2009

Visual Contributions to Spatial Perception during a Remote Navigation Task

Candace Lee Eshelman-Haynes
Wright State University

VISUAL CONTRIBUTIONS TO SPATIAL PERCEPTION DURING A REMOTE NAVIGATION TASK

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

By

CANDACE LEE ESHELMAN-HAYNES
M.S., Wright State University 2004
B.B.S. Hardin-Simmons University 1990

2009
Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

1 MAY 2009

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Candace Eshelman-Haynes ENTITLED Visual contributions to spatial perception during a remote controlled task BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Scott Watamaniuk, Ph.D.
Dissertation Director

Daniel Weber, Ph.D.
Graduate Program Director

John Flach, Ph.D.
Chair, Department of Psychology

Committee on
Final Examination

Valerie Shalin, Ph.D.

Brian Kruger, Ph.D.

Richard Warren, Ph.D.

Joseph F. Thomas, Jr., Ph.D.
Dean, School of Graduate Studies

Abstract

Eshelman-Haynes, Candace L. Ph.D., Human Factors and Industrial/Organizational Psychology Ph.D. Program, Department of Psychology, Wright State University, 2009. Visual Contributions to Spatial Perception During a Remote Navigation Task.

The purpose of this study was to explore the implications of perception and action coupling for the design of control and display interfaces in remotely piloted vehicles. Three experiments were conducted: spatial arrangement, path perception, and remote navigation. The results showed that panning independent of forward motion gives observers a greater sense of depth in a scene and aides in efficient navigation while rotation during forward motion results in ambiguities during passive observation. This research has implications for the design of control and visualization interfaces for remote navigation.

.

Table of Contents

List of Figures

INTRODUCTION

In 1997, a highly skilled and trained Russian cosmonaut was practicing a video based remote manual docking procedure aboard the MIR space station when he very carefully maneuvered his supply ship into the science module of the MIR. The supply ship then bounced off the science module and smashed a solar panel. The incident endangered the safety of the crew as well as the future of international cooperation in space. The Russian response to the incident was to declare it to be the result of operator error and levy fines on the two cosmonauts. If there was more in depth analysis on the part of the Russians, it was not made public. Questions about how the incident occurred seem to have remained unanswered, yet this high profile event highlights some important questions about how human operators work with unmanned vehicles in remote environments.

The use of remotely piloted vehicles (RPVs) on the ground, in the air and under the sea, is growing rapidly. Unmanned Aerial Vehicles (UAVs) logged more than 140,000 combat flight hours in 2007 (Osborn, 2007). These vehicles allow people to extend their presence into environments that are impractical or inhospitable for human operators. The term "unmanned" is misleading when used in the context of UAVs as well as other remotely piloted vehicles. UAV operations involve many people ranging from those in the UAV ground control station and Air Operations Center to those in nearby manned aircraft. Because of the growing use of unmanned vehicles and the continued reliance on human operators as part of the human robot team it is important to understand the interface requirements for the human operator.

Human Interaction with RPV's

As we move through the environment we experience an on-going stream of information from all of our senses. In a remote navigation task all of these senses are limited by the control and display mechanisms provided to the navigator. It may be tempting to assume that the high degree of automation and technological innovation in these UAVs would decrease the risk of accidents, but the numbers suggest otherwise. Unmanned aircraft have a greater number of accidents/incidents than manned aircraft (Williams, 2004), and more than 60% of UAV incidents involve human factors issues (Tvaryannas, Thompson, & Constable, 2005).

In systems that require external pilots, control issues during landing and takeoff have been identified as a major factor in accidents (Williams, 2006). For instance, 67% of human factors related accidents in the Hunter system used by the U.S. Army were attributed to problems with controlling the craft during takeoff and landing (Manning, Rash, Leduc, Noback & McKeon, 2004).

Generally speaking, the field of human factors maintains that automated systems should be designed to fit the limitations as well as the strengths of the human. In other words, the displays and controls of dynamic systems should be designed to keep the operator involved, informed and active. Many of the solutions developed in UAV control have involved increasing automation. Regarding flight-related tasks, Wiener (1988) pointed out the potential for automation to change fundamentally the nature of pilot performance in terms of task demands, ability to oversee and monitor system performance, and crew coordination. Even in highly automated systems, some tasks cannot be automated, thus, requiring the human operator to transition from passive observer to active controller. For unmanned vehicles, specific problems occur in handover of control from automation to human, recovering from error, and mapping of controls.

The increasing use of UAV's and other remotely piloted vehicles (RPV's) brings to the forefront questions about the coupling of perception, action, and environment. RPV's are by definition systems that remove the actor/perceiver from the environment. This means that pilots must maintain an awareness of the vehicles position in space relative to other key elements such as the operator's position, obstacles, targets, and landmarks without the benefit of occupying the action space. The decoupling of perception and action in this way makes the design of control systems slightly more complex regarding the mapping of controls, layout of seating, and transfer rates for information (Wertheim, 1998; Williams, 2006; Muth, Walker, & Fiorello, 2006).

Robots are gaining acceptance as team members in civilian and military operations. Controlling the remote operation of unmanned vehicles is highly dependent upon a human operator's ability to develop situation awareness on the robot, the task being performed and the physical environment. They need to be aware of where the vehicle is, what the vehicle is doing, and how individual parts of a task relate to the end goals. Operators must also consider how the environment affects vehicle status and the ability to complete tasks.

When a remotely piloted vehicle is a stand in for a human observer, the natural dynamic relationship between properties of the scene being explored and the human perceptual system is broken. The decoupling undermines the remote observers’ perception of affordances in the scene (Gibson, 1979) which is illustrated by recent studies of human robot interaction where remote observers experience various difficulties in understanding the environment being traversed by a robotic system (Murphy, 2004). The key hole or soda straw effect in remote working situations refers to the narrow view afforded to operators in remote navigation tasks and results in gaps in mental models of the explored space (Casper & Murphy, 2003). The challenge for the RPV operator is not in point to point navigation but in overall situation awareness. Because the actor is separated from the environment, theoretical questions about perception, action, and environment coupling become more applied in nature and have implications for the design of control surfaces, communication systems, and training for teams responsible for the control of RPV’s (Tvaryannas et al. 2005).

Casper and Murphy (2003) studied the human and robot interactions following the events of 9/11 and relate many difficulties that hindered rescue workers understanding of the remote environment when exploring rubble piles and void spaces. The typical means of navigation are provided by cameras that are usually fix-mounted on the robot, are small in diameter, have small angular fields of view, operate with many frame dropouts, relay low resolution images, and have poor (if any) color rendering. The robot relays this impoverished video imagery to the handler's controller software. It is at this critical point where many perceptual ambiguities inherent in the remote setup can arise or even worsen from the crude integration of the direct video within the interface visualization.

Trying to understand the remote environment through this very literal soda straw undermines just what the human perceptual system excels at in the natural world. Seeing through a remote camera is not the same as having a human observer at a scene. The operator must create a visual understanding based on the constraints of the remote robot agent in the environment. In such cases, ambiguities involving object recognition, judgment of scale, and the absolute position of objects in the remote world are common. Lighting is usually uncontrolled and platforms equipped with their own illumination may alter the perception of color and texture.

As humans, we actively sample the world with effortless independent coordination between heading and gaze. However, with the fixed camera platforms used in robotic search and rescue, gaze and heading are neither independent nor controllable. When no frame of reference for body awareness is provided, there is profound misperception of depth, speed, and scale of obstacles and passages. Many of these ambiguities stem from the impoverished and conflicting cues affecting depth and size perception.

Human Vision and Motion Perception

Within cognitive science, the coupling of perception and action is a significant area of study. Debates within the study of perception and action center on the role of mental representations of space. A perceptual representation of physical space is referred to as visual space. People typically inhabit the visual space in which they are working and, in this context, they can easily track their changing position in the inhabited space. Visual contributions to spatial perception make up only a part of how we gain an understanding of the space that we occupy.

Efficient navigation is thought to require a good representation of body position/orientation in the environment and an accurate updating of this representation when the body-environment relationship changes. Such updating is based on the ability to estimate the speed and amplitude of body displacements. The means by which this is done in personal navigation such as walking or driving include either extracting heading from optic flow (Lappe, Bremmer, & van den Berg, 1999; Warren & Hannon, 1988; Warren, Kay, Zosh, Duchon, & Sahuc, 2001), integrating vestibular acceleration (Ivanenko, Grasso, Israël, & Berthoz, 1997) or using proprioceptive cues (Mittelstaedt, 1999). In a remote navigation task (as previously noted), the actor is often limited to visual information for navigation. This limitation suggests that interface designers should understand how visual information is used to support action.

Representations of visual space also need to take into account information about body movements that displace the retina. Path integration is the process of updating current position during navigation by monitoring internally generated self motion signals, such as vestibular information, efference copy, and proprioception. This process is interrupted during RPV navigation.

One possible consequence of the interruption of path integration in remote navigation tasks is the induced motion illusion. An induced motion illusion is said to occur when motion information is wrongly assigned to self or to objects against a background that is in motion. Such illusions regularly occur in natural settings. One example is interpreting your vehicle to be rolling backward while a larger vehicle next to you moves forward at a traffic light. Another example is watching water flow past as you stand at the edge of a pier or the bow of an anchored boat, and you lose track of what is actually moving (you or the water). These illusions may occur because of the organization of our visual system. As information is transmitted along the neural pathway, it is integrated from one stage of processing to the next. Some researchers believe that the interactions that occur during these integrations result in induced motion illusions (e.g., Hiris & Blake, 1996).

Visual information flows from the eyes to the primary visual cortex (V1) where motion detectors respond to information about speed, orientation, and direction. From V1 the motion information is integrated and passed on to the Middle Temporal area (MT) and Medial Superior Temporal area (MST). Generally, local motion information is integrated in area MT, which then passes that motion information to MST, where motion information is further integrated to reveal global properties of optic flow. All of these areas are highly interconnected with lateral, feed forward, and descending connections (e.g., Beardsley & Vaina, 2001).

Optic flow is the transformation of the optic array that is created as an observer translates through the world (Gibson, 1950). Recent research has shown that observers can recover their direction of self-motion from optic flow (Warren, W. & Hannon, 1988; Warren, Morris, & Kalish, 1988). Further, much recent research has established that certain areas of the visual system are sensitive to the properties of complex motion created by optic flow (Duffy & Wurtz, 1997a,b). Studies have shown that areas MT and MST in the Inferior Parietal Lobe (IPL) are sensitive to the complex properties of motion associated with optic flow such as expansion, contraction, and rotation (e.g., Cornilleau-Perez & Geilen, 1996).

Of course, the existence of induced motion illusions seems counter to the fact that people generally perform adequately when carrying out remote navigation tasks like the spaceship docking mentioned above. In fact, it is quite common for people to make errors in perceptual judgments about the spatial relationships among stimuli, but then interact with the same perceptually illusory stimuli in an accurate and precise manner. For example, Goodale (1998) found that observers would verbally respond that a horizontal bar was shorter than an intersecting vertical bar even though the bars were the same length. However, when asked to form their fingers to grasp the ends of the horizontal bar, observers were more accurate with regard to the size of the grip aperture. Therefore, it is reasonable to ask what sort of visual information emerges as a person interacts in a given environment that allows them to derive actionable understanding regarding the spatial layout of that environment.

There are four different reasons why the image of an external object might move on an animal's retina.

1. The object has moved
2. The animal has moved its location
3. The animal has moved its head relative to other parts of its body
4. The animal has moved its eyes relative to its head

Obviously, two or more of these could happen at the same time; and by coincidence or by design, they might even cancel each other's effects out so that no retinal movement results. It has been demonstrated that optic flow information is sufficient for a person to make judgments about relative self motion (heading), but not object motion (Banks, Ehrlich, Backus, & Crowell, 1996; Lappe, Bremmer, & Berg, 1999b; Royden & Hildreth, 1999).

Though optic flow appears to play an important role in path perception (Warren et al. 1988) and in postural control (Kelly, Loomis, & Bealle, 2005), some questions remain. One question centers around how the visual system treats extra-retinal signals. Extra-retinal signals are those signals created by eye and head rotations independent of body translation. One view is that the visual system recovers direction instantaneously from the retinal flow created by the combination of retinal signals (optic flow) and extra-retinal signals (van den Berg & Brenner, 1994; Warren & Hannon, 1988). The alternate view is that the visual system accounts for these extra-retinal signals through efference copy or feedback signals from the eye-movement system (Banks et al. 1996; Royden, Banks, and Crowell, 1992; Royden & Conti, 2003).

Because optic flow usually arises from the observer's movement through an environment of visually distinct points, contours and surfaces, optic flow inevitably entails retinal flow with motion energy (Adelson & Bergen, 1985). Optic flow is defined in terms of a head-centered reference frame and therefore, does not depend on eye rotations. In contrast, eye rotations cause retinal motions and, thus, alter the correspondence between the retinal flow and non-retinal signals specifying head and eye rotations. For this reason, the physiological processing of optic flow has been assumed to include some differentiation of retinal flow (Loomis, Beall, Macuga, Kelly, & Smith, 2006).

Motion can also be a cue for grouping together objects in the environment (called integration) and the motion of one object can have an effect on the way other objects are perceived to move (Mather, 1998). Things that move together are seen as belonging together and things that are near to objects in motion can be perceived to be in motion themselves (e.g., Ramachandran & Cavanagh, 1987). Motion can also provide information about the structure of objects in our environment. For instance, dots moving in various patterns can create the percept of a 3-dimensional object (e.g., Adelson & Movshon, 1982); dots moving in certain patterns can also create the percept of a human or animal in motion even without lines connecting the dots to create the form, referred to as biological motion (Johannson, 1970).

In order for the visual system to arrive at these percepts, it must parse which moving objects belong together. The process of decomposing a scene into moving parts, such as independently moving objects and self-motion, is referred to as segmentation. This process requires that motion information be assigned to appropriate sources. Duncker's rule (as cited in Wallach, 1982) states that, when both the object and the background are in motion, the motion is assigned to the object. Duncker also distinguishes between object relative change (configural change) and subject relative change. Object relative change refers to changes in the configuration of the visual scene (the positions of objects in the scene relative to each other). Subject relative change occurs as an observer moves through a scene and is roughly analogous to optic flow.

The role of extra-retinal signals in perception of visual motion is of interest to perception-action researchers because it has been hypothesized that these signals are integrated with optic flow information to support spatial updating. Extra-retinal signals refer to information about the position of the eyes obtained from non-retinal sources, including the oculo-motor command to displace the fovea towards a visual target (copy of motor efference) and proprioceptive cues. Extra retinal signals generate an efference copy that is thought to be used as a feed forward mechanism in perception. Generally, efference copy is assumed to modulate cell responses in MT and MST. Efference copy may play a role in moderating effects of induced motion illusions and the lack of efference copy in remote controlled activities may affect the ability to develop a good mental representation of the trajectory of movement in space (Bertin, Israel & Lappe, 2000).

