9) and increased insignificantly to 168.1 ± 2.7 ms with the “long” visual stimulus (t(1783) = 1.08, p = 0.28; unpaired t-test; figure 7.10). Considering only late saccades yields essentially the same results. Across “late” saccades, the average latency to a single 50 ms visual target was 184.4 ± 2.8 ms. This was virtually unchanged at 181.1 ± 3.9 ms with a late visual distracter (t(1030) = 0.70, p = 0.48; unpaired t-test) and 186.1 ± 4.4 ms with the “long” visual stimulus (t(1044) = 0.34, p = 0.74; unpaired t-test).

The results for auditory targets showed modest increases in latency for both the double stimulus and the “long” stimulus. The average latency to a single 50 ms auditory target was 270.3 ± 5.8 ms. This increased, though not significantly, to 281.5 ± 8.0 ms with a late auditory distracter (t(1560) = 1.13, p = 0.26; unpaired t-test; figure 7.11) and 285.2 ± 8.3 ms with the “long” auditory stimulus (t(1548) = 1.49, p = 0.13; unpaired t-test; figure 7.12). Considering only late saccades yielded essentially the same results. Across “late” saccades, the average latency to a single 50 ms auditory target was 320.6 ± 6.5 ms. This increased to 338.7 ± 8.8 ms with a late visual distracter (t(1186) = 1.64, p = 0.10; unpaired t-test) and 340.8.1 ± 9.0 ms with the “long” visual stimulus (t(1181) = 1.82, p = 0.07; unpaired t-test).

Discussion: In this context, without multi-sensory perception actively engaged, we failed to see the same within-modality suppression observed for vision. While some modest suppression was observed within the auditory domain for both the double stimulus, and the “long” stimulus, it failed to reach the significance of the first experiment. These results suggest that the presence or absence of multi-modal stimulus presentation within the experimental block may provide a contextual modulation of the time window associated with saccadic localization decisions.

Experiment 4

In experiment 3, trial blocks contained only a single stimulus modality. In fact, trial blocks contained a single type of stimulus that varied across trials only in length or number of presentations. Under these circumstances we failed to observe the same clear effects from the first experiment. Perhaps the repetition suppression previously observed is a result of the brain’s attempt to perform the localization task within a context that demands integration of separate representations of position information. If that is the case, then perhaps we can elicit the same effect using a second target with a distinct feature difference from the first. In order to investigate this possibility, we measured saccade latency for a red visual target with a subsequent green visual distracter.

Stimuli: Presentation conditions were as follows: In half of the trials, a single red visual target appeared exactly as in the first experiment. In one-quarter of the trials, observers were presented with the same red visual stimulus twice, exactly as in the first experiment. In the remaining quarter of the trials, observers were presented with a single red visual and a second green visual distracter. All relative timings were identical to the first experiment.

Results: In this experimental context, saccades to red visual targets followed by identical red distracters showed no difference in latency. Saccades to red visual targets followed by green visual distracters, on the other hand, showed a significant increase in average latency. The average latency to a single 50 ms visual target was 154.6 ± 1.9 ms. This was virtually unchanged at 153.7 ± 2.9 ms with an identical late visual distracter ($t(1421) = 0.26$, $p = 0.79$; unpaired t-test; figure 7.13). With a distinct (green) late visual distracter, average latency increased to 163.3 ±

3.9 (t(1427) = 2.24, p = 2.5 x 10^{-2}; unpaired t-test; figure 7.14). Considering only late saccades yields essentially the same results. Across "late" saccades, the average latency to a single 50 ms visual target was 178.4 ± 3.1 ms. This was virtually unchanged at 176.9 ± 4.8 ms with an identical late visual distracter (t(718) = 0.27, p = 0.79; unpaired t-test). With a distinct (green) late visual distracter, average latency increased to 195.4 ± 6.6 (t(733) = 2.65, p = 8.2 x 10^{-3}; unpaired t-test).

Discussion: Within this experimental context, visual saccade suppression with identical repetition was not observed, however repetition with a feature-distinct distracter did exhibit suppression. It appears that the experimental context strongly affects the degree to which repetition suppression is observed.

General Discussion

Saccadic responses to visual and auditory targets are suppressed by identical late distracters in a context in which observers are challenged with multi-modal stimuli. For an explanation of this phenomenon, we borrow from Colonius and Diederich's (2004) time-window-of-integration (TWIN) model. The model, in brief, postulates that there is an initial stage during which sensory information from separate modalities is processed independently, and a second stage in which those independent percepts are integrated in the preparation of an ocular motor response. Integration only occurs if the independent processes of the first stage all terminate within a certain time interval. Furthermore, the dynamics of the integration are a function of the spatial and temporal proximity of the percepts, such that coincidence results in facilitation and disparity results in suppression. The model allows that the time window involved is variable, depending on the stimuli and context.

We propose that this TWIN model is more general, applying not only to perceptual events in different modalities, but to different stimulus representations within modality as well. While it is convenient when explaining multi-sensory phenomena to presume segregation between unimodal and multimodal processing, we instead propose the following. The initial step in visual motor processing involves the generation of object specific motor-saccade representations of all events. Those alternate, potential eye movement plans are then integrated, regardless of modality, over some time window that depends on stimulus condition and perceptual context. In this model, competing representations of the repeated target (with temporal disparity between the representations of the two presentations) results in suppression if the time window of integration is sufficiently large. Perhaps with context-specific suppression by repetition, we are observing how the time window of integration changes as a function of experimental context.

When all experimental stimuli in a block are presented in a single modality, and all stimulus features are essentially the same (experiment 3), attention can be concentrated in one domain on a specific combination of features, and the time window of integration is potentially minimized. Within this kind of experimental context we found no effect of late stimuli in the visual domain. We still observed some modest effect (not significant) in the auditory domain, and this could be evidence of longer integration times implicit for auditory localization. With the addition of feature variation to the experimental context (experiment 4), there is an added component of representational disparity. If the facilitation or suppression that results from integration is a function of disparity between the representations, then perhaps the distinct-color distracter exhibits a greater suppressive effect within modality than an identical distracter. For experimental conditions in which multi-sensory demands were consistently placed upon the observer (experiment 1), we suggest that the time window for integration was sufficiently large to magnify the effect of feature-identical distracters that were only temporally disparate. In this context, suppression by repetition was observed.

The findings of suppression of visual-evoked saccades by subsequent auditory targets and facilitation of auditory-evoked saccades by subsequent visual targets are not unique (Meredith, Nemitz, and Stein, 1987; Kirchner and Colonius, 2005). These results may at first seem to contradict the model: the later stimuli, provided across the same temporal disparity and modal disparity should produce the same effect, at least if the delay between stimulus presentations and the forming of saccade motor-plan representations is the same for both stimuli. But it is not. While detection of auditory targets (as reported by button press) is traditionally faster than for visual targets, auditory-evoked saccade reaction times are much larger than those for visual-evoked saccades. Sounds are detected faster, but it takes much longer to form a visual-motor plan for orienting. In cross-modal stimulus conditions, a sound may be detected before a synchronously presented light, but the formation of its orienting saccade likely occurs later in the superior colliculus. Thus, the saccade program of a late auditory distracter may occur much later than that of a just-prior visual target, while the saccade program of a late visual stimulus could even occur simultaneous with the saccade program of its just prior auditory target.

This area needs further study to determine the specific parameters of experimental presentation and context that determine the relevant integration windows involved, but provides an interesting and surprising finding that must be explained by future models.

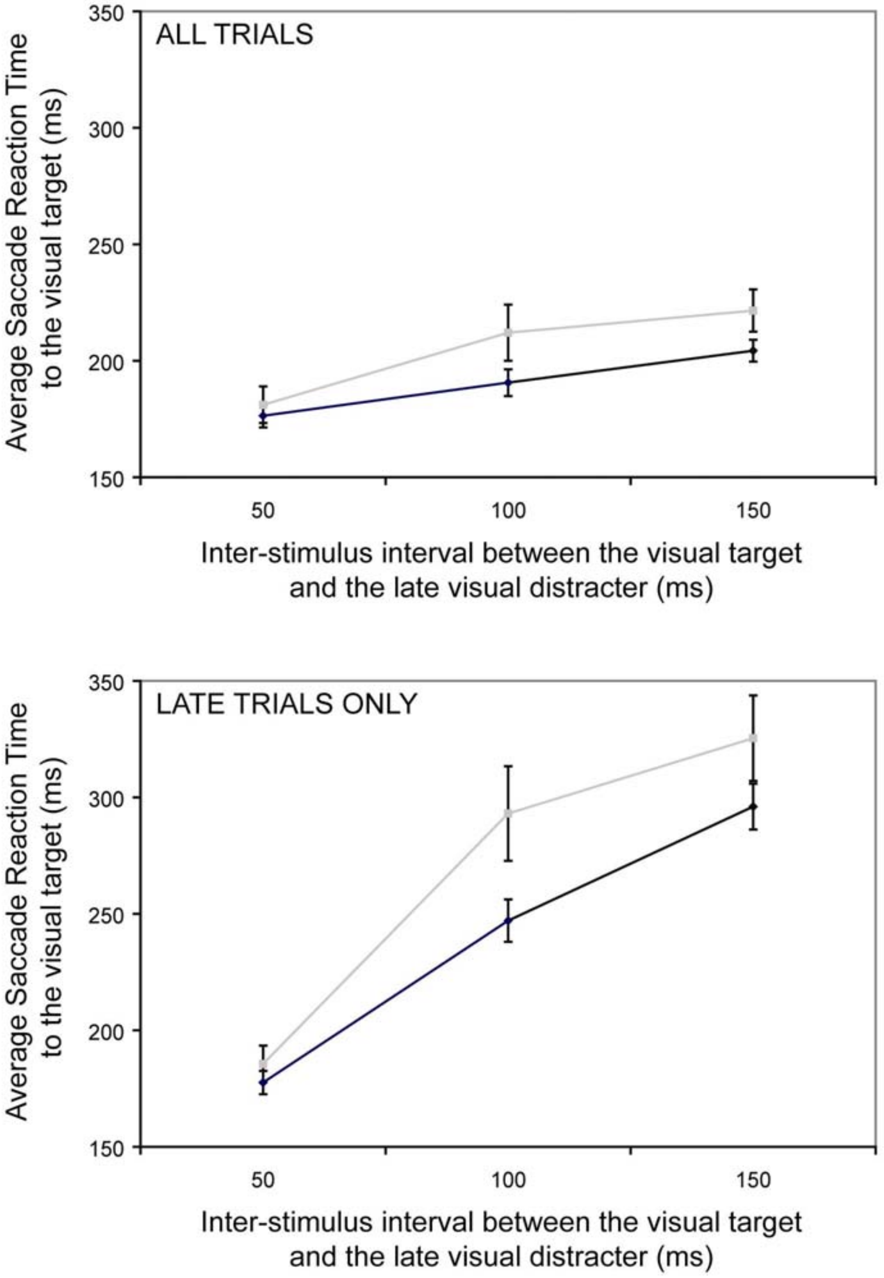


Figure 7.1: <u>Average saccade reaction time (SRT) to a visual target presented alone (black) or with a subsequent, identical visual distracter (gray) in experimental blocks including late visual and auditory distracters</u>.