One approach that perceptual researchers have taken to reveal the underlying mechanisms of the visual system has been to study visual illusions. The assumption is that understanding the mechanisms of the visual system will provide insight into the construction of mental representations. If the mental representation is the object being acted upon, then action responses should show evidence of perceptual illusions. There are several examples of visual illusions that manifest themselves in action. One such illusion is induced motion. The induced motion illusion occurs when the self or stationary objects are perceived to have moved when in reality the background moved (Duncker, 1929; Warren & Rushton, 2007). Such illusions regularly occur in natural settings. One example is interpreting your vehicle to be rolling backward while a larger vehicle next to you moves forward at a traffic light. Another example is watching waves flow past as you stand at the edge of a pier or the bow of an anchored boat, and you lose track of what is actually moving (you or the water). These illusions are hypothesized to occur because of the organization of the visual system.

Duffy and Wurtz (1993) suggested that an illusion they reported in which the focus of expansion in optic flow displays became displaced in the direction of planar motion may also arise because of interactions specifically in MT or MST. In their experiment, observers viewed random dot optic flow stimuli (rotation, expansion, and contraction) with an overlapping field of horizontal planar motion appearing in the fronto-parallel plane. Observers consistently misjudged the focus of expansion to be displaced in the direction of the planar motion. Pack and Mingolla (1998) further suggested that these types of illusions are the result of a two-stage neural mechanism. The local interactions between MT cells form the first stage of this mechanism. The second stage of the mechanism is a global subtraction occurring in MST and is thought to have developed to deal with image motion due to eye movements. Pack and Mingolla measured a shift in the focus of expansion in the presence overlapping planar motion and found that the global mechanism accounted for about 80% of the illusion with the remainder attributed to local interactions.

Vision for Perception and Vision for Action

The two systems model of perception and action is based on the organization of the visual system. According to this model, visual information is processed in two neural streams dedicated to different functions; one stream supports perception and one stream supports action (Milner & Goodale, 1995). Visual information arrives at the primary visual area, and then moves to two extra-striate regions: an occipito-temporal system (the ventral, or 'what' stream) concerned with object recognition and an occipito-parietal system (the dorsal, or 'where' stream) concerned with spatial characteristics (Schneider, 1969; see Figure 1). Proponents of this model claim that information in the dorsal stream is used to support action and the information in the ventral stream is used to support perception and object identification. The action system is not subject to illusions which are manifested in the dorsal (perceptual) system (Ungerleider & Mishkin, 1982; Milner & Goodale, 1995). Mental models are proposed to be part of the perceptual experience, but are not the object of action.

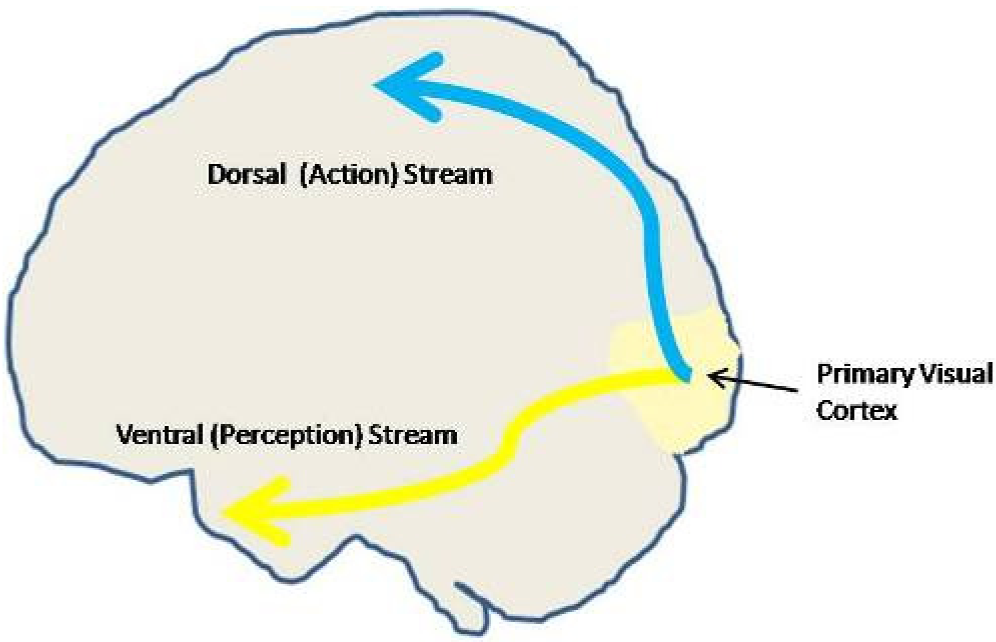

Figure 1: Two cortical visual systems - information flows from primary visual cortex to the dorsal and ventral streams

According to Milner and Goodale (1995), this arrangement allows spatially directed behavior to be rapid and efficient because it is implemented by a dedicated processor operating solely on the here-and-now goal of action. The perceptual system, in contrast, specializes in recognizing and remembering the identities of objects and patterns and their spatial interrelationships based on comparisons with prior knowledge. Since the sensorimotor and perceptual systems normally lead to motor actions and perceptual experiences that are consistent with each other, evidence for their dissociability is most likely to emerge when this congruence is disturbed, because of either experimental intervention in normal subjects or certain types of brain injury in clinical patients. Thus, research testing the two-system model has focused on trying to find differential performance between making perceptual judgments versus carrying out an action response.

One way to perturb perception in normal subjects is to expose them to illusions such as the Mueller-Lyer figure (see Figure 2). This illusion consists of two lines of equal length each with arrows on the end. One line has arrow heads pointing inward and the other has arrow heads pointing outward. When asked to judge the lengths of the two lines, viewers typically claim that the line with inward pointing arrows is longer. Observers who experience this illusion are nevertheless able to point (open-loop) accurately at its endpoints (Mack et al., 1985; Gillam & Chambers, 1985).

Evidence of a perception/action dissociation in healthy observers was reported by Aglioti, DeSouza, and Goodale(1995). They found that grasping was minimally affected, if at all, by visual illusions. Their study employed the Ebbinghaus Illusion (see Figure 3) in which two circles identical in size are surrounded by circles of different sizes. Observers typically judge the central circle surrounded by smaller circles to be larger than the central circle surrounded by larger circles. They found that perceptual judgments of size were dramatically influenced by the illusion, but the size estimates used in grasping the same objects were only marginally influenced. As mentioned earlier, Goodale and Milner (1998) found that observers would verbally respond that a horizontal bar was shorter than an intersecting, equal length vertical bar. However, observers would more accurately space their fingers to grasp the ends of the horizontal bar. This dissociation between perceiving the size of an object and grasping it was interpreted as strong evidence for the two-system model which posits that the ventral stream, but not the dorsal stream, is affected by visual illusions.

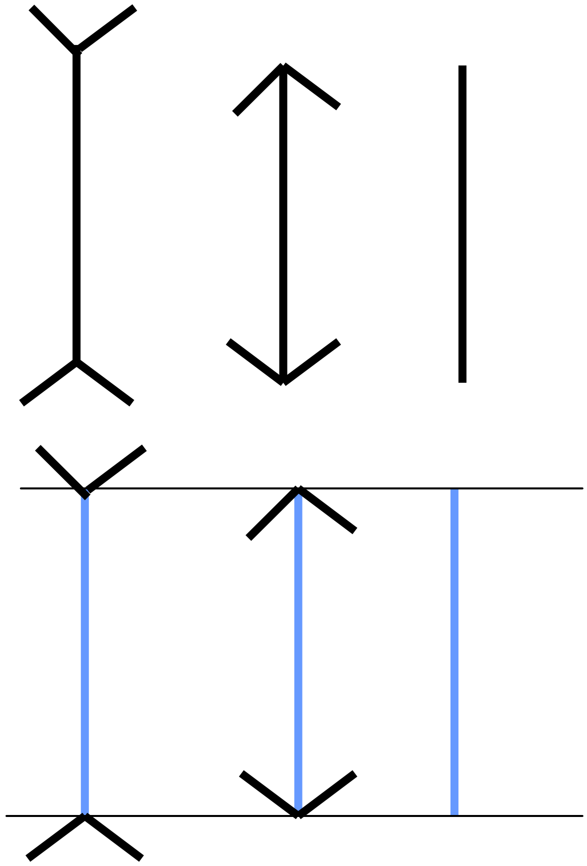

Figure 2: The Mueller-Lyer Illusion. The vertical lines as shown in the top panel appear to be different lengths based on the type of end-point. As illustrated in the bottom panel the vertical lines are all the same length.

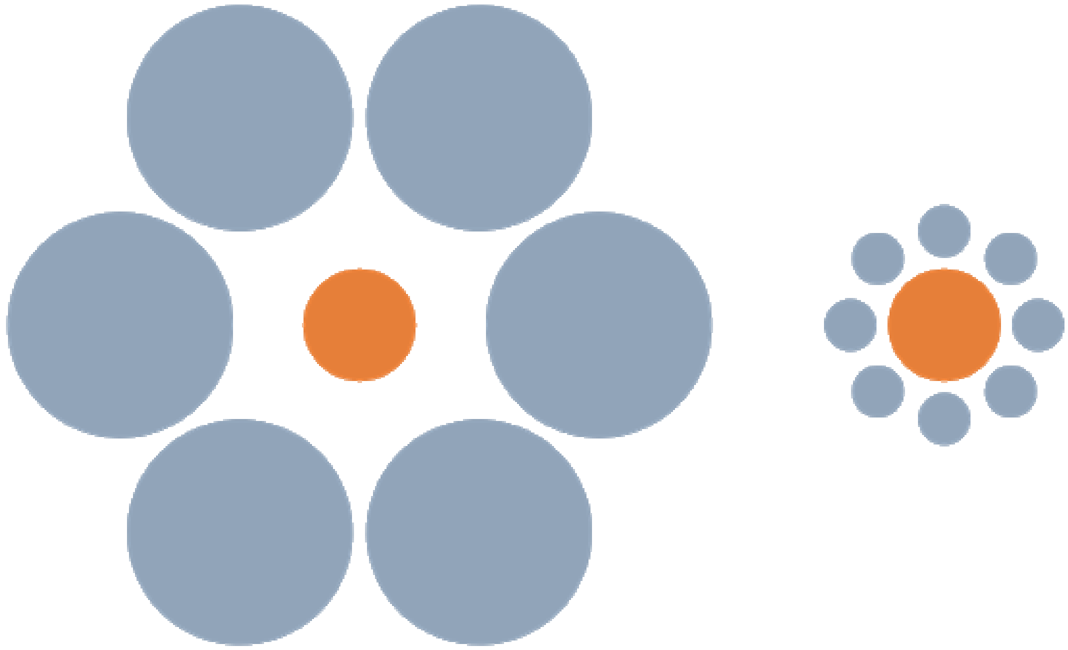

Figure 3. The Ebbinghuas Illusion. The central blue circle in the left and right figures are identical in size but the sizes of the surrounding circles causes a misperception such that the blue circle on the left seems larger than the one on the right.

As noted earlier, motion is processed in the dorsal stream. If the distinguishing factor for illusory responses is that the stimuli must be processed in the ventral stream, then illusions should disappear if the stimuli are altered so that the stimuli are processed in the dorsal (action) stream. In fact, Watamaniuk (2005) found that changing a stimulus from a static image to a moving one resulted in the negation of an illusory effect. He transformed the Poggendorf Illusion (see Figure 4) into a moving stimulus by replacing the oblique line with a dot that moved from left to right at an angle across the screen, disappearing behind an opaque occluder and then reappearing and continuing to move along its angled path on the other side. Observers were asked to judge if the path of the moving dot on the right side of the occluder was above or below alignment with the path of the moving dot on the left side of the occluder. For the static stimulus, a robust perceived misalignment between the oblique line segments was obtained, but when the oblique line was replaced with a dot moving in an oblique direction observers' judgments of alignment were close to veridical.

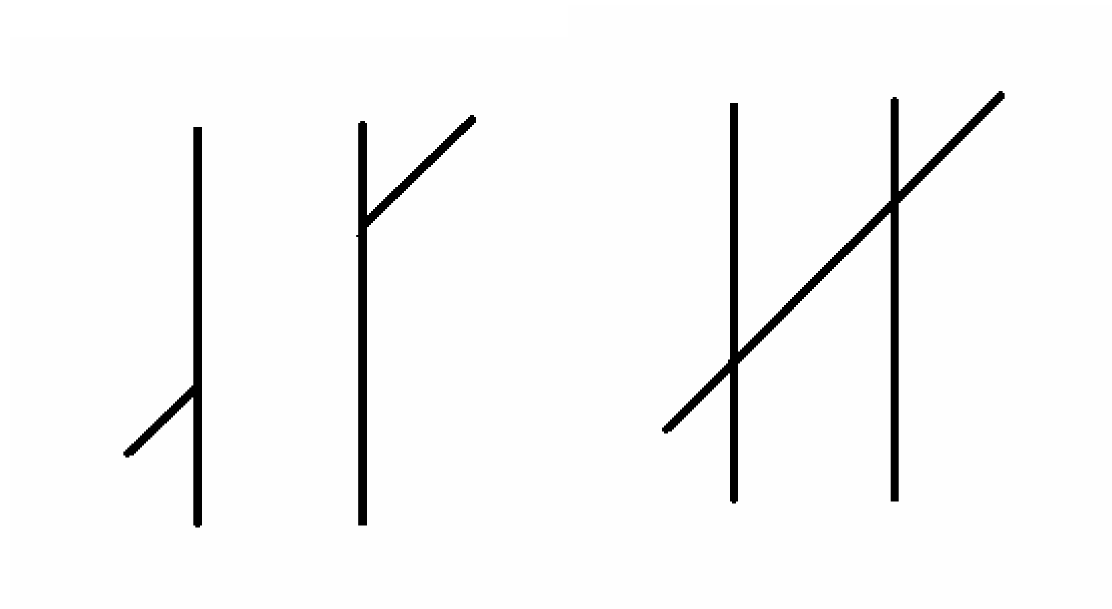

Figure 4: The Poggendorf Illusion. On the left, the two ends of a straight line segment passing behind an obscuring rectangle appear offset when, in fact, they are aligned. On the right, the middle segment is drawn in illustrating that the segments are aligned.

EXPERIMENTATION

The present set of experiments is intended to examine the relationship between perceptual judgments and performance in a remote navigation task. Specifically, these experiments were meant to begin exploring how perception of spatial layout of the environment changes as information about the environment emerges from initial static images to an actively explored environment with only visual information being made available. The work will add to previous research in two important ways. First, the images were not the idealized computer generated images of previous work, but were unprocessed images of physical scenes taken with a camera. This will make extrapolation to the real world context of remote navigation more direct because the images will contain the level of detail and "noise" that is common in real world applications. Second, the final experiment employing the navigation task will allow us to isolate the available strategies and test which strategy is more comfortable for participants to use in a remote navigation task.

The three experiments described in the following chapters explore mental representations of space and motion and their possible effects on navigation in a remote environment. The first two experiments explore mental representations in a traditional experimental paradigm showing a discrete stimulus event and requiring a perceptual judgment from observers. Specifically, the first experiment examines allocentric representation of space and the second focuses on egocentric perception of heading from optic flow. The third experiment utilizes an action-based paradigm to discern the ability to act based on different control mechanisms for obtaining visual information about the environment.

Apparatus

A Logitech™ camera designed for LEGO™ used in conjunction with the LEGO™ Mindstorms RCX (the RCX is the "brain" or CPU of the Mindstorms robotic system) and ROBOLAB software was used to create all of the visual stimuli for the experiments. A robotic vehicle built from the LEGO system that could be maneuvered by remote control was also employed. The Logitech camera had a viewing angle of 45.24° X 31.65° and a resolution of 360 x 240 pixels. All experiments and generated scenes occurred inside an arena 84 inches long by 70 inches wide (see Figure). The walls of the arena were painted flat black and the floor of the arena was covered with a flat black cotton sheet, which was sprinkled with self-adhesive ¾" diameter white dots in order to provide visual texture. Wooden dowels (6 x 1 inch) painted either red, yellow, or light blue were used as targets within the arena. The reported speed of forward motion of the remote vehicle is an approximated average because the servo motors on the LEGO™ vehicle did not run at a consistent rate due to variations in the flooring and friction caused by the configuration and gearing. While the control mechanisms on the LEGO™ system are not as precise as some other remote systems, this system was chosen because it offered the best flexibility for the different configurations needed for the experiments.

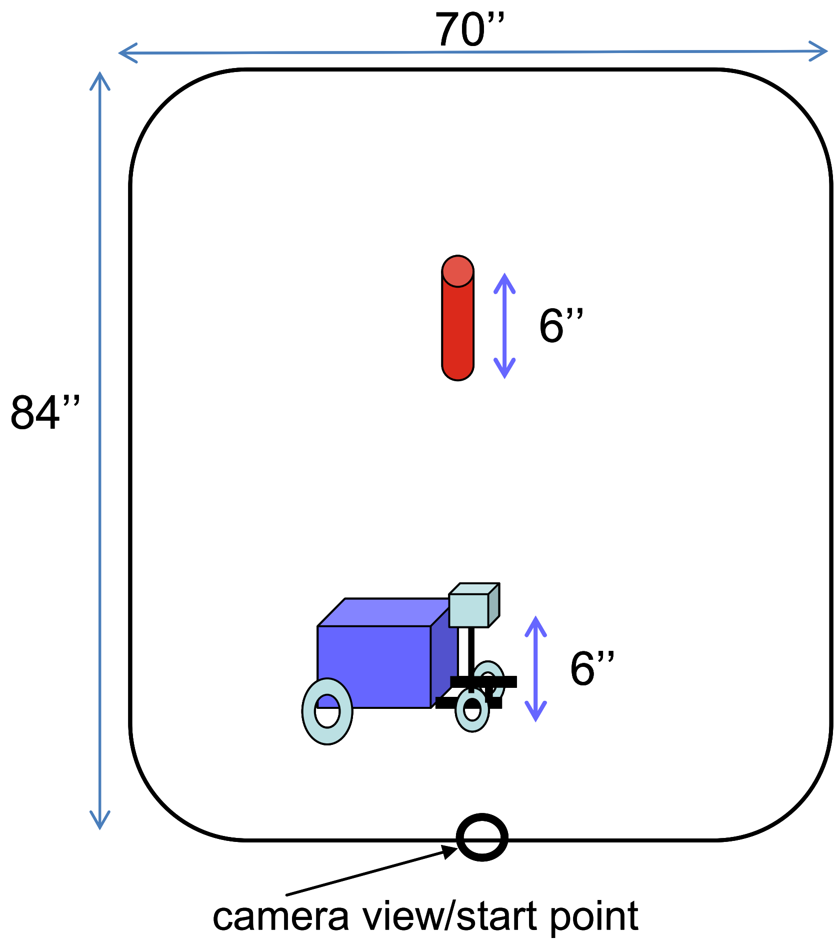

Figure 5: Schematic representations of the car and arena used for all experiments. The viewpoint of the camera mounted on the car was 6 inches as were the cylinder targets. The camera starting point was the same for all experiments and was marked on the arena floor with a piece of clear tape.

Experiment 1: Placement in Still and Moving Scenes

This experiment is intended to examine how perception of spatial arrangement of a set of targets changes as the stimulus changes from a static image to a motion video. The use of camera based images in this study allows the study of changes in perception based on properties of optic flow that are not associated with changes in body, head, or eye movement typically referred to as extra-retinal cues. Subjects regularly report depth compression when viewing still images, video that simulates forward motion, and in virtual environments. Cutting and Vishton (1995) point out that people use a large number of cues to derive structure in a natural scene, but no single source of information can account for the accuracy of human performance across a wide range of conditions. The narrow viewing angle often associated with navigating a remote vehicle with a video feed results in what is referred to as the keyhole or soda straw effect. These experiments are initial steps in trying to understand how navigation using a video medium. Observers viewed still images, from various camera-viewing angles, and moving images in which the type of motion being portrayed (forward, panning right and forward panning) was manipulated between conditions. The addition of camera motion provided additional cues such as motion parallax that could improve judgments about the spatial layout of a scene. Based on this information the following hypothesis was formed.

Hypothesis 1: Observers viewing moving images will place objects more centrally, in both X and Z dimensions, than when viewing still images.

Therefore, the goal was to assess subjects' perception of the spatial layout of a scene based on exposure to a single static picture or a brief motion video.

Method

Participants

Sixteen observers (4 males and 12 females) were drawn from Midwestern university psychology classes for this study. All subjects had normal or corrected to normal vision.

Stimuli

The stimuli were created by placing three targets on the floor of the arena in a centralized area that had been marked off into an invisible 3x6 position grid. The grid was created using metal washers glued to a poster board and hidden beneath the arena floor cover. Magnets were then glued to the bottoms of the targets so that targets would always be precisely placed within the grid positions. Rows and columns of the grid were spaced 4 inches apart so that the grid was twelve inches deep by twenty-four inches wide. This arrangement is illustrated in Figure 5. For each stimulus, one target was placed in each of the three rows, and each scene was coded for target placement and color. There were six camera conditions (static right +12°, static center, static left –12°, forward motion, forward motion with panning, and panning with no forward motion).

All stimuli were digitally captured using the LEGO™ system and Logitech™ camera mounted on the remote control vehicle. Static images were captured at a distance of 83 inches from the back wall of the arena. Panning movies were captured at a distance of 83 inches from the back wall of the arena; the camera panned to the right at a rotation rate of 22.6° per second. Movies that included forward motion were recorded with the initial starting point of the camera at 83 inches from the back wall and the camera moved forward at approximately 7 inches per second. The presented stimuli were 360 x 240 pixels and occupied the full screen of a 17 inch monitor (ViewSonic PF775, 31.5 x 23.5 cm screen size) with a refresh rate of 75Hz. Observers sat directly in front of the display and viewed it from a distance of 57 cm with their heads stabilized with a chin rest.

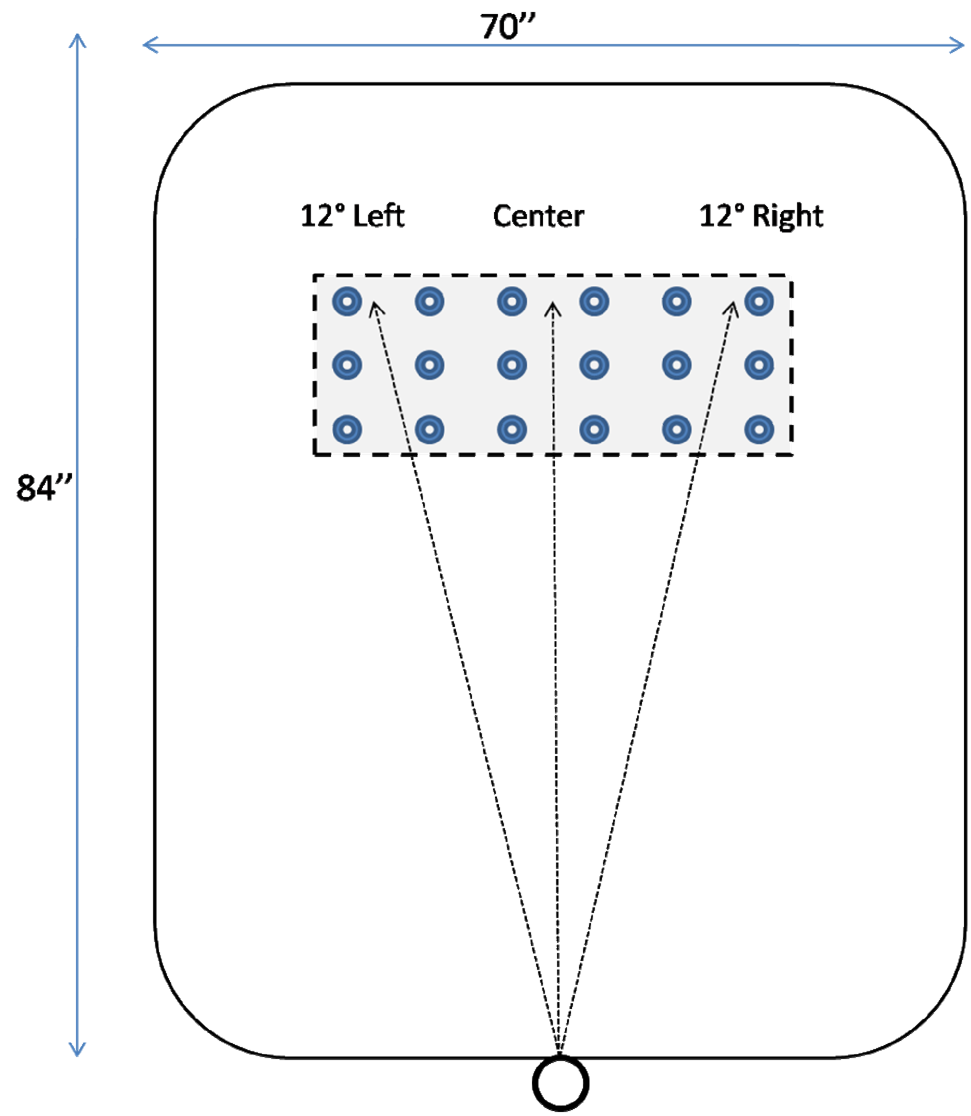

Figure 6: Arena layout for target placement experiment. The blue circles represent the possible target positions. The circle at the bottom center represents the observer's viewpoint. In any given scene, three targets were placed such that only one target was in each row and no two targets occupied the same column.

Procedure

Observers viewed 10 different scenes of each of the 6 camera conditions for a total of 60 trials and did not see the same target configuration twice. Observers were shown each scene for duration of 1 second. To account for any possible interactions between viewing the moving and static images in identifying target positions, the scenes were blocked according to image type. Within each block, trials were randomized and shown as an electronic slide presentation. This resulted in two slide presentations with 30 images in each. Half of the observers viewed the static stimuli first, and half viewed the moving stimuli first. After viewing each stimulus, they were asked to mark where each of the targets was located on a representation of the arena that included a grid and a mark representing the viewing position of the camera (see Appendix B).

Results

Analyzing this data presented a significant challenge. The goal of the experiment was to examine the specific influence of viewing angle and motion on spatial arrangement perception. It turned out that characterizing the nature of the responses in order to tease out the relevant differences was difficult. Several strategies were used in an attempt to quantify the judged spatial arrangement of the three targets (see Appendix C for a more thorough discussion). It was determined that the best solution was simply to take the X (lateral) and Z (depth) placements of each of the three targets. By using X and Z as separate variables, it was possible to observe lateral shifts and shifts in depth attributable to the manipulated camera conditions.

To analyze the effects of camera conditions, the average X and average Z positions were calculated for the judged target placements in each condition. The response grid was numbered in both the X- and Z-axes with 0,0 at the upper left corner of the diagram and the camera starting point at 10,24. For each trial, an X and a Z coordinate was recorded for each of the three targets in the scene. Then the three X coordinates were averaged together and the three Z coordinates were averaged together to yield an average X and Z coordinate for each trial for each observer. The X and Z coordinate data were analyzed separately in order to explore observer's perceived placement in depth and perceived placement to the right or left.

To test for the effect of stimulus order (viewing static vs. motion stimuli first), a Multivariate ANOVA was conducted with average X and average Z as dependent variables. Condition order was entered first, and then condition (grouped as static or moving scenes) entered second. There was no main effect of order for average X ($F(1, 947) = .132$, $p = .717$) or average Z ($F(1, 947) = .014$, $p = .906$) indicating that the order in which observers were presented the static and motion stimuli had no effect on their spatial arrangement judgments. Since there was no effect of order, we repeated the analysis ignoring order of presentation.

A repeated measures ANOVA was conducted to compare the effect of conditions on average X and average Z. There was a statistically significant difference in average X ($F (5, 9) = 130.082$, $p < .001$) and average Z ($F (5, 953) = 9.246$, $p < .001$) based on camera condition. A Tukey's HSD was then performed on each data set to identify homogeneous subsets of conditions for average X and average Z. Tukey's HSD revealed three subsets for average X; static images with the camera pointing to the right or left each produced significantly different X judgments (shifted to the left and right respectively, see Figure 7 for average X placements by condition) which were different from those produced in all other conditions (see Table 1). Hypothesis1, therefore, was partially supported since the still images with the camera centered also produced more centered judgments (along the x-axis) that were not significantly different from those for the moving stimuli.

Table 1. Tukey HSD results for average X judgments.

		Subsets for alpha = .05		
Condition	N	1	2	3
right still	160	7.3458		
forward pan	160		10.5958	
forward	160		10.9771	
pan	160		11.0604	
center still	159		11.1132	
left still	160			13.2521

Means for groups in homogeneous subsets are displayed. Right still and left still each fall into separate subsets while all other camera conditions fall into the same subset.