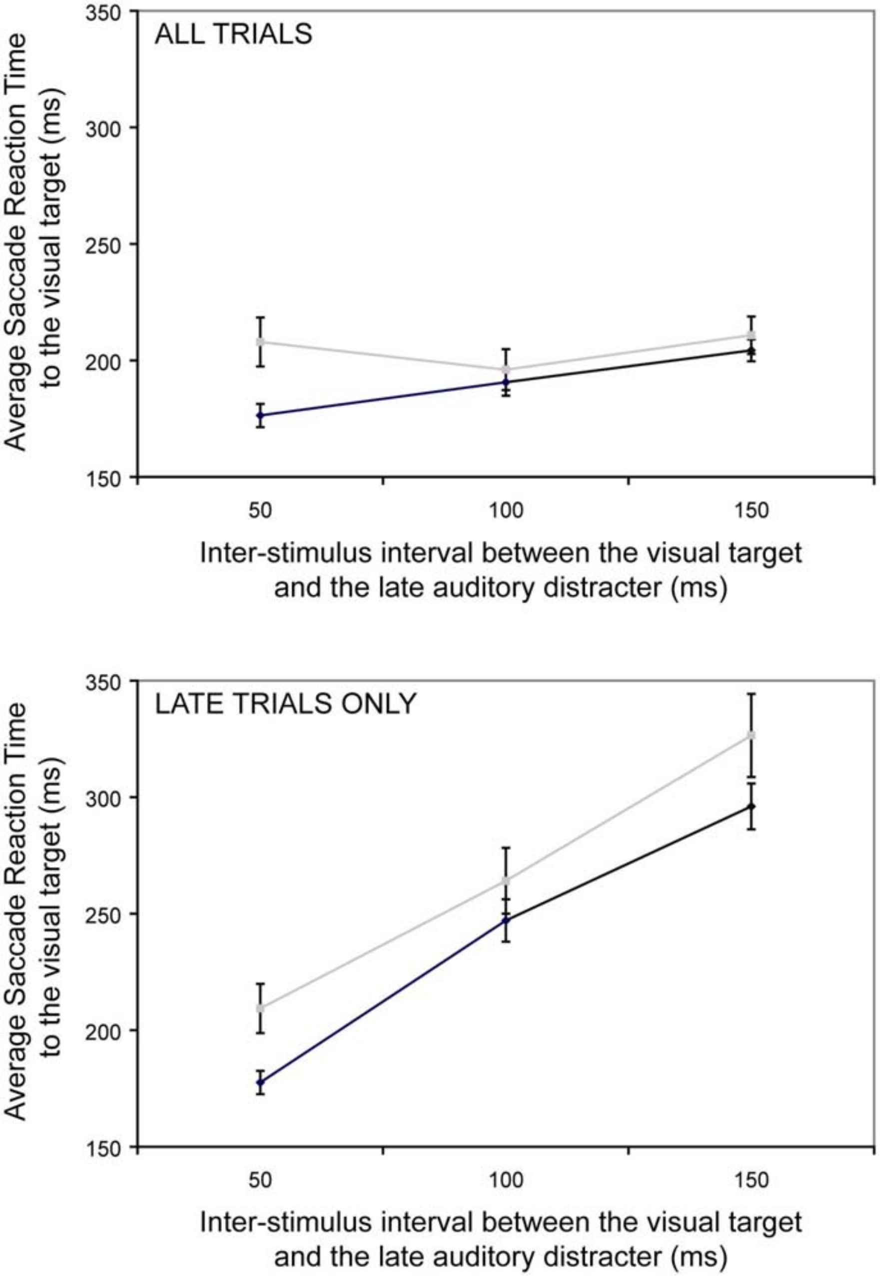
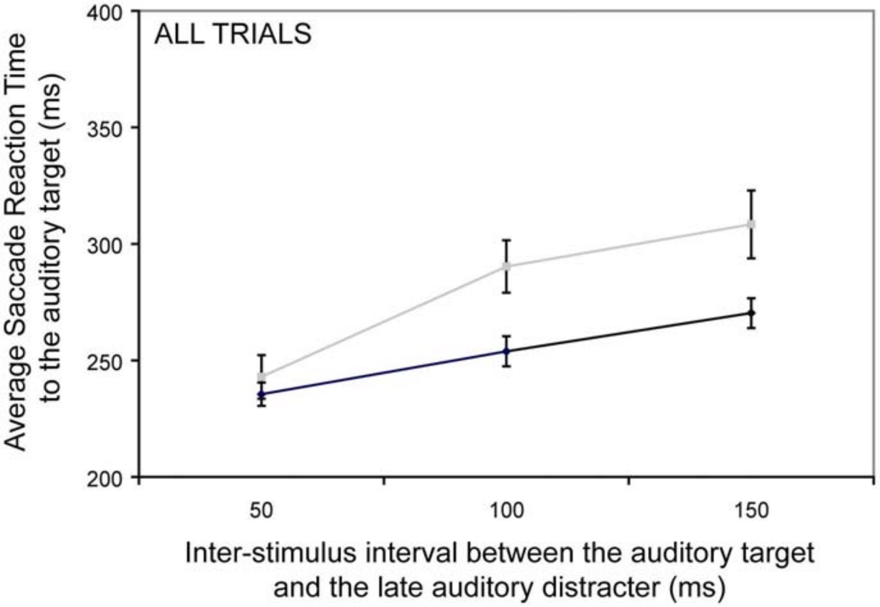


Figure 7.2: Average saccade reaction time (SRT) to a visual target presented alone (black) or with a subsequent, auditory distracter (gray) in experimental blocks including late visual and auditory distracters.



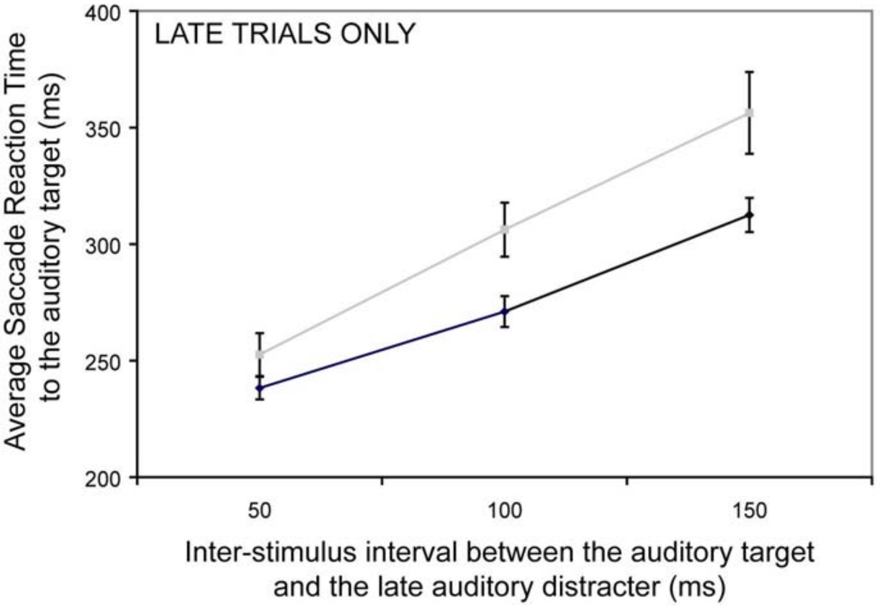
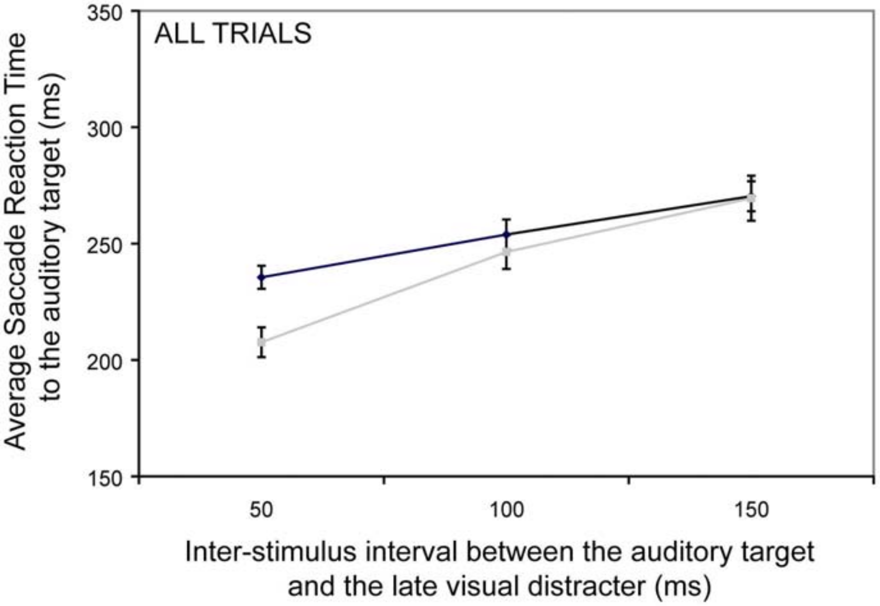


Figure 7.3: Average saccade reaction time (SRT) to an auditory target presented alone (black) or with a subsequent, identical auditory distracter (gray) in experimental blocks including late visual and auditory distracters.



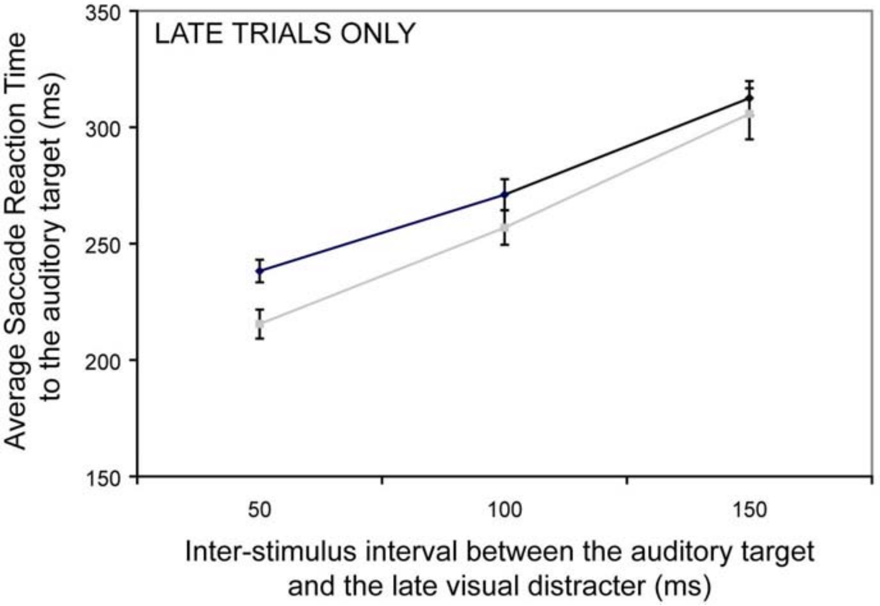


Figure 7.4: Average saccade reaction time (SRT) to an auditory target presented alone (black) or with a subsequent, visual distracter (gray) in experimental blocks including late visual and auditory distracters.

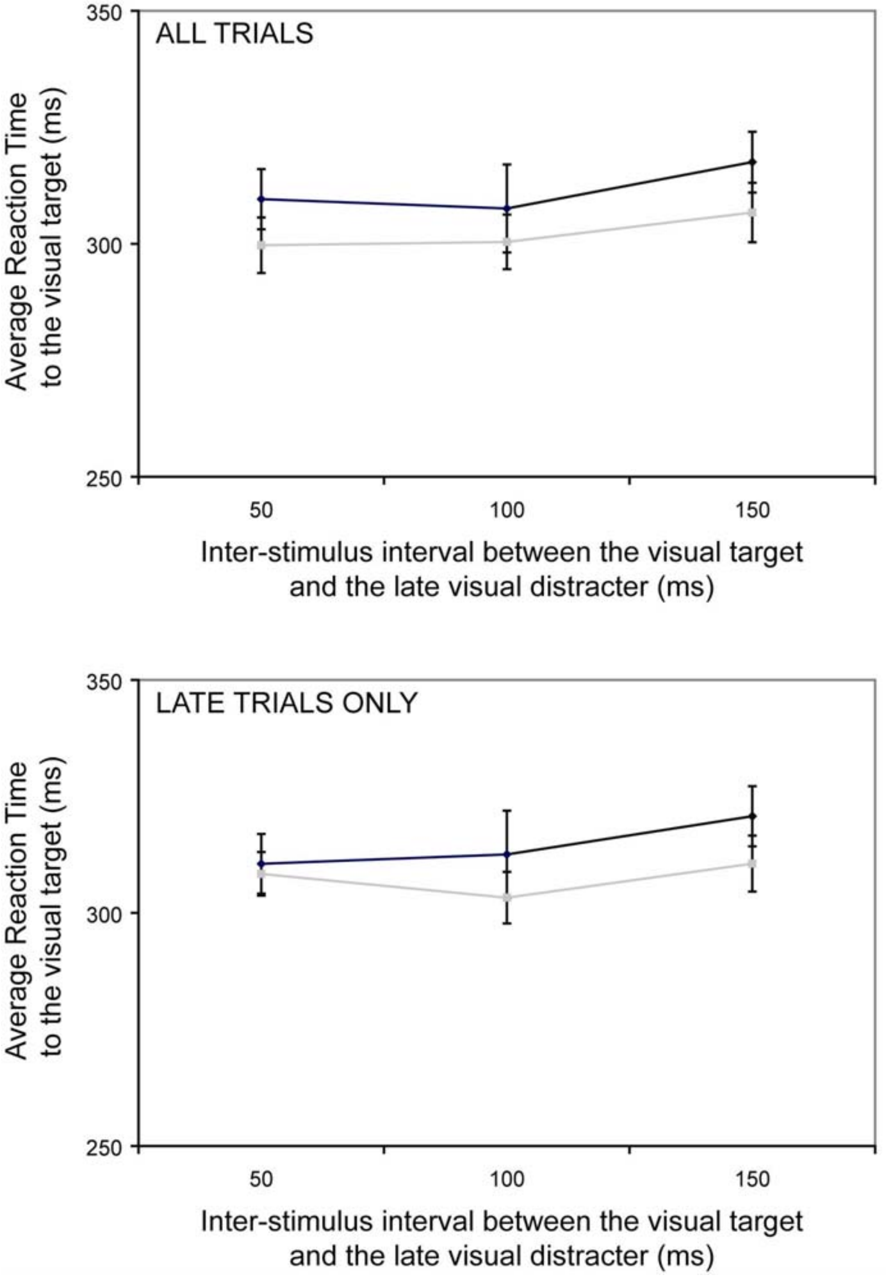


Figure 7.5: Average reaction time (measured via button press) to a visual target presented alone (black) or with a subsequent, identical visual distracter (gray) in experimental blocks including late visual and auditory distracters.

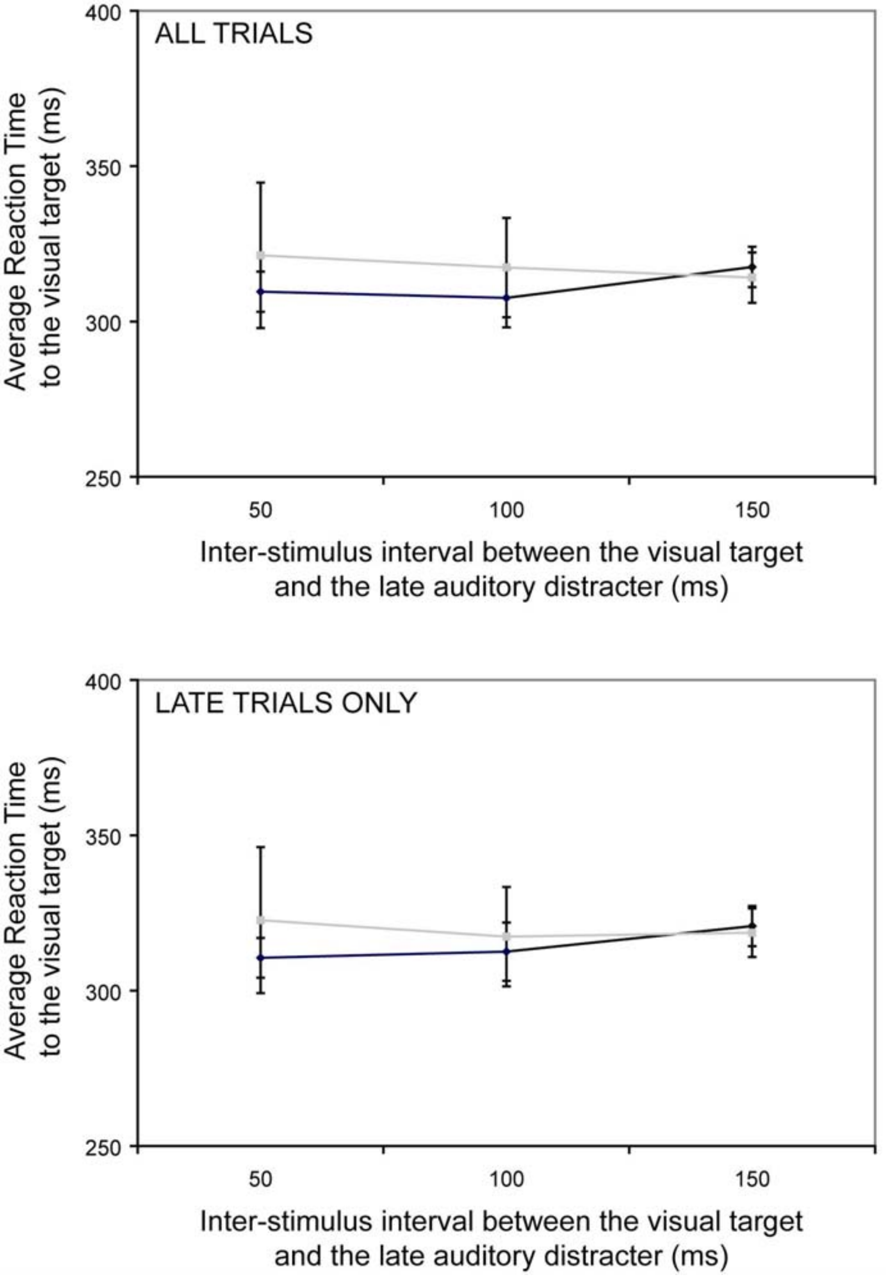
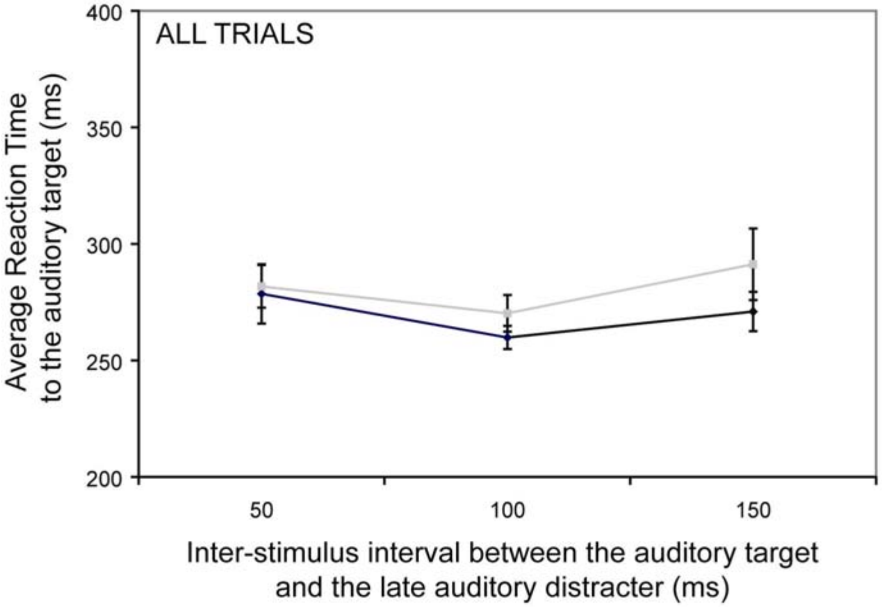


Figure 7.6: Average reaction time (measured via button press) to a visual target presented alone (black) or with a subsequent, auditory distracter (gray) in experimental blocks including late visual and auditory distracters.



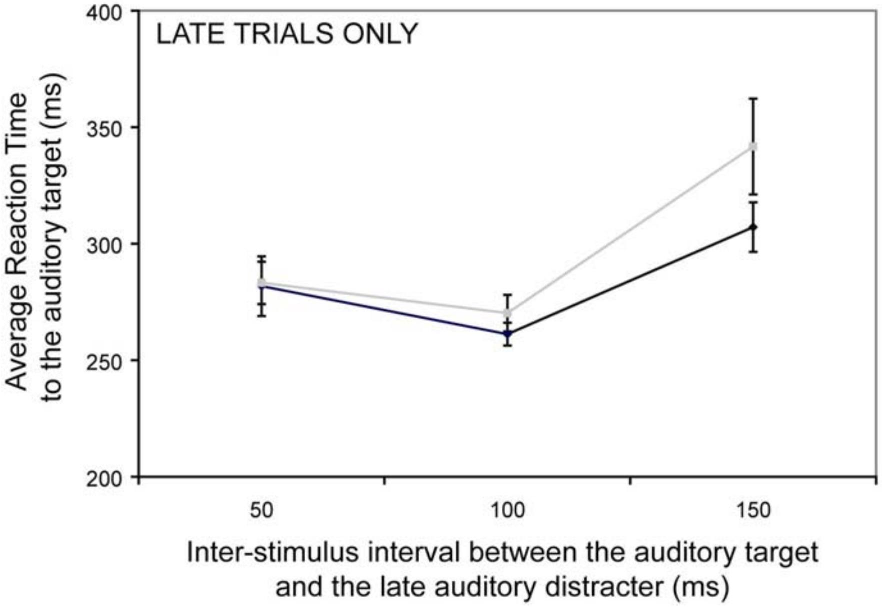


Figure 7.7: <u>Average reaction time (measured via button press) to an auditory target presented alone (black) or with a subsequent, identical auditory distracter (gray) in experimental blocks including late visual and auditory distracters</u>.