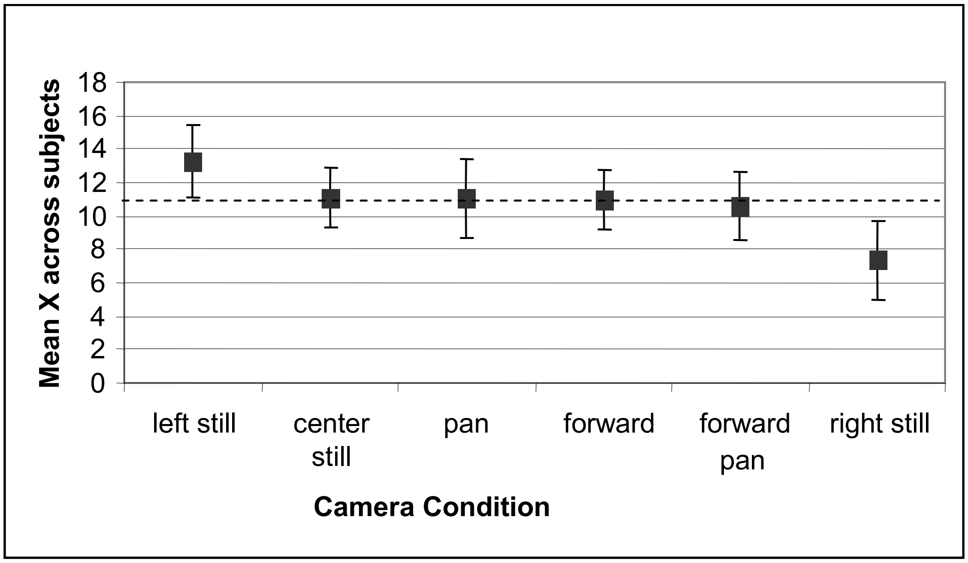

Figure 7: Average lateral judgments of target positions based on the X grid coordinate by condition. The expected average is represented by the dashed line. The error bars represent +/- 1 SD. Lower X coordinates indicate leftward lateral placement on the response grid relative to the camera view. Higher X coordinates indicate rightward lateral placement on the response grid relative to the camera view. Mean X grid coordinate for the left still condition was 13.25 and that for the right still condition was 7.35.

Two subsets of conditions for average Z (see Figure 8) were revealed by the Tukey's analysis (see Table 2), however, these did not follow the pattern expected. It was expected that camera motion (simulated self-motion) would result in more centered depth judgments for spatial layout than still images. The post-hoc analysis showed, however, that the panning condition produced a significantly different average depth placement of the targets than the rest of the conditions. While all of the motion conditions did result in depth judgments that were further from the observer than the still images, the panning condition produced perceived target depths very near veridical.

Table 2. Tukey HSD results for average Z judgments.

		Subsets for alpha = .05	
Condition	N	1	2
pan	160	9.1875	
forward pan	160		10.4146
forward	160		10.6708
left still	160		11.0583
center still	159		11.5031
right still	160		11.5917

Means for groups in homogeneous subsets are displayed. Smaller numbers reflect placement further from the camera position. The panning condition falls into a subset by itself while all other conditions fell into a single subset.

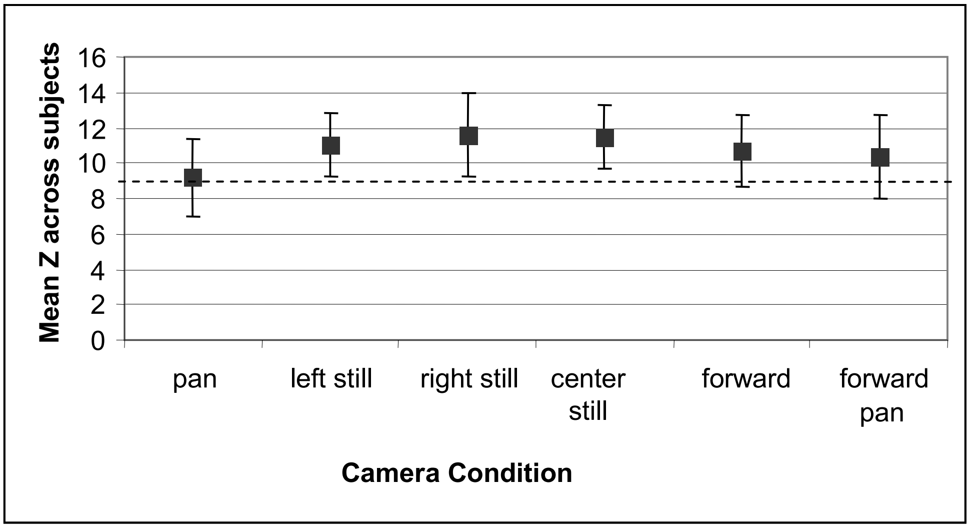

Figure 8: Average judgments of target positions in depth based on the Z grid coordinate by condition. The expected average is represented by the dashed line. The error bars represent +/- 1 SD. Mean Z coordinate for the panning condition was 9.19. Lower Z coordinates indicate targets being placed further from the camera on the response grid.

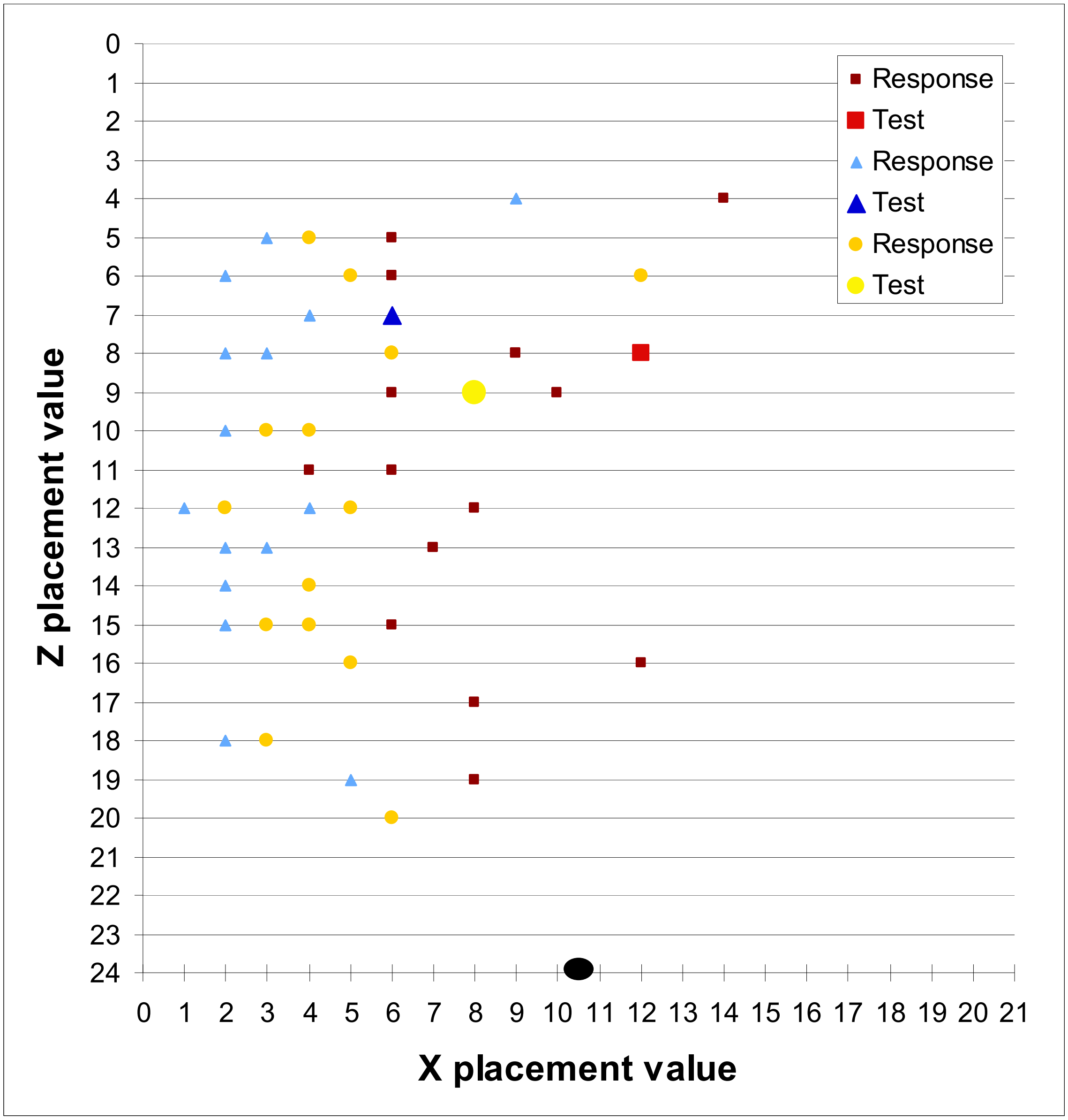

Figure 9: Individual responses to a static right camera angle test stimulus. The red square responses were for the red square test, blue triangle responses for the blue triangle test, and yellow circle responses for the yellow circle test. Test refers to the three target locations presented in a trial. The axes are arranged so the camera origin (marked by the black dot) is at 10, 24. Responses with higher Z values were closer to the camera than responses at lower Z values. While there is a great deal of variance in the individual responses, it can be seen that they are mostly shifted to the left and forward.

Discussion

The results indicate that when the images being viewed were static, observers exhibited a bias in their judgments of target positions within the environment. This was expected because the observers had no context providing information about the camera direction and thus when, for example, the targets were on the left of the image they judged the targets to be on the left of the environment. There was quite a bit of individual variation in the data as illustrated in Figure 9. This may have been due to the complexity of the task. Thorton & Gilden (2001) showed that tasks involving more visual complexity were subject to greater levels of variation in performance.

The surprise in this data was the effect of panning. While observers viewing panning in conjunction with forward motion still exhibited a sense of depth compression (objects were perceived closer to the camera), panning motion alone seemed to mitigate the depth compression effect. Camera panning appeared to add a greater sense of depth to the scenes and observers judged the targets to be further away from the camera than in all other conditions. Figure 10 illustrates the overall pattern of the data, showing the rightward, leftward, and compressed tendencies in the responses.

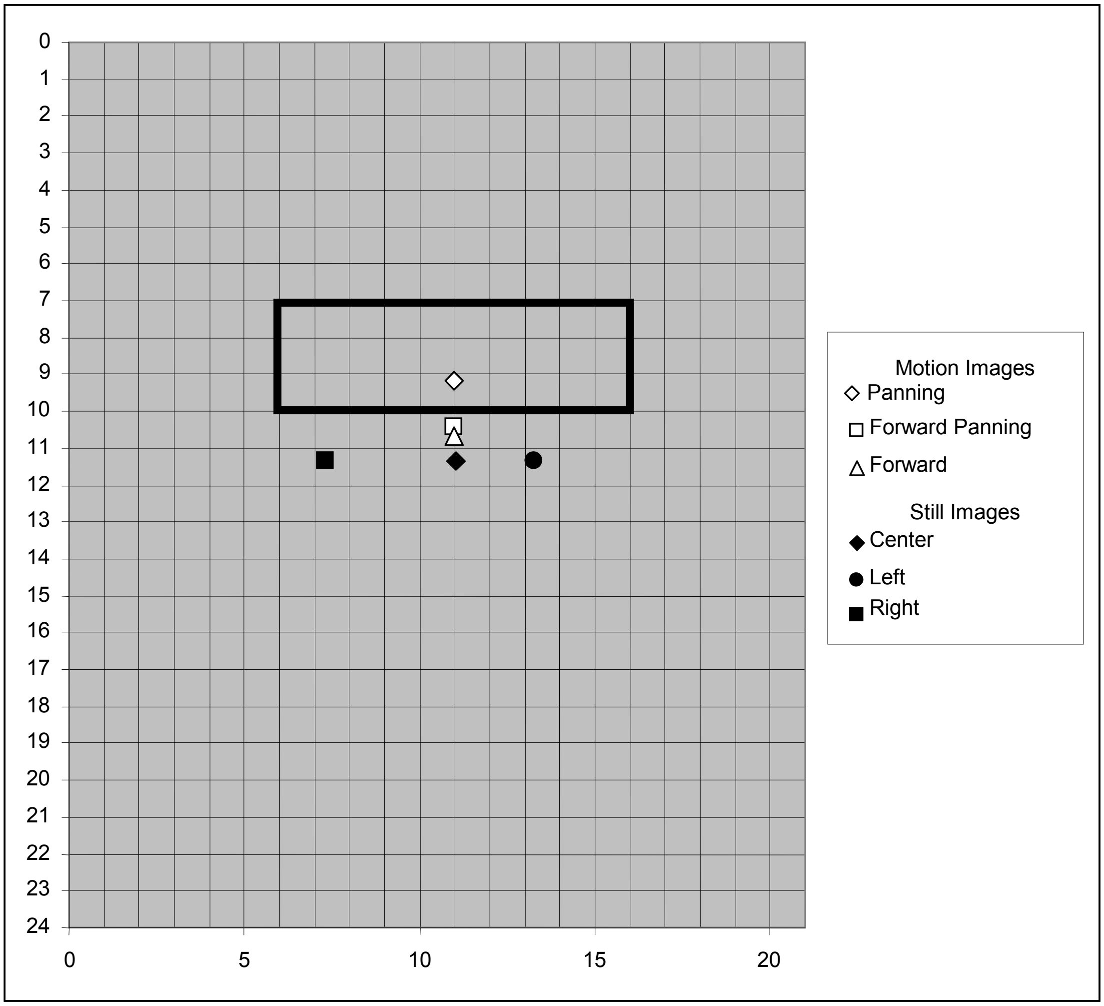

Figure 10: Averaged X and Z response coordinates for all conditions. The box represents the region of physical target placement in the stimuli based on a scale conversion to grid coordinates. Because there was no significant difference in the Z judgments of static scenes the Z judgments were averaged in the static conditions (filled symbols). The x placements of the moving conditions (open symbols) were averaged because there was no statistically significant difference in the x placements in the motion conditions.

The most likely explanation for the greater sense of depth in the panning condition is motion parallax. Previous research (Nawrot, 2003) has shown a linkage between eye movements and perception of depth from motion parallax. In fact, the ability to benefit from motion parallax was tied specifically to eye movements and not to other types of visual image rotations. The results obtained here suggest that efference copy generated by eye movements is not required to support depth perception from motion parallax. However, the benefit of motion parallax from panning appeared to be smaller when paired with forward motion.

These results suggest that video sources for remote controlled systems can better support the creation of a mental representation by providing a flexible point of view. Motion parallax is a natural consequence of moving through the environment, but panning through the scene clearly supports greater depth perception by observers than panning combined with forward motion. It is not clear how independent panning might alter the soda straw effect, but this should be further explored. The greatest benefit of panning, therefore, may be in applications that involve smaller overall movements and close-in situations such as tele-robotic repair work.