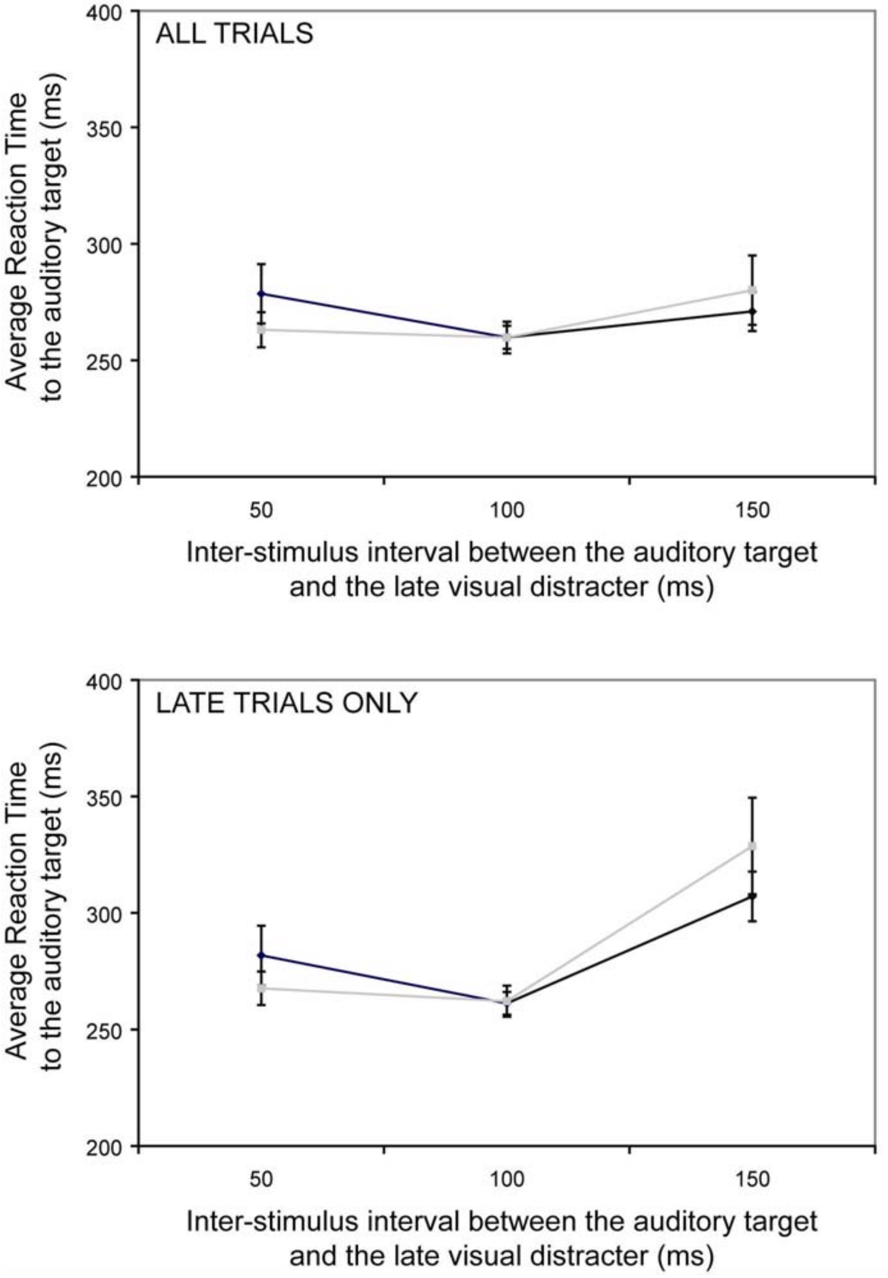


Figure 7.8: <u>Average reaction time (measured via button press) to an auditory target presented alone (black) or with a subsequent, visual distracter (gray) in experimental blocks including late visual and auditory distracters</u>.

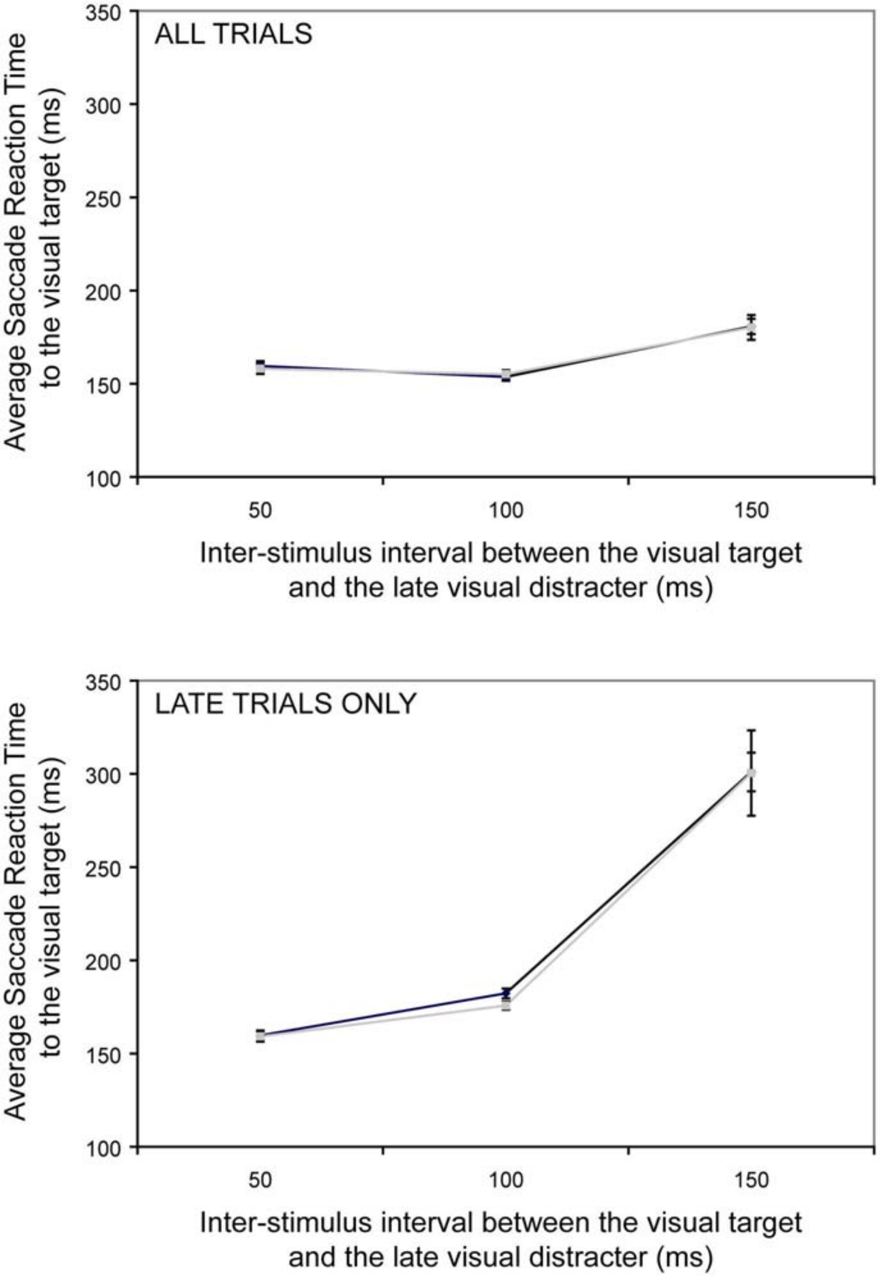
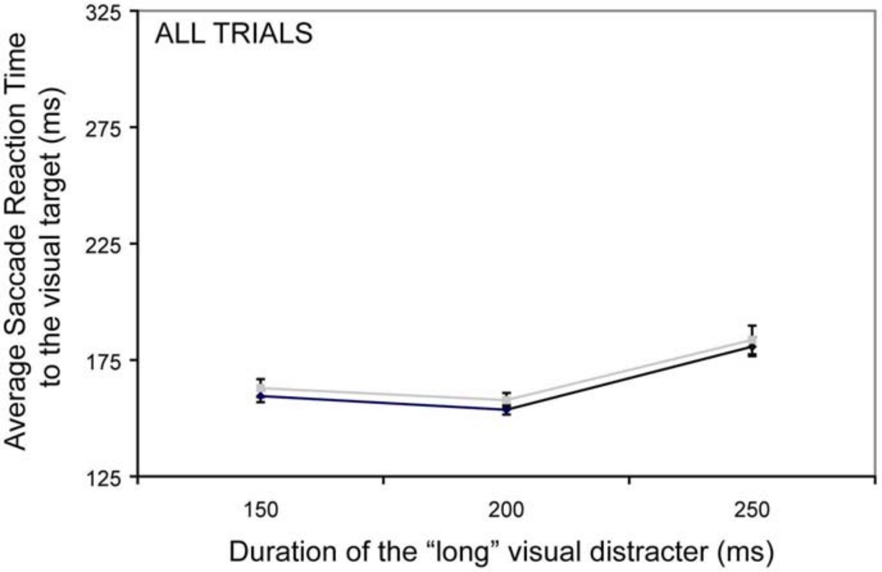


Figure 7.9: Average saccade reaction time (SRT) to a visual target presented alone (black) or with a subsequent, identical visual distracter (gray) in experimental blocks including late visual distracters and "long duration" visual targets.



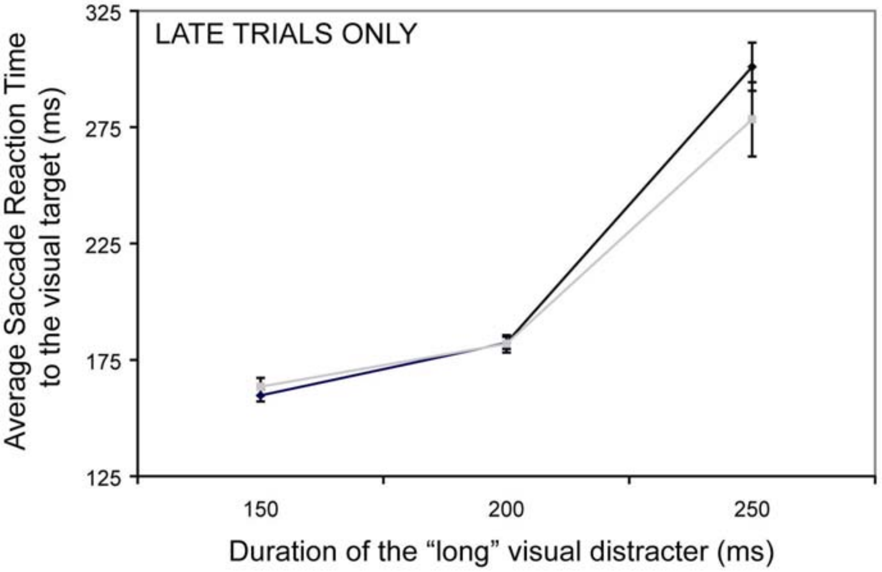
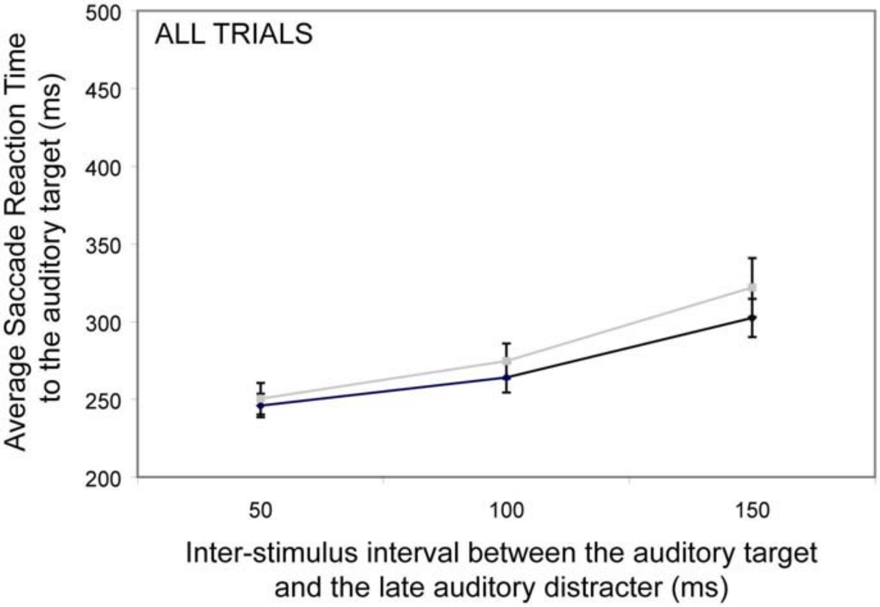
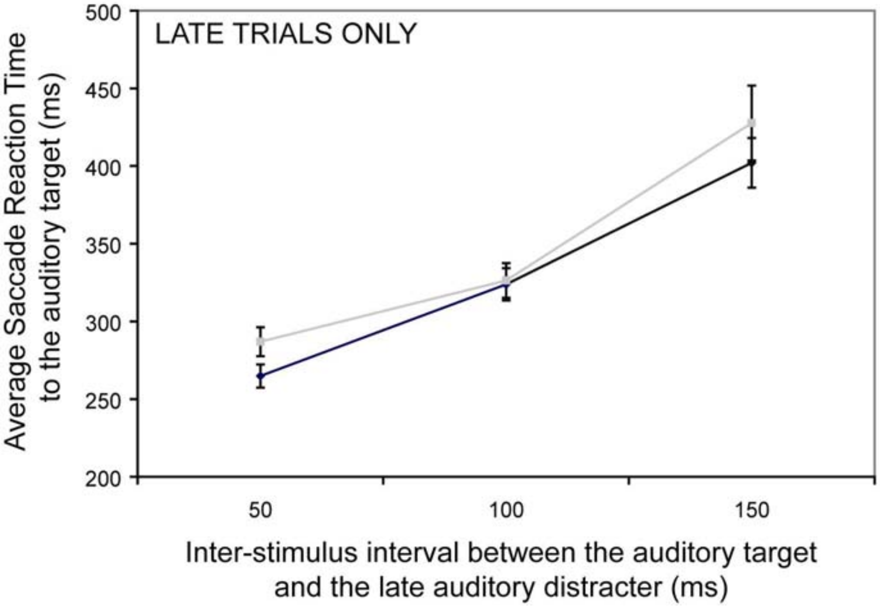
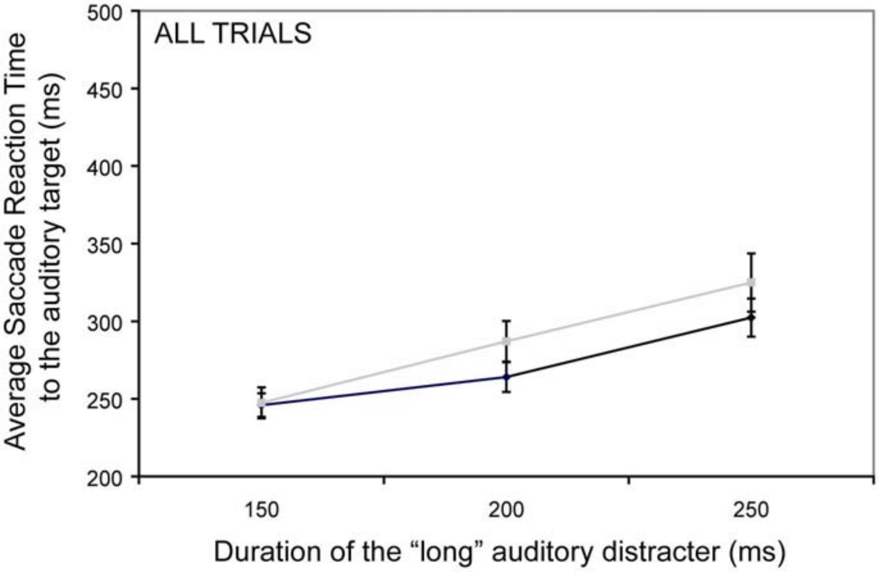


Figure 7.10: Average saccade reaction time (SRT) to a 50 ms visual target (black) or a "long duration" visual target (gray) in experimental blocks including late visual distracters and "long duration" visual targets.





Chapter 7, Figure 11: Average saccade reaction time (SRT) to a auditory target presented alone (black) or with a subsequent, identical auditory distracter (gray) in experimental blocks including late auditory distracters and “long duration” auditory targets.



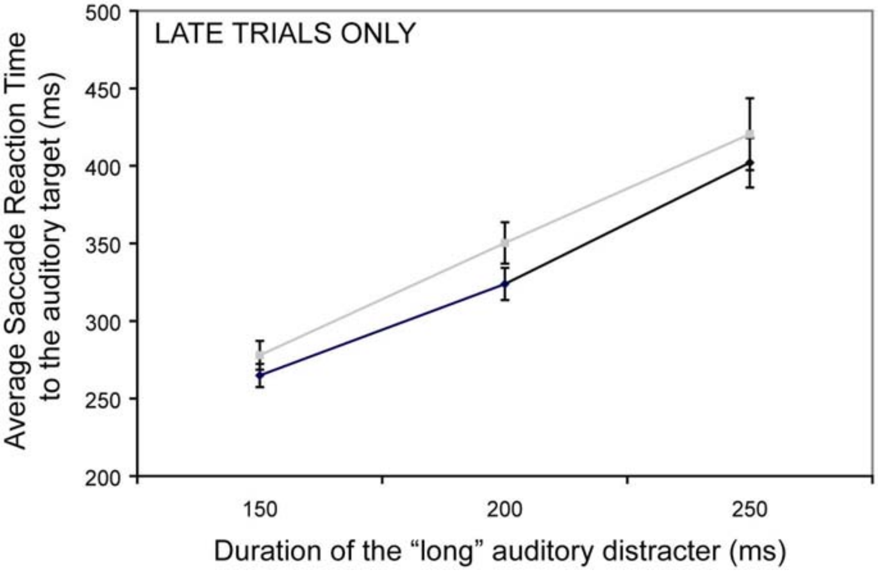


Figure 7.12: <u>Average saccade reaction time (SRT) to a 50 ms auditory target (black) or a “long duration” auditory target (gray) in experimental blocks including late auditory distracters and “long duration” auditory targets</u>.

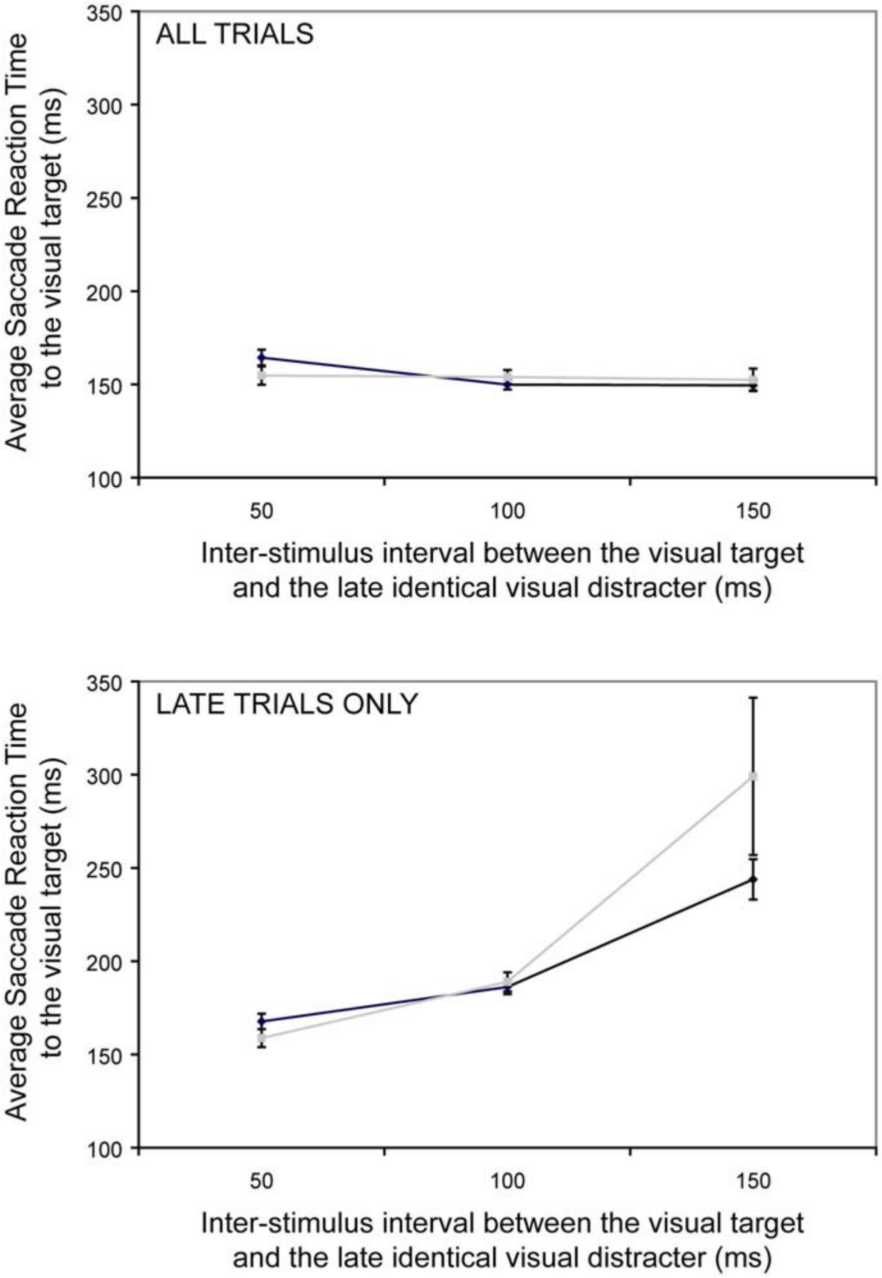


Figure 7.13: Average saccade reaction time (SRT) to a red visual target presented alone (black) or with a subsequent, identical red visual distracter (gray) in experimental blocks including late red visual distracters and late color-distinct green visual distracters.