Experiment 2: Judging heading during rotation

This study is based on perceived heading research reported by Li and Warren (2004), Banks et al. (1996), and Royden, Banks, and Crowell (1992). Though these studies, and others, have shown that optic flow is useful for navigation, some questions remain. For instance, how does the visual system treat extra-retinal signals? Extra-retinal signals are those signals created by eye and head rotations independent of body translation. One view is that the visual system recovers direction instantaneously from the retinal flow created by the combination of optic flow and a fixations on objects in the environment (parallax) (van den Berg & Brenner, 1994; Warren & Hannon, 1988). The alternate view is that the visual system accounts for these extra-retinal signals through efference copy or feedback signals from the eye-movement system (Banks et al., 1996; Royden et al., 1992; Royden & Conti, 2003).

The most common paradigm for testing these two hypotheses has been to simulate eye/head rotations using a computer display and then ask observers to make judgments about heading based on the simulation. Generally, observers make errors regarding heading when eye or head rotations are simulated on computer displays, but the results have been mixed in terms of the magnitude of the errors. Royden, Crowell, and Banks (1994) found that observer estimates of heading were more accurate when eye rotations were executed than when they were simulated on a computer screen. The magnitude of observer error in heading estimation during simulated eye movements was proportional to the simulated eye rotations (up to 12°). Li and Warren (2003) found smaller errors of 1° to 3° using displays with greater detail and reference objects that provided motion parallax information. Factors thought to affect the results have included the wording of the instructions (whether observers were told to expect a straight or curved path or were given neutral instructions) and the degree of depth and detail available in the scene (Li & Warren, 2004). Scenes presented with random dots generated in the fronto-parallel plane resulted in greater heading misalignment than scenes containing monocular depth cues and motion parallax. Li and Warren (2004) found that depth range and dense motion parallax were not essential for accurate perception, but that expectations of curved or straight path travel did affect judgments. When viewing a simulated straight path of travel and observers expected a straight path of travel their responses were more accurate than when they expected a curved path of travel or were given neutral instructions.

It should be noted, even if heading errors are as small as 1° of visual angle, 1° can functionally be very large depending on the distance at which the objective is being viewed. For instance, at 57 cm 1° of visual angle is equal to 1 cm, while at 57 meters 1° of visual angle would be equal to 1 m.

The exact nature of the role of extra-retinal signals becomes important in the examination of operator performance in video based remote navigation tasks. This type of task requires the operator to recover heading information using only the visual information on the screen. If, as Pack and Mingolla (1998) suggest, induced motion illusions are attributable to the same neural mechanism that subtracts extra-retinal signals, then it would be reasonable to expect observers to make errors in heading judgments in video based remote navigation tasks because of the absence of meaningful extra-retinal signals. This led to the following two experimental hypotheses.

Hypothesis 2a: Observers will accurately judge heading from forward camera motion (virtual self-motion) as long as the camera direction is fixed relative to the observer's point of view.

Hypothesis 2b: Perceived heading will be displaced in the opposite direction of panning for stimuli created with camera panning coupled with forward camera motion.

Method

Participants

Four observers, 3 males and 1 female ranging in age from 21 to 40 years, participated in this experiment. All subjects had normal or corrected to normal vision.

Stimuli

Observers were shown a movie of a scene that contained a single target 66 inches from the starting position of the camera lens (see Figure 11) and two reference objects positioned 17 inches beyond the target and each offset by ½ inch from the center of the target, one to the left and one to the right. The movies were recorded by the Logitech™ camera as it, attached to the remote LEGO™ vehicle, moved forward through the arena at approximately 7 in/sec for one second. The vehicle/camera traveled along one of five paths that, if the vehicle continued beyond the one-second stimulus duration, would have passed by the target by 5° or 10° (left or right) or run directly into the target. The camera was set on the remote vehicle to have a constant angular difference from the direction of vehicle travel by 0° (centered), +7°(right offset), or −7°(left offset). There was also a camera panning condition in which, during the recording of the movie, the camera panned once through the scene from right to left at a rate of 22.6° per second as the vehicle moved forward. This resulted in a total of five path conditions and four camera conditions.

Procedure

The movies were ordered randomly to create 5 full screen electronic slide shows. Each slide show contained 100 movies lasting 1 second each. Observers were shown each movie clip and asked to judge whether the camera would pass to the right or left of the target. Observers started each timed trial by pressing a button. When the trial was complete, the observer was immediately shown a blank screen that remained until the start of the next trial. Observers were not permitted to replay a trial once it had been shown. Each block contained 100 trials (5 paths x 4 camera conditions x 5 trials of each condition) and each observer completed five blocks. The point of subjective equality (PSE), the direction of vehicle travel that resulted in 50% judgments that the camera would travel to the left or right of the target, was determined from the percent correct data for each observer and camera condition using Probit analysis (Finney, 1971).

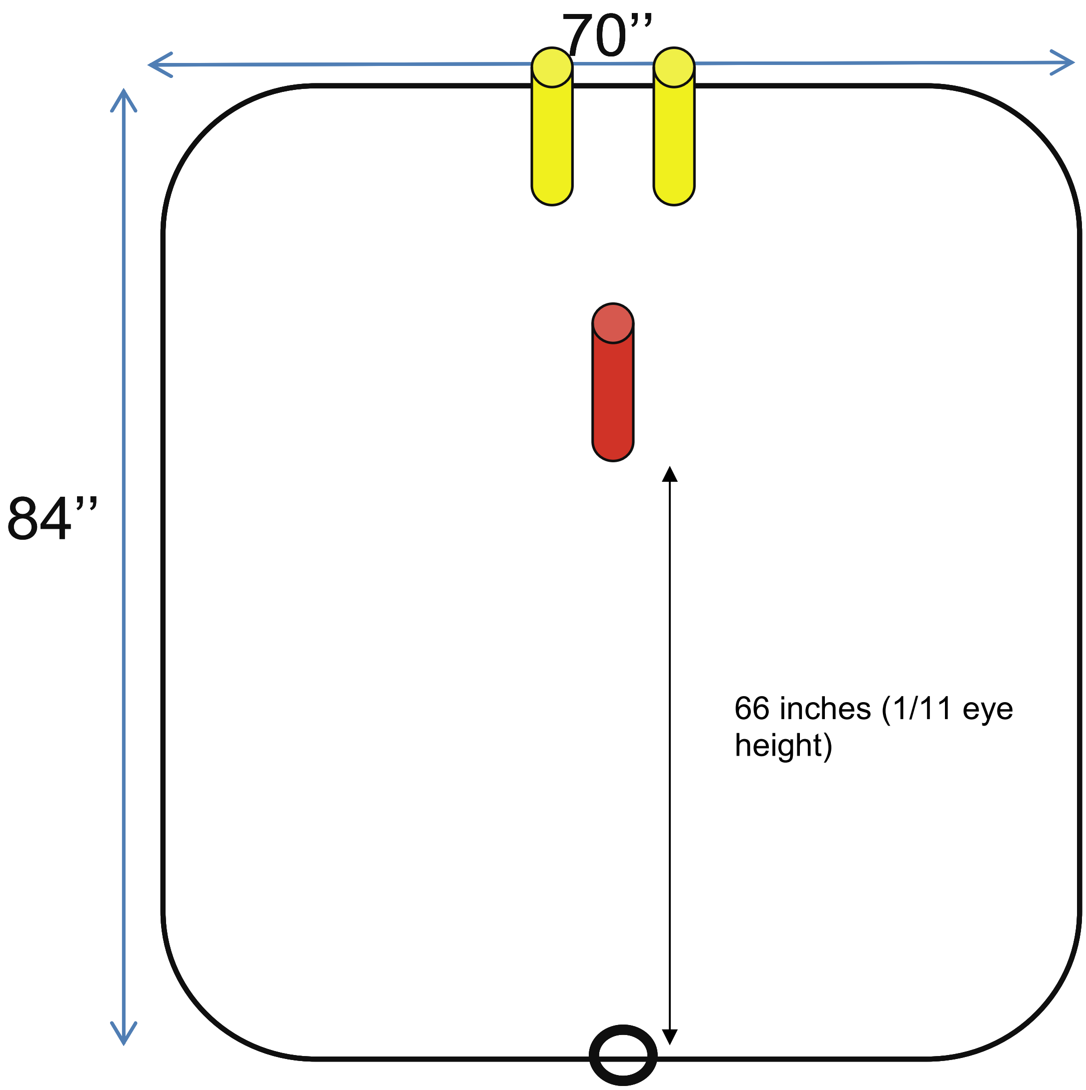

Figure 11: Arena layout for path perception experiment. Observers were asked to indicate if the path of the camera would pass to the right or the left of the target in the middle of the arena. Two reference objects were placed at the back of the arena to support motion parallax while the camera was moving.

Results

No PSE's could be determined for the panning condition because observers never perceived the camera as passing to the right of the target for the range of path conditions used. For all other camera conditions, the PSE's ranged from -4.4° to 2.5° (see Figure 12) meaning that performance in this experiment was noisier than what has been observed in the past for non-rotating conditions. This performance is not consistent with hypothesis 2a that observers would accurately judge heading from forward camera motion (virtual self-motion) as long as the camera direction was fixed relative to the observer's point of view.

The fact that observers never saw the camera as passing to the right in the camera panning condition suggests that they were unable to discern the forward path of the camera in the presence of visual rotation. Observers commented that in the panning condition they always perceived a curved path passing to the left of the target and were not able to perceive paths to the right of target. This shows an unequivocal effect of visual rotation on heading judgments, and is consistent with hypothesis 2b stating that perceived heading would be displaced in the opposite direction of the panning camera motion. Unfortunately, it was not possible to quantify this effect with any accuracy because PSE's for the panning condition could not be determined because the range of path angles tested was too small for observers to reliably perceive a change in the path direction. However, this implies that in this condition, PSEs were larger than 10°, the largest angular offset between the target and path tested.

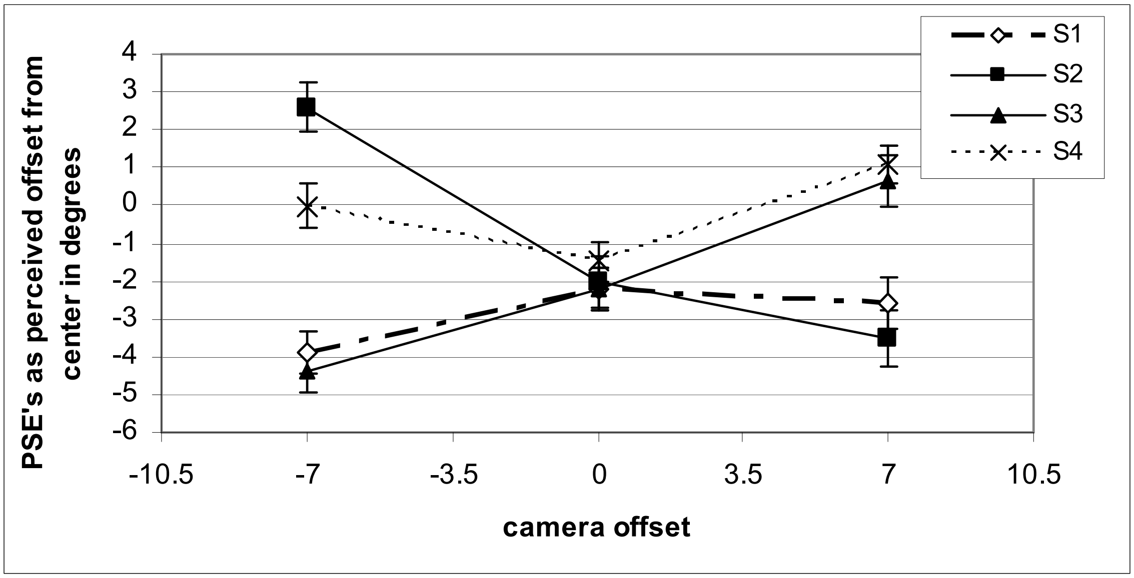

Figure 12: Observer mean PSE's are shown for -7 (left offset), 0 (centered),+7 (right offset) camera conditions. The error bars represent ±1 standard error. In this plot, negative numbers represent a leftward response and positive numbers represent a rightward response. The data show that heading discrimination performance was inconsistent across observers for the different camera offsets.

Discussion

Observers never judged the camera to be passing to the right when the camera panned to the left (panning condition). These results are consistent with models of visual processing that utilize a subtraction process in which extra-retinal signals are subtracted from the retinal signal to produce a percept (e.g., Barlow, 1980). It also suggests that models relying solely on optical variables would not be sufficient to account for path perception in these circumstances. Such models would predict that observers would have little difficulty in interpreting the direction of heading even in the presence of visual rotations (e.g. Li & Warren, 2000). Our observers exhibited an effect of rotation consistent with Royden et al. (1994) who reported that the magnitude of observer error in heading estimation during simulated eye movements was proportional to the simulated eye rotations. Studies showing more accurate performance used rotation rates of less than 1°/second (e.g. Warren & Hannon, 1988). The rotation rate in this study, 22.6°/sec, is arguably more consistent with rotation rates used by Banks et al. (1996) and Royden et al.(1994). Some studies have shown that when response markers were further away in virtual displays, observer judgments of heading were more affected by simulated rotations (Ehrlich, Beck, Crowell, Freeman, & Banks, 1998). Ehrlich et al. examined the effect of depth information on observer judgment in heading estimation during simulated rotation and found that neither binocular not monocular depth cues improved performance in the absence of extra-retinal information. However, when the posts used in the display to mark the point at which observers were to declare left or right passage were depicted as being further from the observer, the effect of simulated rotations was greater. The ranges of viewing distances for the Ehrlich et al. study were from 250 cm to 2000 cm. This is not an adequate explanation for the increased size of the effect of rotation seen in the present study because the starting viewing distance was 66 inches or 167.64 cm, which is well below the smallest viewing distance of 250 cm and should therefore be associated with a smaller effect of rotation. If depth of the response marker were an issue in this study, the effect should have been smaller rather than larger. The most plausible explanation of the large effect observed in this study is the higher rotation rate.

Asking observers to estimate the speed of an expanding flow pattern in an head mounted display, Durgin, Gigone, and Scott (2005) found that subjective magnitude estimation of speed from visual flow could be reduced both by active self-motion (regular and treadmill walking) and by passive self-motion (e.g. being pushed forward or backward on a chair). They attributed this self-motion induced reduction in perceived speed to a "subtractive" operation, rather than a reduction in gain. This is consistent with inhibition theory (Barlow, 1990). According to Barlow, highly correlated events such as walking and an expanding flow pattern mutually specify each other. Consequently, the perceptual system uses this redundancy to modify its sensory coding. The results obtained in this study are consistent with models based on inhibition theory.

Crowell, Banks, Shenoy and Anderson (1998) performed a series of experiments to quantify the effectiveness of different extra-retinal cues in mediating accurate self-motion judgments during head turns. As well as measuring performance during eye pursuit, three possible sources of extra-retinal information were examined independently and in combination: efferent information about motor commands to the neck muscles, proprioceptive information from the neck muscles and vestibular semicircular canal information about head rotation. They found that neither neck proprioception nor vestibular information alone was sufficient for accurate perception of heading. The combination of all three extra-retinal sources, however, supported veridical judgments of direction of locomotion. Again, the results of the present study suggest that efference is important for judging heading.