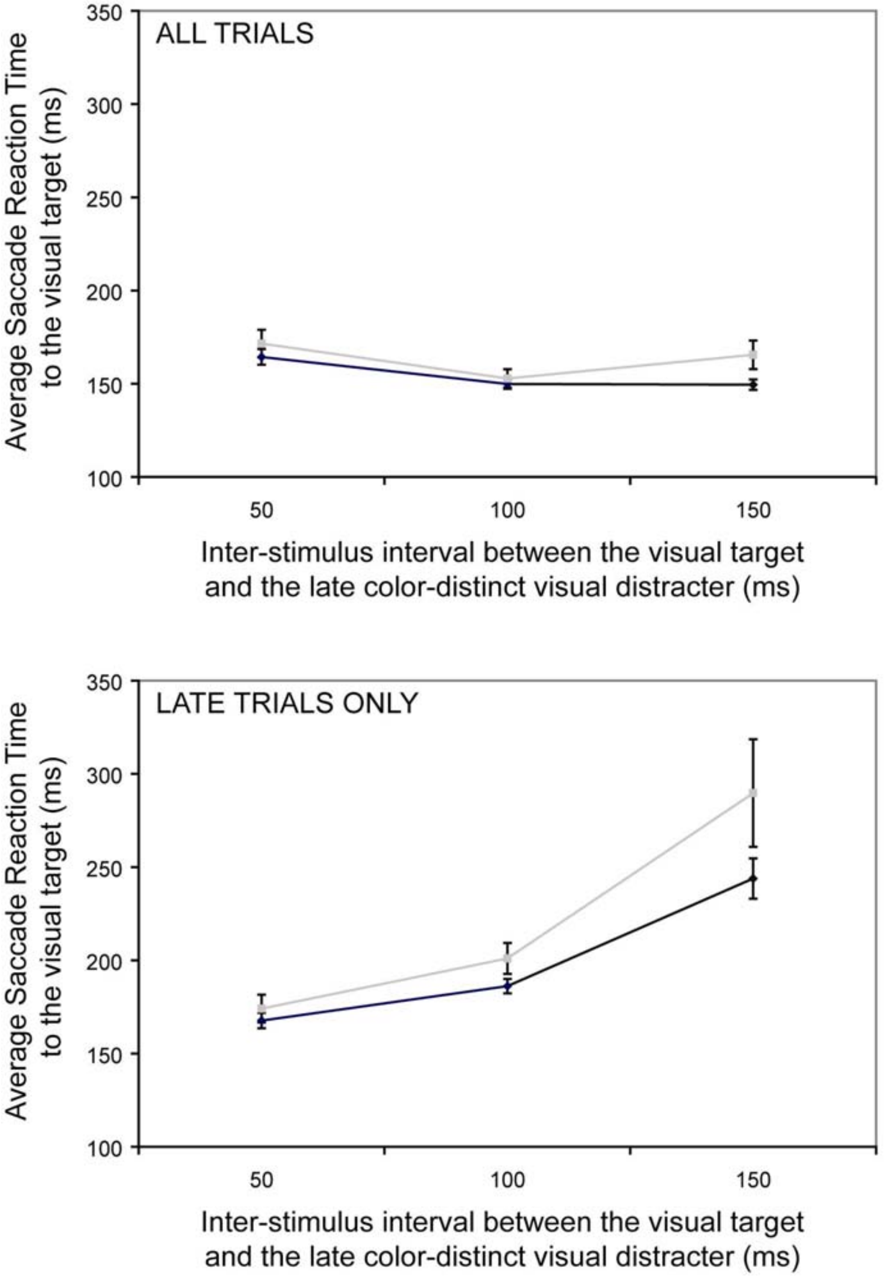


Figure 7.14: Average saccade reaction time (SRT) to a red visual target presented alone (black) or with a subsequent, color-distinct, green visual distracter (gray) in experimental blocks including late red visual distracters and late color-distinct green visual distracters.

Chapter 8

General Discussion

The first study presented here, on gaze-dependent aftereffects in color and depth, showed how powerfully extra-retinal information, like gaze direction, can modulate the visual experience. By simultaneously adapting opposed visual stimuli in the same retinal location under different gaze conditions, we demonstrated the robustness and power of coordinate transformation neural processes. The second study, on cyclopean flash-lag illusion, indicated how general and universal the mechanisms of stimulus integration are, finding flash-lag with stereoscopically defined, third-order visual motion stimuli. The third study, which introduced the turn-point phantom, a new visual motion illusion, suggested that integration between features of a target—its motion and the position of its turn-point—can evoke the same sorts of stimulus integration issues that distinct targets face. This suggested that integration is not necessarily associated with distinct stimuli so much as with distinct internal representations of sensory information. The fourth study, on vestibular induced visual and auditory target mislocalization, provided evidence for broad windows of integration (upwards of a second) and another example of postdictive awareness in multisensory perception including the vestibular sense. The fifth study, on temporal factors related to visual-motor ventriloquism, again demonstrated time windows of integration that spread beyond the presentation of relevant stimuli to task-irrelevant distracters. And the sixth study, on the saccade suppression by repetition in a multisensory context, introduces a special case of unimodal integration that is once again postdictive, and strongly argues for context-dependent scaling of integration windows. Taken together, these studies offer clues to a generalized model of stimulus integration that operates both within and across sensory modalities.

Towards a Generalized Model of Stimulus Integration

The sixth study discussed a model of stimulus integration that was adapted from models of multisensory integration (Colonius and Arndt, 2001; Colonius and Diederich, 2004). Here we discuss some of the details and ramifications of that model.

Colonius and Diederich (2004) proposed a two-stage model of multisensory integration that includes an initial stage, during which sensory information from separate modalities is processed independently, and a second stage in which those independent percepts are integrated in the preparation of an ocular motor response. Integration only occurs if the independent processes of the first stage all terminate within a certain time interval. Furthermore, the dynamics of the integration are a function of the spatial and temporal proximity of the percepts, such that coincidence results in facilitation and disparity results in suppression. The model allows that the time window involved is variable, depending on the stimuli and context.

Given the results presented here for repetition suppression, in which stimuli within a single modality demonstrate integration timing effects when presented within a certain context, we argue that the neural machinery of integration is much more general. While it is convenient to presume an initial, unimodal stage in which within-modality stimuli are resolved and then fed into a second stage multisensory integration process, there is no necessity to do so. We suggest instead that all stimuli generate task-specific representations of perceptual space that are continually integrated over a time window that is context and stimulus dependent. On a task-specific basis, representations that occur within the time window exert influence on the final percept.

Specific results suggest numerous features of this model. Let us assume that the activity associated with a particular representation follows a Gaussian pattern. The fact that auditory stimuli are detected faster than visual stimuli, but that the saccadic reaction time is greater in auditory localization tasks, implies that the Gaussian for representations evoked by auditory stimuli is much broader in the time domain, and that the threshold for stimulus detection is lower than the threshold for localization (see figure 8.1). As a result, auditory stimuli reach detection thresholds faster than visual stimuli, but reach localization thresholds later. If the temporal distance between peak activity levels contributes to the perceptual compatibility of the stimuli (i.e., whether integration involves enhancement or depression), then this model predicts the suppression of visual targets followed by auditory distracters (figure 8.2), but the facilitation of auditory targets followed by visual distracters (figure 8.3). As well, it explains suppression as the result of integration across the representations of repeated stimuli (figures 8.2 and 8.3), provided both representations occur within the time window of integration.

Our results in other experimental contexts suggest additional features of both the time window, and the tolerance of perceptual compatibility. The fact that repetition suppression was not observed within modality for time-varied stimuli but was for feature-changed stimuli suggests that the time window of integration scales globally with stimulus variability across trials. If the time window is sufficiently small such that late distracter representations occur outside of it, then they should have no effect on target perception. These same results also suggest that the perceptual compatibility that determines enhancement or depression from integration is context sensitive. In experimental blocks that include auditory stimuli, temporal discrepancy seems to have a greater impact on stimulus compatibility, perhaps owing to the higher temporal resolution of the auditory sense (though it failed to reach significance, some degree of suppression with auditory repetition was observed). In experimental blocks that included only visual stimuli, temporal discrepancy had less of an impact, but color discrepancy was sufficient to induce suppression.

Future Directions

Repetition suppression, even within such a specific context, is new and very surprising. The specific limits of the contexts that result in sufficiently large integration windows are still not known. It is possible that multisensory stimulus presentation is not sufficient, but that the regular correlation in time or space of multimodal stimuli is required. In study 6, the location of the distracter was always the same as the location of the target. Some pilot studies have suggested that uncertainty with regard to the position of distracter provides sufficient context for within-modality repetition suppression. This needs to be confirmed. The results here also suggest that the degree of perceptual discrepancy between stimuli results in suppression while sufficient perceptual likeness can result in facilitation. If this is the case, then a color-distinct distracter presented in a multimodal context may result in even larger suppression through repetition than was observed here. Similarly, if separation in color-space can evoke suppression, then perhaps an auditory correlate can be demonstrated using pure tones of different frequencies in place of white-noise bursts.

In the case of ventriloquism, the paradigm provides relative information about both the time scale, and the resulting percept of integration. A surface examination of the distribution of saccade magnitudes suggested two different types of solution to the integration problem: 1) an averaging across stimuli, and 2) one stimuli or the other winning out. It appeared that which solution was employed depended largely on the perceptual compatibility of the stimuli. More rigorous examination of the parameters that lead to each strategy may yield insight as to the boundaries of perceptual compatibility, particularly if the determination of integration strategy correlates with saccadic response latency.

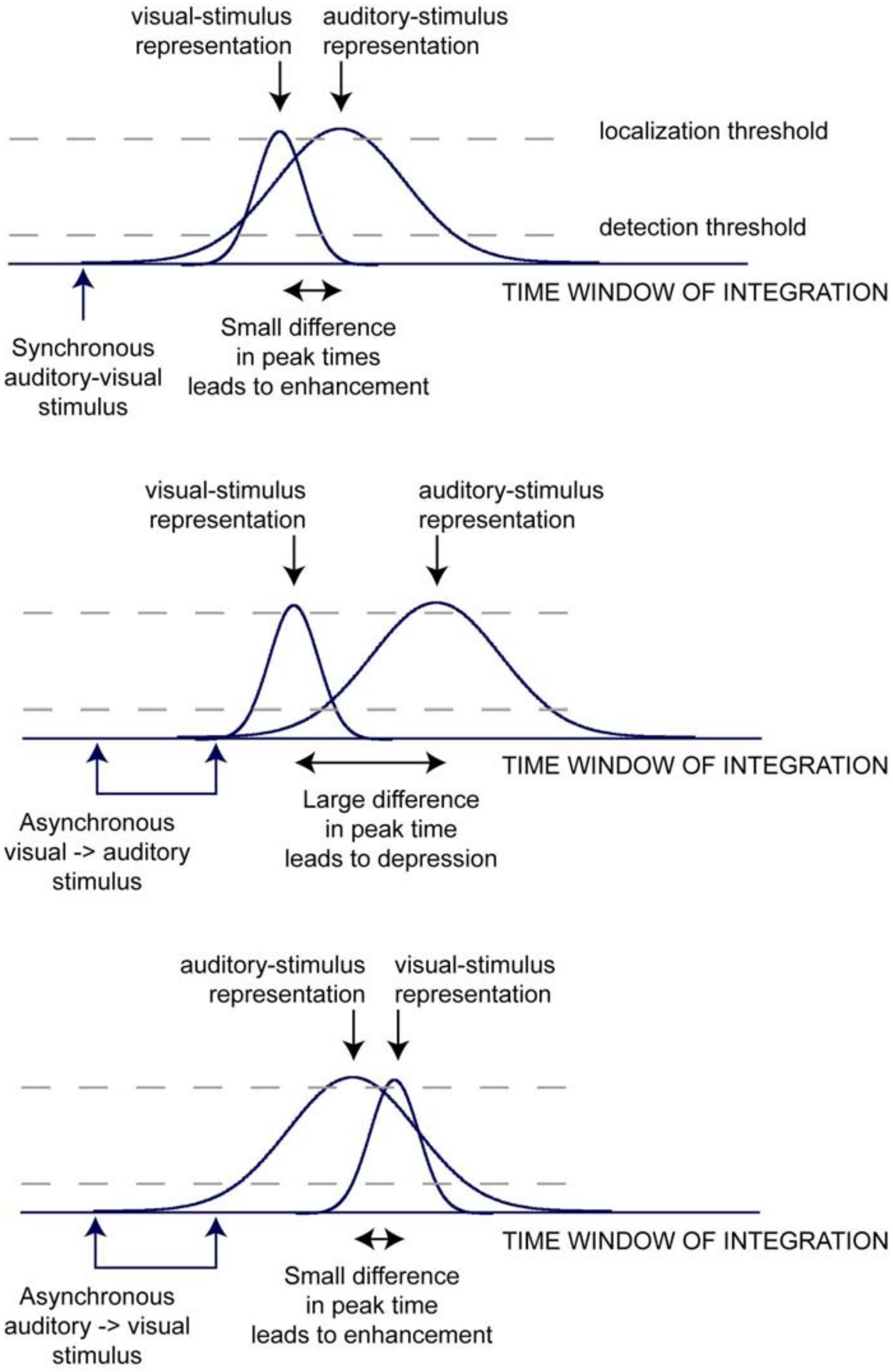


Figure 8.1: Schematic representation of a time window of integration model for synchronous and asynchronous visual and auditory stimuli.

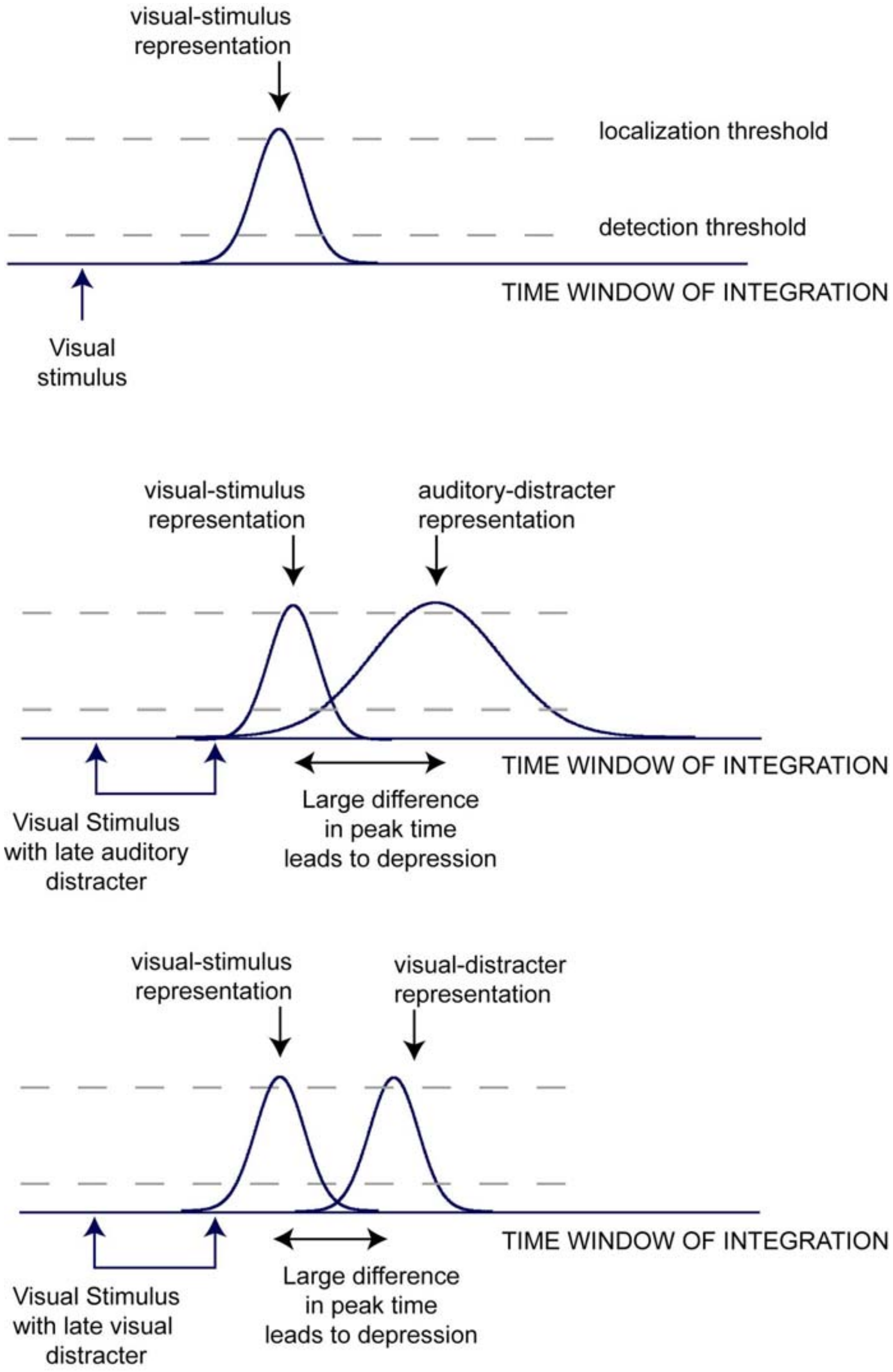


Figure 8.2: Schematic representation of a time-window-of-integration model for a visual target alone, with a late visual distracter, or a late auditory distracter.

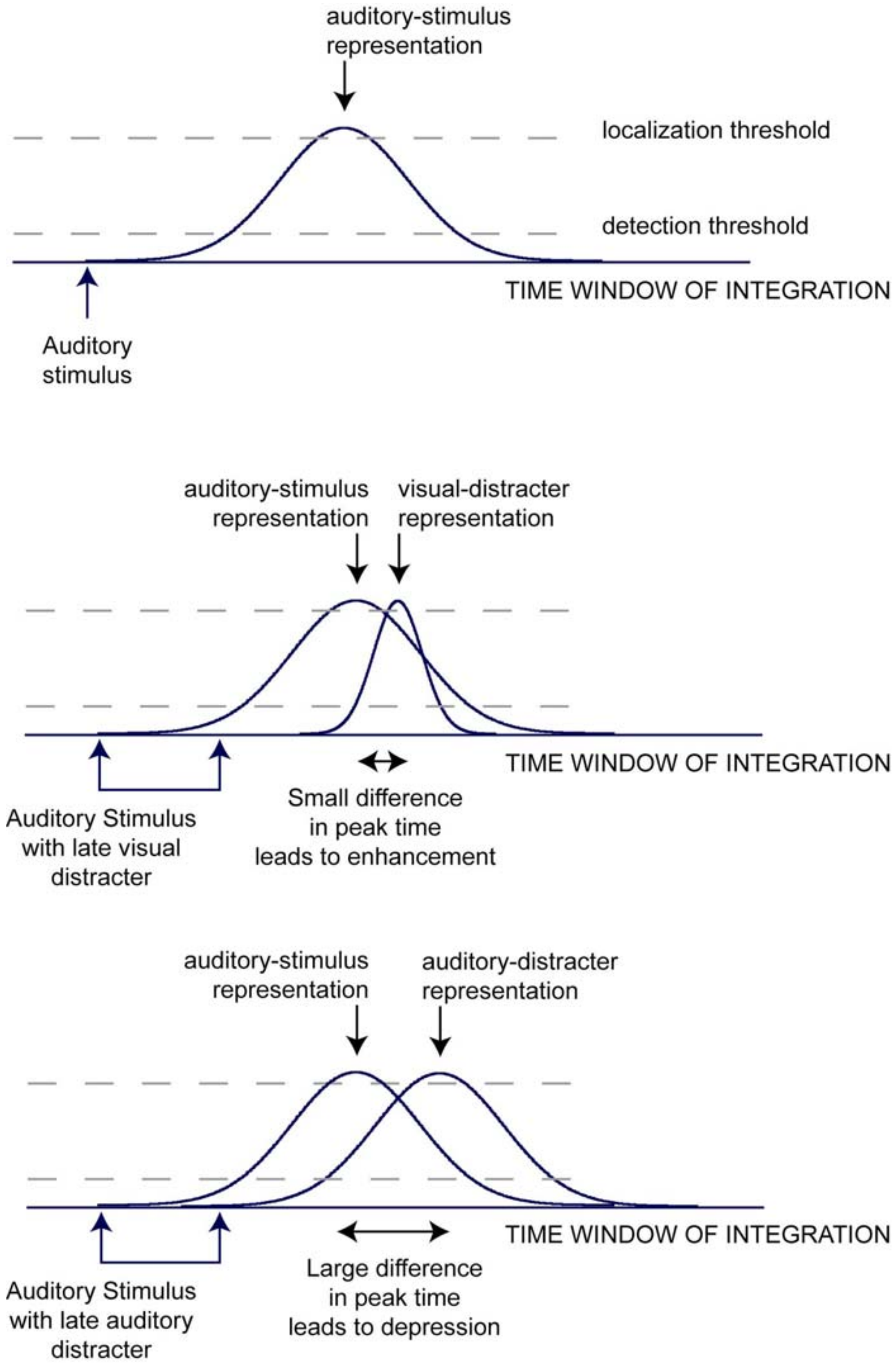


Figure 8.3: Schematic representation of a time-window-of-integration model for an auditory target alone, with a late visual distracter, or with a late auditory distracter.

References

Actis-Grosso R, Stucchi N (2003) Shifting the start: Backward mislocation of the initial position of a motion. *Journal of Experimental Psychology: Human Perception and Performance* 29:675–691.

Amlot R, Walker R, Driver J, Spence C (2003) Multimodal visual-somatosensory integration in saccade generation. *Neuropsychologia* 41:1–15.

Andersen RA, Essick GK, Siegel RM (1985) Encoding of spatial location by posterior parietal neurons. *Science* 230:456–468.

Andersen RA, Mountcastle VB (1983) The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience* 3:532–548.

Anstis SM, Harris JP (1974) Movement aftereffects contingent on binocular disparity. *Perception* 3:153–168.

Bachman T (1994) *Psychophysiology of Visual Masking*. Nova Science, Commack, NY.

Baldo MV, Klein SA (1995) Extrapolation or attention shift? *Nature* 378**:**565–566.

Barlow HB (1953) Summation and inhibition in the frog’s retina. *Journal of Physiology* 119:69–88.

Berry MJ, Brivanlou IH, Jordan TA, Meister M (1999) Anticipation of moving stimuli by the retina. *Nature* 398:334–338.

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

Blakemore C, Julesz B (1971) Stereoscopic depth aftereffect produced without monocular cues. *Science* 171:286–288.

Blouin J, Gauthier GM, van Donkelaar P, Vercher JL (1995) Encoding the position of a flashed visual target after passive body rotations. *Neuroreport* 6:1165–1168.

Blouin J, Gauthier GM, Vercher JL (1995) Failure to update the egocentric representation of the visual space through labyrinthine signal. *Brain and Cognition* 29:1–22.

Blouin J, Gauthier GM, Vercher JL (1997) Visual object localization through vestibular and neck inputs. 2: Updating off-mid-sagital-plane target positions. *Journal of Vestibular Research* 7:131–143.

Blouin J, Labrousse L, Simoneau M, Vercher JL, Gauthier GM (1998) Updating visual space during passive and voluntary head-in-space movements. *Experimental Brain Research* 122:93–100.

Braddick O (1980) Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society London* B 290:137–151.

Brainard DH (1997) The psychophysics toolbox. *Spatial Vision* 10:433–436.

Cavanagh P (1992) Attention-based motion perception. *Science* 257:1563–1565.

Cavanagh P, Mather G (1989) Motion: The long and short of it. *Spatial Vision* 4:103–129.

Colonius H, Arndt P (2001) A two-stage model for visual-auditory interaction in saccadic latencies. *Perception and Psychophysics* 63:126–174.

Colonius H, Diederich A (2004) Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience* 16:1000–1009.

Coltheart M (1973) Colour-specificity and monocularity in the visual cortex. *Vision Research* 13:2595–2598.

Coren S, Hoenig P (1972) Effect of non-target stimuli upon length of voluntary saccades. *Perceptual and Motor Skills* 34:499–508.

Doris MC, Munoz DP (1995) A neural correlate for the gap effect on saccadic reaction time in monkey. *Journal of Neurophysiology* 73:2558–2562.

Doris MC, Pare M, Munoz DP (1997) Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. *Journal of Neuroscience* 17:8566–8579.