Stone and Perrone (1997) pointed out that in these studies of heading estimation subjects were not asked to estimate their current heading, but rather asked to project their path into the future. Therefore Stone and Perrone asked observers to adjust a dial indicating their current direction of travel. Using this methodology, observers reported their instantaneous heading accurately at rotation rates up to 16°/sec. This seemingly slight variation in instructions is an important distinction because it hints at the difference between open loop and closed loop performance. In closed loop actions, it is only necessary to be able to respond to perceptual cues with sufficient speed for the task and environment. In open loop performance, the ability to accurately project into the future becomes important.

It may be that the more important question here is the ability to judge rotation rates. Perception of rotation rate (e.g., how fast am I turning) seems to be more important for projecting movement in the future than instantaneous heading (e.g., which way am I pointing right now). The closest approximation of open-loop behavior in a remote navigation task is the act of finding things that are not in the current field of view such as a target that is off screen or knowing where the operator has been rather than where they are going. Further research is required to fully explore this phenomenon; however, the next experiment addresses this to some extent.

Experiment 3: Navigation in a remote environment

As mentioned in the introduction, a prominent question in the literature regards the use of optic flow as a strategy for navigation. The fact that people ***can*** use optic flow to determine direction of heading (Warren & Hannon, 1990) does not mean that an optic flow strategy is the dominant or optimal strategy for visual navigation. In fact, several optical variables have been shown to support optimal strategies for action such as optic flow, rate of expansion, or Tau. However, optimal strategies are not necessarily the strategies used by the majority of human performers. As an alternative to the optic flow strategy, the target-direction strategy was explored by Rushton, Harris, Lloyd & Wann (1998). Rushton had observers wear prism goggles that imposed an offset to target-direction, but did not disturb optic flow. The expectation was that if observers were relying on an optic flow strategy, they should follow a relatively straight path to a target placed directly ahead of them. Alternatively, if observers were using a target-direction strategy, they should follow a curved path based on constantly correcting for the visual displacement caused by the prism goggles, which is what participants did.

Harris and Carre (2001), using the same prism goggle methodology, found that when a greater level of ground texture was available, participant performance did not show the same degree of path disturbance. Specifically, observers were instructed to direct their gaze at the ground on some trials. On those trials, the paths showed less of the characteristic curvature. Harris and Carre argued that the more dense texture provided by the carpeted ground allowed participants to use an optic flow strategy.

This experiment was intended to explore the proposed navigation strategies of optic flow and target-direction identified by Lappe, Bremmer van den Berg, A. V. (1999a&b), Rushton et al. (1998), and Harris and Carre (2001). Here, the participants had to navigate a remotely operated vehicle through an obstacle course under two camera conditions: a fixed-camera condition in which the camera providing visual feedback always pointed directly in front of the vehicle and a yoked-camera condition in which the camera turned in correspondence with eth steering wheels. The optic flow strategy was approximated by the fixed camera view which is common for search and rescue robot platforms, and the target direction strategy was approximated by the camera view that was yoked to steering.

Hypothesis 3: When navigating a remote vehicle with visual feedback provided by only a camera attached to the remote vehicle, observers will perform better when the camera view is coupled to steering (yoked-camera condition) than when the camera view is fixed relative to the orientation of the remote vehicle (fixed-camera condition).

Method

Participants

Observers for this study included 5 males and 1 female. All subjects had normal or corrected to normal vision. Subjects ranged in age from 20 to 41. These observers did not participate in the previous experiments.

Stimuli

In this experiment, observers were asked to remotely navigate a LEGO robotic vehicle through an obstacle course comprised of gates identified by colored cylinders and then make contact with a target cylinder with the front of the LEGO vehicle (see Figure X for layout) without knocking over the target. The height of each of the cylinders was 6.1 inches or equal to one eye height relative to the height of the camera placement. Observers controlled the vehicle using a standard LEGO Mindstorm multi-button programmable handheld remote control (model 9738) which controlled two bi-directional motors: one moved the steering mechanism left and right while the other engaged the driving wheels forward and backward. It should be noted that this remote control does not re-center when the control buttons are released. Thus, the steering mechanism did not automatically return to the straight-ahead position after being turned right or left, but this was done manually by the observer via the remote control. Environmental layout information was provided exclusively by visual feedback provided by a video camera mounted on the remote vehicle. There were two camera conditions: 1) a yoked camera condition in which the camera viewing direction was yoked to the steering wheels (supporting a target-direction strategy), and 2) a fixed camera condition in which the camera always pointed straight ahead relative to the direction of the remote vehicle (supporting an optic flow strategy).

Procedure

Subjects completed three sessions of 10 trials for 30 trials each. Subjects were told at the beginning of each trial, which of the two intermediate gates (left or right) to go through first and which target (left or right) to contact (see Figure 13). An incomplete trial resulted if, a) the final target was not contacted within 5 minutes, or b) if the vehicle was disabled by over-steering. Incomplete trials were repeated at the end of the block. In order to encourage accuracy and speed during the trial, an observer's performance on each trial was graded with points awarded for successfully moving through each gate and for making contact with the target cylinder. Points were subtracted for knocking over cylinders other than the target, and a final bonus was awarded for completing the trial within specified time limits (see Appendix B for sample scoring sheet).

Navigation performance on every trial was recorded by a camera located directly above the navigation arena, providing a 'bird's eye view' of the movement of the remote vehicle through the course. Once during each of the three blocks/sessions of trials the subject was asked to rate their own performance in the most recent trial based on knowledge of the space that they were acting in and their ability to control the vehicle. Two interviews were conducted during the first block of trials in which the participants were asked to think aloud as they completed the navigation task. The observers were asked to provide a commentary of their thoughts as they were navigating the vehicle in the arena employing a think aloud protocol. The think aloud trials occurred after the second and eighth trial of the first block of trials. The think aloud protocol was repeated in the third trial in each of the two final blocks of trials.

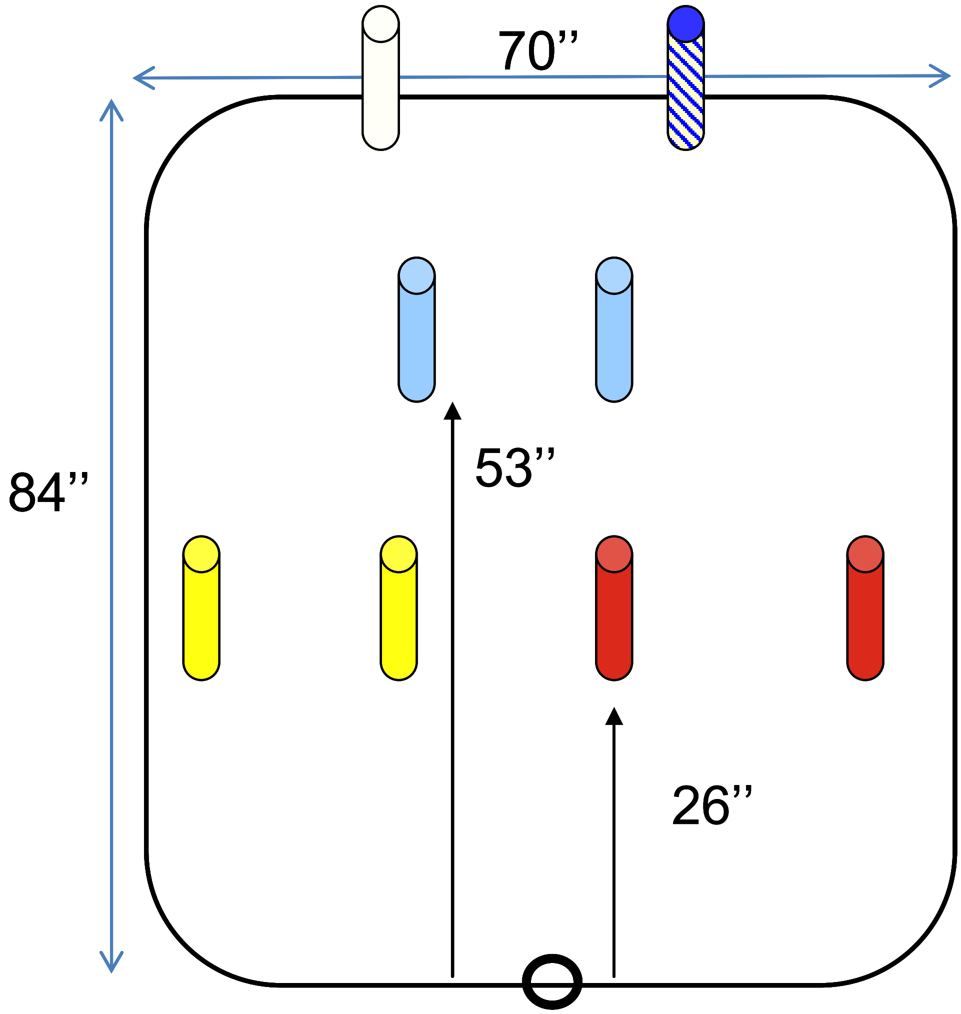

Figure 13: Schematic representation of the navigation experiment layout of obstacles

To evaluate the video data, the path of the RPV for each trial was digitally recorded and transferred into coordinates on an X-Y plot. This was accomplished by viewing each of the overhead videos on a computer and, using the display's pixels as a localizing grid, recording the coordinates of a specific point on the front of the RPV every ten frames (1/3 sec). Quantifying the vehicle's path of travel in this way provided a visual representation of the path traveled in addition to supporting statistical analysis of the data.

Results

Both groups were able to complete the task and improved over the three sessions. The fastest and slowest trial completions were 15 and 201 seconds respectively. The average time to completion for all subjects in all conditions was 49.24 seconds. An analysis of variance was condition to test for differences in performance based on time to completion, average speed, and points awarded. The analysis for this data was done in the traditional way for psychophysics, in which we have a small number of subjects, but every trial is considered and independent measurement and the error term for effects is the residual without any subject variability removed. The analysis revealed that there was a significant main effect of condition for average speed ($F(1,174) = 8.261$, $p = .005$), time to completion ($F(1,174) = 28.934$, $p < .001$), and average points per trial ($F(1,174) = 10.949$, $p < .001$). Speed differs from time in that it incorporates forward and backward movement while time is based solely on completion of the task. Points actually reflect the overall precision of navigation as well as time to completion (see Appendix B for sample score sheet). In general, the yoked-camera condition, which supported a target direction strategy, seemed better suited to instinctual navigation; observers in the yoked-camera condition performed better than those in the fixed-camera condition on all measures (average speed, time to completion, and number of points, see Figures 14 & 15).

An additional analysis was conducted using a measure called tortuosity. Tortuosity is a way of describing the "*twistedness*" of a curve. The simplest way of estimating tortuosity is the arc-chord ratio or the ratio of the length of the curve (L) to the distance between the ends. Such a method has been used as a way of describing complexity in locomotion and performance differences in control interfaces for remote navigation (Voshell, Woods, & Phillips, 2005). We used this method to calculate a

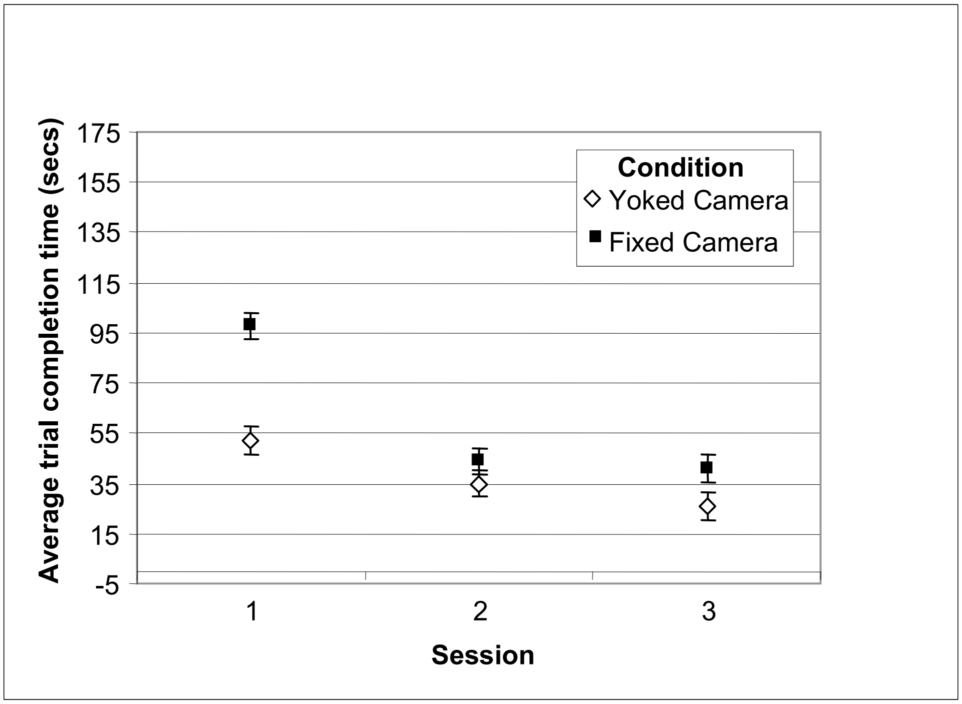

Figure 14: Average time to complete the navigation task is displayed in seconds by condition and session. The error bars represent +/- 1 SE. Observers in the yoked-camera condition had lower average completion times through all sessions and showed steady improvement through the third session. Observers in the fixed-camera condition had much higher completion times in the first session, showed a more dramatic drop in completion times in session 2 and minimal improvement between the second and third session.