Eagleman D, Sejnowski TJ (2000) Motion integration and postdiction in visual awareness. *Science* 287:2036–2038.

Favreau OE, Emerson VF, Corballis MC (1972) Motion perception: A color-contingent aftereffect. *Science* 176:78–79.

Findlay JM (1982) Global visual processing for saccadic eye movements. *Vision Research* 22:541–555.

Fischer B, Ramsperger E (1984) Human express saccades: Extreme short reaction times of goal directed eye movements. *Experimental Brain Research* 55:232–242.

Frens MA, van Opstal AJ, Van der Willigen RF (1995) Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception and Psychophysics* 57:802–816.

Freyd JJ, Finke RA (1984) Representational momentum. *Journal of Experimental Psychology—Learning Memory and Cognition* 10:126–132.

Fröhlich FW (1923) Über die Messung der Empfindungszeit (Measuring the time of sensation). *Zeitschrift für Sinnesphysiologie* 54:58–78.

Fu YX, Shen Y, Dan Y (2001) Motion-induced perceptual extrapolation of blurred visual targets. *Journal of Neuroscience* 21:RC172.

Galletti C, Battaglini PP (1989) Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *Journal of Neuroscience* 9:1112–1125.

Geldard FA, Sherrick CE (1972) The cutaneous “rabbit”: A perceptual illusion. *Science* 178:178–179.

Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA (2003) Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience* 15:20–29.

Harrington LK, Peck CK (1998) Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Experimental Brain Research* 122:247–252.

Heron J, Whitaker D, McGraw F (2004) Sensory uncertainty governs the extent of audio-visual interaction. *Vision Research* 44:2875–2884.

Howard IP (1982) *Human visual orientation*. Wiley, Toronto.

Hughes HC, Reuter-Lorenz PA, Nozawa G, Fendrich R (1994) Visual-auditory interactions in sensorimotor processing—saccades versus manual responses. *Journal of Experimental Psychology—Human Perception and Performance* 20:131–153.

Israel I, Fetter M, Koenig E (1993) Vestibular perception of passive whole-body rotation about horizontal and vertical axes in humans: goal-directed vestibule-ocular reflex and memory contingent saccades. *Experimental Brain Research* 96:335–346.

Jack CE, Thurlow WR (1973) Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Perceptual and Motor Skills* 37:967–979.

Jay MF, Sparks DL (1984) Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature* 309:345–347.

Jay MF, Sparks DL (1987a) Sensorimotor integration in the primate superior colliculus. I. Motor convergence. *Journal of Neurophysiology* 57:22–34.

Jay MF, Sparks DL (1987b) Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. *Journal of Neurophysiology* 57:35–55.

Jonides J (1981) Voluntary versus automatic control over the mind’s eye’s movement. In: Long J, Baddely A (eds.) *Attention and Performance IX*. Erlbaum, Hillsdale NJ, pp. 187–203.

Julesz B (1971) *Foundations of cyclopean perception.* Chicago: University of Chicago Press.

Kanai R, Sheth BR, Shimojo S (2004) Stopping the motion and sleuthing the flash-lag effect: spatial uncertainty is the key to perceptual mislocalization. *Vision Research* 44:2605–2619.

Kapadia MK, Westheimer G, Gilbert CD (1999) Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proceedings of the National Academy of Sciences USA* 96:12073–12078.

Kennet S, Taylor-Clarke M, Haggard P (2001) Noninformative vision improves the spatial resolution of touch in humans. *Current Biology* 11:1188–1191.

Kirchner H, Colonius H (2004) Predictiveness of a visual distractor modulates saccadic responses to auditory targets. *Experimental Brain Research* 155:257–260.

Kirchner H, Colonius H (2005) Interstimulus contingency facilitates saccadic responses in a bimodal go/no-go task. *Cognitive Brain Research* 25:261–272.

Kolers P, von Grunau M (1976) Shape and color in apparent motion. *Vision Research* 16:329–335.

Livingstone M, Hubel D (1984) Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience* 4:309–356.

Lu ZL, Sperling G (1995a) Attention-generated apparent motion. *Nature* 377:237–239.

Lu ZL, Sperling G (1995b) The functional architecture of human visual motion perception. *Vision Research* 35**:**2697–2722.

Lueck CJ, Crawford TJ, Savage CJ, Kennard C (1990) Auditory-visual interaction in the generation of saccades in man. *Experimental Brain Research* 82:149–157.

Mayhew JE (1973) After-effects of movement contingent on direction of gaze. *Vision Research* 13:877–80.

Marr D (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman and Company, New York.

McCullough C (1965) Colour adaptation of edge-detectors in the human visual system. *Science* 149:1115.

McDonald JJ, Teder-Salejarvi WA, Hillyard SA (2000) Involuntary orienting to sound improves visual perception. *Nature* 407:906–908.

McGurk H, Macdonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748.

Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors. *Journal of Neuroscience* 7:3215–3229.

Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221:389–391.

Meredith MA, Stein BE (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research* 365:350–354.

Mignard M, Malpeli JG (1991) Paths of information flow through visual cortex. *Science* 251:1249–1251.

Moore C, Enns J (2004) Object updating and the flash-lag effect. *Psychological Science* 15:866–871.

Munoz DP, Wurtz RH (1993a) Fixation cells in monkey superior colliculus. I. Characteristics of cell discharge. *Journal of Neurophysiology* 70:559–575.

Munoz DP, Wurtz RH (1993b) Fixation cells in monkey superior colliculus. II. Reversible activation and deactivation. *Journal of Neurophysiology* 70:576–589.

Müsseler J, Aschersleben G (1998) Localizing the first position of a moving stimulus: The Fröhlich effect and an attention-shifting explanation. *Perception and Psychophysics* 60: 683–695.

Nakayama K (1985) Biological image motion processing: A review. *Vision Research* 25:625–660.

Nawrot M, Blake R (1989) Neural integration of information specifying structure from stereopsis and motion. *Science* 244:716–718.

Nieman DR, Hayashi R, Andersen RA, Shimojo S (2005) Gaze direction modulates visual aftereffects in depth and color. *Vision Research* 45:2885–2894.

Nieman D, Nijhawan R, Khurana B, Shimojo S (2006) Cyclopean flash-lag illusion. *Vision Research* 46:3909–3914.

Nijhawan R (1994) Motion extrapolation in catching. *Nature* 370:256–257.

Nijhawan R, Kirschfeld K (2003) Analogous mechanisms compensate for neural delays in the sensory and the motor pathways: evidence from motor flash-lag. *Current Biology* 13:749–753.

Nishida S, Motoyoshi I, Andersen RA, Shimojo S (2003) Gaze modulation of visual aftereffects. *Vision Research* 43:639–649.

Ottes FP, van Gisbergen JAM, Eggermont JJ (1984) Metrics of saccade responses to visual double stimuli: two different modes. *Vision Research* 24:1169–1179.

Ottes FP, van Gisbergen JAM, Eggermont JJ (1985) Latency dependence of colour-based target vs. nontarget discrimination by the saccadic system. *Vision Research* 25:849–862.

Patterson R (1999) Stereoscopic (cyclopean) motion sensing. *Vision Research* 39:3329–3345.

Poggio G (1994) Stereoscopic processing in monkey visual cortex: a review. In: Papathamos T, Chubb C, Gorea A, Kowler E (eds.) *Early vision and beyond*. MIT Press, Cambridge, MA, pp. 43–53.

Posner MI, Cohen Y (1984) Components of visual orienting. In: Bouma H, Bouwhuis DG (eds.) *Control of language processes. (Attention and performance X).* Erlbaum, Hillsdale NJ, pp. 531–556.

Pouget A, Sejnowski TJ (1997) A new view of hemineglect based on the response properties of parietal neurones. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 352:1449–59.

Purushothaman G, Patel SS, Bedell HE, Ogmen H (1998) Moving ahead through differential visual latency. *Nature* 396:424.

Radeau M, Bertelson P (1977) Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception and Psychophysics* 22:137–146.

Regan D, Beverley KI (1972) Disparity detectors in human depth perception: Evidence for directional selectivity. *Science* 181:877–9.

Reichardt W (1957) Autokorrelations-auswertung als funktionsprinzip des zentralnewvensystems (Auto-correlations as a functional principle of the central nervous system). *Z. Naturf.* 12b:447–457.

Roll R, Velay J-L, Roll JP (1991) Eye and neck proprioceptive messages contribute to the spatial coding of retinal input in visually oriented activities. *Experimental Brain Research* 85:423–431.

Saslow MG (1967) Effects of components of displacement step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America* 57:1024–1029.

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. *Nature* 385:308.

Shams L, Kamitani Y, Shimojo S (2000) Illusions: What you see is what you hear. *Nature* 408:788.

Shenoy KV, Bradley DC, Andersen RA (1999) Influence of gaze rotation on the visual response of primate MSTd neurons. *Journal of Neurophysiology* 81:2764–2786.

Sheth B, Nijhawan R, Shimojo S (2000) Changing objects lead briefly flashed ones. *Nature Neuroscience* 3**:**489–495.

Slutsky DA, Recanzone GH (2001) Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12:7–10.

Smith AT, Scott-Samuel NE (1998) Stereoscopic and contrast-defined motion in human vision. *Proceedings of the Royal Society London* B 265**:**1573–1581.

Sparks DL, Mays LE (1980) Movement fields of saccade-related burst neurons in the monkey superior colliculus. *Brain Research* 190:39–50.

Sparks DL, Nelson JS (1987) Sensory and motor maps in the mammalian superior colliculus. *Trends in Neuroscience* 10:312–317.

Sparks DL (1988) Neural cartography: sensory and motor maps in the superior colliculus. *Brain, Behavior and Evolution* 31:49–56.

Squatrito S, Maioli MG (1997) Encoding of smooth pursuit direction and eye position by neurons of area MSTd of macaque monkey. *Journal of Neuroscience* 17:3847–60.

Stein BE, Wallace MT (1996) Comparisons of cross-modality integration in midbrain and cortex. *Progress in Brain Research* 112:289–299.

Tassinari G, Aglioti S, Chelazzi L, Peru A, Berlucchi G (1994) Do peripheral non-informative cues induce early facilitation of target detection? *Vision Research* 34:179–189.

Thornton IM (2002) The onset repulsion effect. *Spatial Vision* 15:219–243.

Trotter Y, Celebrini S (1999) Gaze direction controls response gain in primary visual-cortex neurons. *Nature* 398: 239–42.

Vroomen J, de Gelder B (2000) Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology – Human Perception and Performance* 26:1583–1590.

Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology* 76:1246–1266.

Whitney D, Cavanagh P (2000) Motion distorts visual space: Shifting the perceived position of remote stationary objects. *Nature Neuroscience* 3:954–959.

Whitney D, Murakami I (1998) Latency difference, not spatial extrapolation. *Nature Neuroscience* 1:656–657.

Wurtz RH, Albano JE (1980) Visual-motor function of primate superior colliculus. *Annual Review of Neuroscience* 3:189–226.

Yantis S, Jonides J (1984) Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance* 10:601–621.

Zahn JR, Abel LA, Dell'Osso LF (1978) The audio-ocular response characteristics. *Sensory Processes* 2:32–37.

Zahn JR, Abel LA, Dell'Osso LF, Daroff RB (1979) The audio-ocular response: Intersensory delay. *Sensory Processes* 3:60–65.

Zambarbieri D, Schmid R, Magenes G, Prablanc C (1982) Saccadic responses evoked by presentation of visual and auditory targets. *Experimental Brain Research* 47:417–427.

Zipser D, Andersen RA (1988) A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* 331:679–84.