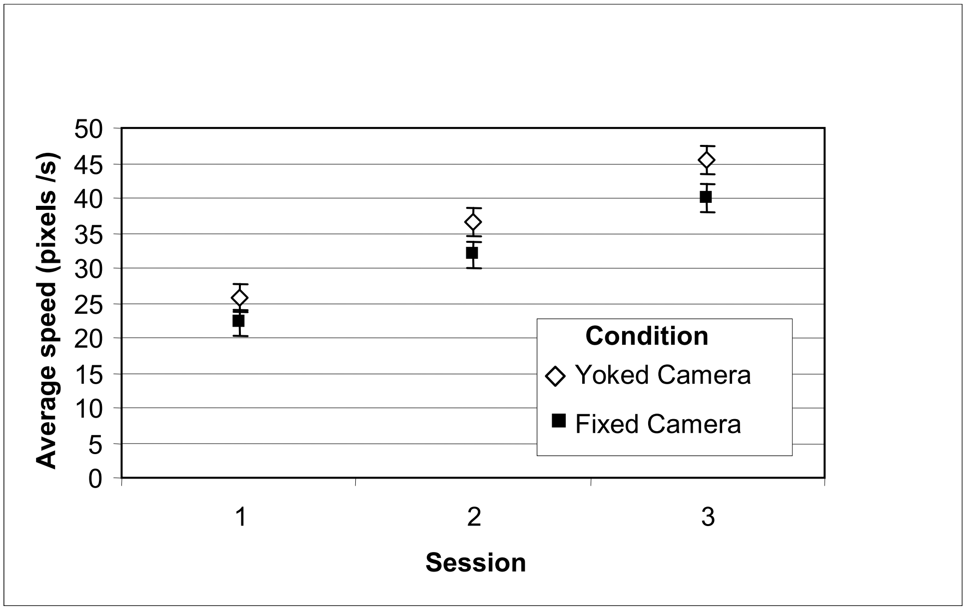

Figure 15: Average vehicle speed (in cm/s) is shown for the yoked and fixed camera conditions. The error bars represent +/- 1 SE. While both groups showed improvement in performance over the 3 sessions, the speed of the vehicle increased in all conditions, the yoked-camera condition always supported faster vehicle speeds than the fixed-camera condition. Speed was calculated from the raw data which is based on pixels/334 msec - 1 cm is equal to 4.12 pixels.

tortuosity score for each trial. The results show a significant effect of camera condition on performance ($F(1, 4) = 11.608$, $p = .027$, see Appendix C for the ANOVA table). Observers in the yoked-camera condition performed better (had less twisted paths) than those in the fixed-camera condition. As can be seen in Figure 16, performance of observers in the fixed condition was noisy in comparison to that of observers in the yoked camera condition.

The data from the think aloud protocol turned out to be unusable. This was primarily due to the recordings being largely inaudible and in some cases non-existent. The recording equipment seemed sensitive enough during tests, however, when subjects were carrying out the navigation task, they tended to mutter or grunt rather than actually narrate their thoughts. In some cases they did not speak out loud even after being given several prompts to think aloud. The navigation task may have required a level of concentration that interfered with the ability to narrate concurrently with the task.

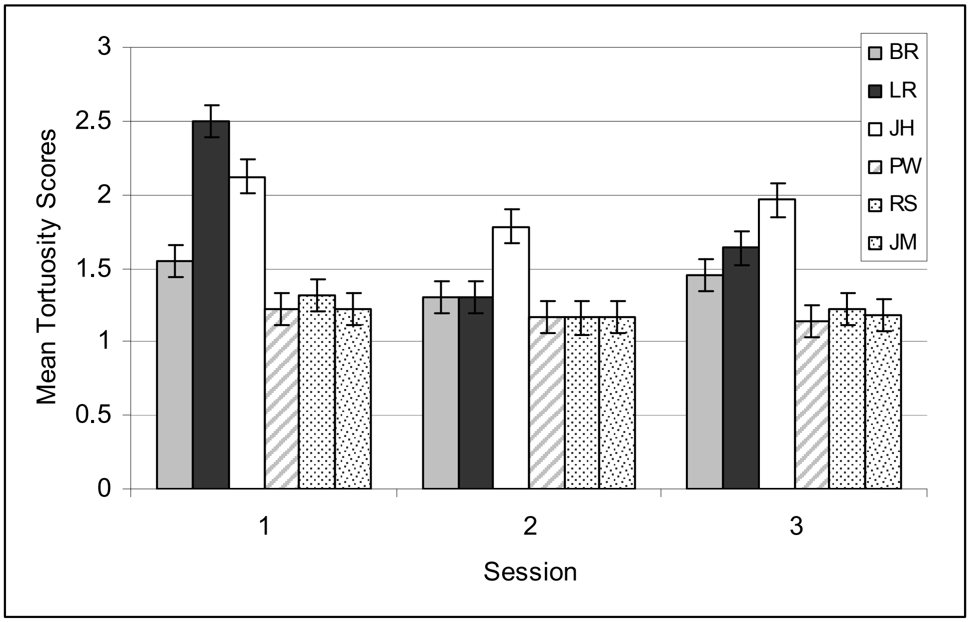

Figure 16: The mean tortuosity scores as a function of session for each observer. The error bars represent +/- 1 SE. The solid bars indicate that the observer was in the fixed camera condition and the patterned bars indicate the observer was in the yoked camera condition.

Observers in both camera conditions were able to complete the course in the allotted time. However, subjects in the yoked camera condition were able to complete the course with lower tortuosity scores than subjects in the fixed camera condition. The performance in the remote navigation task does not offer a straight forward interpretation. There was some evidence of difficulty in judging depth and rate of rotation. However, subjects in both camera conditions were able to complete the task in the allotted time and it is possible that the differences between the two camera conditions would have disappeared had we measured performance over additional sessions. Given that the input for turns and for forward and backward motion was a button push, it is not likely that the learning of the system can be explained by efference copy.

Discussion

This navigation experiment was conducted to begin an exploration of the strategies for navigation of a remotely piloted vehicle. The yoked-camera condition was designed to support a target direction (point and go) model of steering. The fixed-camera condition was designed to support an optic flow model of steering. Subjects in the yoked-camera condition completed the task more quickly than subjects in the fixed-camera condition. The difference in performance between the two groups persisted through all trials of the experiment. Although there was a statistically significant difference in performance, it is not clear that this represents a practical difference. Both groups successfully completed the task and it is possible that the differences in performance would have disappeared with continued sessions. The question of practical difference must be answered in order to gain better understanding of the requirements and benefits of different types of navigation interface.

As was seen in Figure 16, performance was somewhat irregular for observers in the fixed camera condition. This might be evidence of complexity (Thorton & Gilden, 2001) such as was seen in the path perception experiment. Just as in the path perception experiment and in the forward panning condition of the spatial arrangement experiment, forward motion and rotation occur simultaneously.

Murphy and Burke (2005) clearly identified situation awareness or the building of mental models as a major challenge in human robot interaction. All observers in our experiment reported that they felt as if they had a good mental model of the space that they were navigating through and a good mental model of the capabilities of the remote vehicle. However, the fixed-camera group exhibited a pattern of navigation that could be interpreted as showing a discomfort with interpreting either spatial information or vehicle dynamics. Observers in the fixed-camera group regularly passed far enough through the gates to ensure adequate distance for backing up and making the turn for the next gate or target. This pattern illustrated in Figures 17 - 19 suggests several things. First, it suggests that these subjects may have been uncomfortable with judging the rate of turn of the vehicle. Rather than trying to make a gradual turn during forward movement, this group would typically stop, turn the wheels slightly while backing up and then move forward as if to check alignment. They would repeat this sequence until satisfied with the alignment and then move forward through the gate. These results suggest that while extra-retinal information is not necessary for successful performance, it likely makes performance more efficient.

The backing/turn pattern could also indicate that subjects were not comfortably familiar with the capabilities of the remote vehicle since they avoided the optimal strategy of small turning adjustments during forward travel that was adopted by the subjects in the yoked-camera group. Arguably, this strategy should have worked equally for both groups because it simply relies on aligning the focus of expansion with the center of the viewing field or optic flow. Yet, the participants in the fixed camera condition never adopted behaviors indicating use of the optic flow strategy.

Finally, the pattern may indicate some discomfort with depth judgments. If subjects were comfortable with their ability to judge distance of forward self motion they should have been able to judge when they had passed through the gates and thus started their backing/turn procedure earlier. The narrow viewing angle of the camera made it difficult for all observers to judge forward progress relative to things that had passed from view, but the group with the fixed camera appeared to have adopted this unique strategy to locate objects in the periphery. Another behavior pattern consistent with a lack of confidence about depth judgments was shown by the yoked-camera group who would stop and turn the camera view (by turning the steering wheels) as they approached the gates presumably to check position relative to the gate markers.

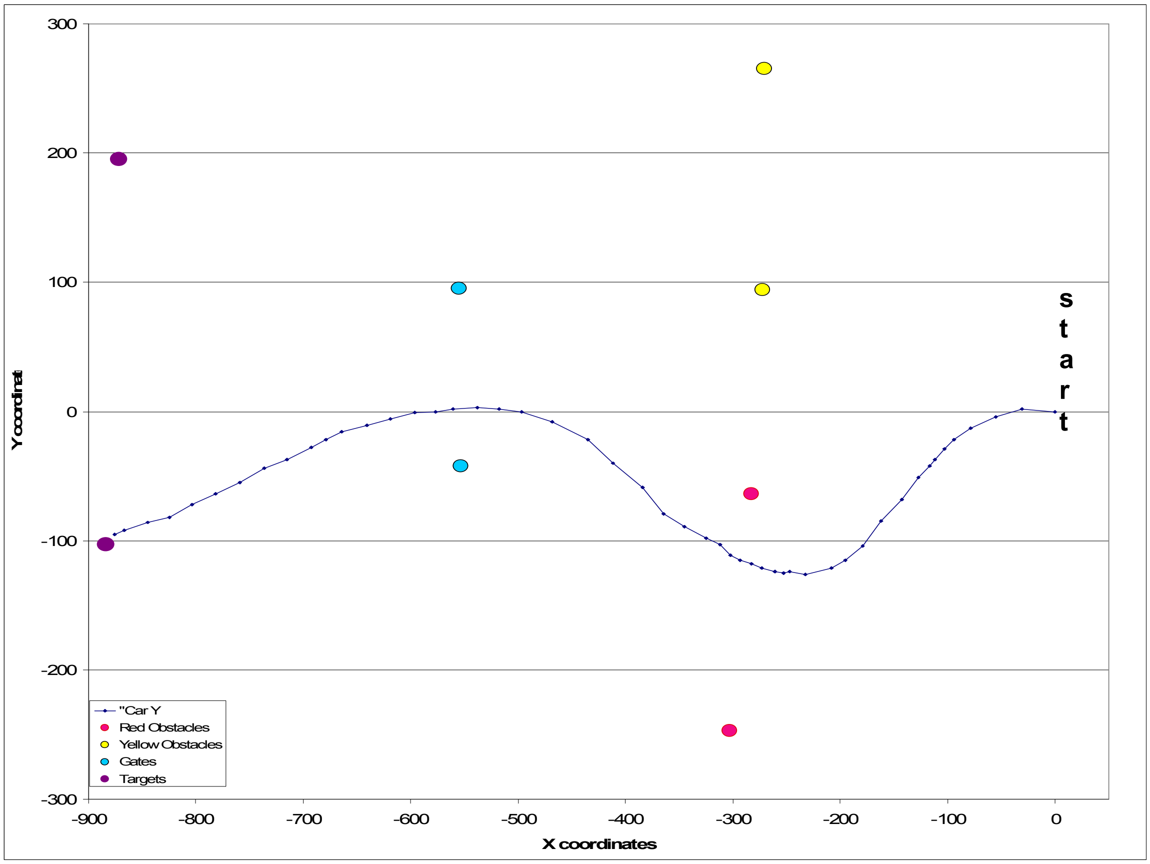

Figure 17: This figure depicts the actual path taken by an observer in the yoked camera condition. The path is near optimal with smooth navigation and a well regulated speed.

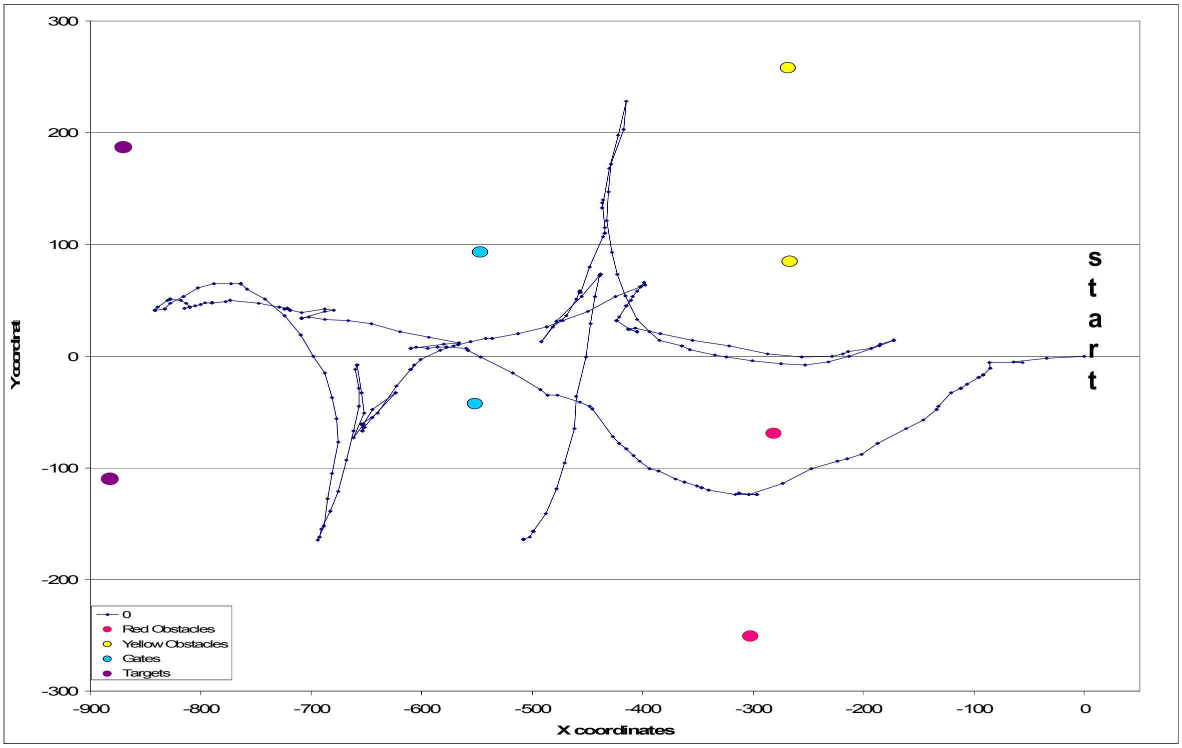

Figure 18: Path during an incomplete trial in the yoked camera condition. This subject reported that they had forgotten which target they were supposed to contact. The subject realized the mistake and then tried to correct it. Looking at the path as drawn it appears that the subject was not aware of position in the trial space and ended up traveling backwards resulting in an incomplete trial.

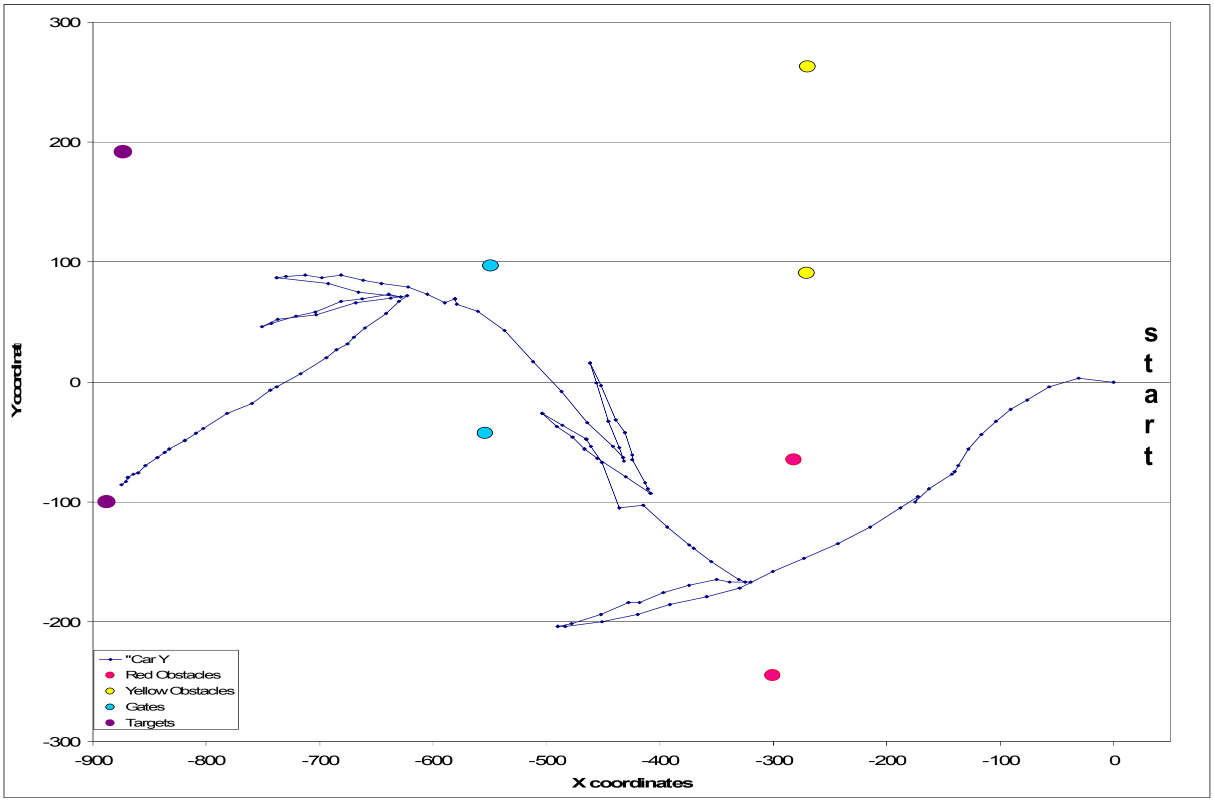

Figure 19: The shown path is typical for observers in the fixed camera condition. Observers in this condition would travel forward ensuring enough room to back up without running into the gate markers and then back and turn to line up their path to the next goal.

It appears that people resort to using mental representations of space when the intended goal or target is not visible. Based on observations from the present experiments, people seem to look for reference points that are consistent with a mental representation based on relative position (i.e. in front, behind, to the left or to the right) of the environment in order to reorient. This could be construed as an attempt to build a mental representation; however, the representation is not a representation of seen space. Rather, it appears to be a representation of things not seen. Such a representation would be useful in recovering from errors. However, as long as the navigational goal can be viewed directly, a mental representation is not necessary for action.

GENERAL DISCUSSION

The research undertaken in this project was designed to begin exploring the strategies human operators use while navigating in a remote environment. The goal was to explore the performance during the navigation task and look for evidence of effects from the two perceptual judgment tasks. It seems as if a key factor for all of the experiments is panning or rotation. The exact requirement for panning is not clear, but reflection on the results obtained here leads one to conclude that further investigation of this specific element of navigation could be fruitful. In other research (e.g. Murphy & Burke, 2005) it was shown that building mental models is difficult in real world remote navigation tasks. There is some evidence that this was true for our experiments as well and that this may be linked to panning and rotation. In order to understand this I will first briefly review the experimental results.

Experiment 1

One of the most interesting results from all of the experiments was the enhancement of perceived depth shown in the camera-panning condition of the spatial arrangement experiment. In this experiment, the task was to describe the spatial arrangement of the objects relative to the camera position. The camera motion in this experiment was meant to simulate the optic flow that might be created as an observer moves through a scene. As previously discussed, the greater depth perception under panning conditions is most likely attributable to motion parallax. Because the parallax was created by camera motion rather than body motion such as a head turn or eye rotation, the increased depth perception provided by simply panning through the scene cannot be due to efference copy. Consistent with the ecological account of perception and action coupling this suggests that the panning motion itself carries the depth information. The improvement was greatest when not combined with forward motion. Thus, use of the panning mechanism in tele-operation might best be used in environments that allow the operator to stop and scan the environment.

Experiment 2

In the path perception experiment, observers were not asked about visually perceived space, but about the future position of the camera. In this experiment, it was shown that observers had difficulty separating visual rotation from forward motion. In fact, they never perceived the stimuli in the simulated rotation condition to pass to the left of the target despite the fact that in 40% of the trials the camera would pass to the left of the target by as much as 10 deg. Results such as these indicate that some sort of compensation such as kinesthetic cueing (Ellis, Adelstein, & Welch, 2002) may be useful when working in environments that result in misalignment due to visual rotations.

Experiment 3

The remote navigation experiment provided a rich data set to study how two different view control mechanisms (yoked- versus fixed-camera) affect navigation performance. The yoked-camera provided support for a target direction (point and go) navigation strategy while the fixed-camera provided support only for an optic flow based strategy. Observers in both camera conditions were able to complete the course in the allotted time with no difference in the number of failed trials. However, observers in the yoked-camera condition were able to complete the course faster and with higher overall scores than observers in the fixed-camera condition. These results confirm the usefulness of having a panning capability that can be controlled independently of forward motion as was suggested after the first experiment.

Panning and Building Mental Models

It should be noted that if an observer loses sight of the target, recovery becomes quite difficult. Burke, Murphy, Coovert, and Riddle (2004), reported that the main human-robot interaction problem was not remote navigation per se, but rather understanding the situation the robot had encountered. This seems to be precisely the right description for the recovery problem observed in our experiment. Anecdotal evidence in the present research suggested that neither of the steering strategies investigated provided optimal support for developing a mental representation of allocentric position which seemed to be necessary for recovery from error. A control interface which supports independent control of the camera view may be a possible solution to the soda straw effect.

The first experiment showed that panning was useful in overcoming depth compression in a passively viewed scene. The navigation experiment showed that the having the camera view rotate with steering and not dependent on forward motion produced more efficient performance in a remote navigation task. However, the independent rotation in the path perception experiment created ambiguities regarding the direction of travel. These results suggest that while panning is useful for understanding space it must be incorporated in a way that does not interfere with judging self motion.

Summary and Future Research

One of the original goals of this research was to take some very traditional approaches to experimentation and determine if their results had any relevance to a more realistic task. The first two of these three experiments were meant to look at pieces of a remote navigation task (i.e. viewing space remotely through a medium and judging spatial layout and motion path) in order to identify effects that might emerge during a navigation task and affect the successful completion of the task. Depth compression and difficulty projecting the motion path into the future under conditions of rotation emerged as likely consequences of working in a remote environment. The remote navigation experiment was primarily designed to explore the use of optic flow for navigation and secondarily, as a platform to provide a practical assessment of all results in the context of a real world requirement.

Reviewing all results revealed some evidence that depth compression can be linked to performance problems in the navigation task. Specifically, the characteristic backward maneuvering of observers in the fixed camera condition appears to have been an attempt to overcome depth compression in the absence of the ability to pan through the scene independent of forward motion. Visual rotation appears to be more problematic for planning than for immediate action. It is possible that difficulties in recovering rate of rotation information interrupted the ability to build mental models which could be used to support recovery from error. Observers in both conditions appeared to have problems with such recovery; however, the navigation experiment was designed in such a way that only successful trial information was collected.

In the Introduction, the two systems model of perception and action (Milner & Goodale, 1995) was described as a possible way to account for differences between perceptual judgments and motor behavior. Relevant to human factors, DeLucia (2008) provides a practical framework for applying the two visual systems approach to control and display interfaces. The framework is predicated upon the notions of task, distance and time. The further something is in time and space the greater its dependency on the ventral system. The closer the task is in time and space the more it draws on the dorsal system. This could be beneficial for evaluating the results and recommendations reported here, but it requires further investigation. For instance, there is evidence in the navigation experiment that a different interface may be required for recovery from error which would incorporate the notions of planning and problem solving utilizing resources of the ventral stream. Such a framework might also explain the influence of the rotation in projecting future position.

In terms of applying the two systems model to the present experiments, it seems that the experimental data do not show consistent agreement. Specifically, the spatial arrangement experiment elicited a perceptual judgment showing that visual motion carries unique information that can facilitate more accurate spatial layout perception. The remote navigation experiment required an action response and showed that optic flow alone was not sufficient for efficient navigation. These results are the exact opposite of that predicted by the two systems account. Taken together, however, the present results are in agreement with a systems view, like that of DeLucia (2008) that is based on the time/space dimensionality of a task and this should be further explored.

Decoupling the perceiver from the environment, as is the case of remotely piloted vehicles, clearly affects human performance. The design of control interfaces for remotely piloted vehicles needs to be specific for the task being supported. When working in a close and relatively stable environment, having the ability to independently pan through a scene may provide support for better depth perception. When working in larger and faster paced environments a direct connection between the steering mechanism and the viewing medium appears to support more efficient navigation behavior. Finally, based on the anecdotal evidence obtained in these studies, when working in larger environments observers appeared to use mental representations to aid recovery when the target was lost or not in view.

Though not directly tested in this set of experiments, observers would probably benefit from having the ability to directly and independently control point of view in many situations. To the extent that panning was done independent of forward motion, it provided a greater depth perspective to observers in the spatial arrangement experiment. When panning was coupled with forward motion in Experiment 2, observers were unable to separate the two signals in order to determine heading direction. Taken together these results suggest that independent and flexible operator control of point of view would better serve precise control in close work and in recovery from disorientation during navigation. Mental representations seem to be more important in terms of relative position such as in front, behind, to the left, or to the right rather than in terms of absolute location.

The purpose of this study was to explore some of the factors relevant to navigating remotely piloted vehicles. Future research should focus on further identifying task variables that best predict the type of behavior required for successful performance. Such research could benefit from employing a framework such as that posited by DeLucia (2008) based on time, space, and task. Further, recovery from error provides a potentially rich and relevant area for exploration in perception and action coupling so future research should employ methodologies that are designed to capture this information specifically. Understanding how the errors occur and what people believe to be true about their own position within the environment requires a data collection method that is designed specifically for such a purpose. Murphy and Burke (2005) showed that the teams of two operators were more efficient in teleoperated search and rescue tasks than was a single operator. The natural interaction of such a team may also be helpful in data collection for future research.

REFERENCES

Adelson, E.H. & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, 2*(2), 284-299.

Adelson, E.H. & Movshon, J.A. (1982). Phenomenal coherence of moving visual patterns. In S. Yantis, (Ed.), *Visual perception: Essential readings. Key readings in cognition.*

Aglioti S., DeSouza J.F.X., & Goodale M.A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, 5(6), 679-685.

Banks, M.S., Ehrlich, S.M., Backus, B.T., & Crowell, J.A. (1996). Estimating heading during real and simulated eye movements. *Vision Research, 36,* 431–443.

Barlow, H.B. (1990). A theory about the functional role and synaptic mechanism of visual after-effects. In *Vision: Coding and Efficiency*, C B Blakemore (Ed). Cambridge University Press

Beardsley, S.A. & Vaina, L.M. (1998). Computational modeling of optic flow selectivity in MSTd neurons, *Network: Computational Neural Systems*, 9, 467-493

Bertin R.J.V., Israël, I., Lappe, M. (2000). Perception of two-dimensional, simulated ego-motion trajectories from optic flow. *Vision Research* 40: 2951-2971

Burke, J.L., Murphy, R.R., Coovert, M.D., and Riddle, D.L. (2004). Moonlight in Miami- a field study of human robot interaction in the context of an urban search and rescue disaster response training exercise. *Human Computer Interaction*, *19*, 85-116.

Casper, J. and Murphy, R. R. (2003). Human-Robot Interaction during the RobotAssisted Urban Search and Rescue Response at the World Trade Center. *IEEE SMC B*, 33, 367-385.

Cornilleau-Perez, V. & Gielen, C.C.A.M.(1996). Interactions between self-motion and depth perception in the processing of optic flow. *Trends in Neuroscience* 19(5), 196-202

Crowell, J., Banks,M., Shenoy, K.V., & Andersen, R.A. (1998). Visual self-motion perception during head turns. *Nature Neuroscience* **1**, 732 - 737 doi:10.1038/3732

Cutting, J.E & Vishton, P.M. (1995). Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.) *Perception of* *space and motion* (pp. 69-117). San Diego, CA: Academic Press.

Duffy C.J. & Wurtz R.H. (1993). An illusory transformation of optic flow fields. *Vision research* 33(11):1481-90.

Duffy, C.J. & Wurtz, R.H. (1997a). Medial superior temporal area neurons respond to speed patterns in optic flow. *Journal of Neuroscience*, 17(8), 2839–2851.

Duffy, C.J. & Wurtz, R.H. (1997b). Planar directional contributions to optic flow responses in MST neurons. *Journal of Neurophysiology*, 77, 782–796.

Durgin, F.H., Gigone, K. & Scott, R. (2005). Perception of Visual Speed While Moving. *Journal of Experimental Psychology: Human Perception and Performance,* (31,2) 339–353. DOI: 10.1037/0096-1523.31.2.339

Dunker, K. (1929). Über induzierte bewegung. *Psychologische Forschung, 12,* 180–259.

Ehrlich, S.M., Beck, D.M., Crowell, J.A., Freeman, T.C.A., & Banks, M.S. (1998). Depth information and perceived self-motion during simulated gaze rotations. *Vision Research*, 38, 3129-3145

Ellis, S.R., Adelstein, B.D., & Welch, R.B. (2002). Kinesthetic Compensation for Misalignment of Teleoperator Controls Through Cross-Modal transfer of

Movement Coordinates, *Proceedings of the Human Factors Society*, October, 2002, Baltimore MD

Finney, D.J. (1947), *Probit Analysis*. (1st edition) Cambridge University Press, Cambridge, UK.

Gibson, J.J. (1950). *The Perception of the Visual World* Boston: Houghton-Mifflin.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton-Mifflin.

Gillam, B. & Chambers, D. (1985) Size and position are incongruous: Measurements on the Müller-Lyer figure. Perception & Psychophysics, 37:549-556.

Goodale, M.A. (1998). Visuomotor control: Where does vision end and action begin? *Current Biology, 8*, R489-R491.

Harris, M.G. & Carre, G. (2001). Is optic flow used to guide walking while wearing a displacing prism? *Perception*, 30, 811-818.

Hiris, E. & Blake, R. (1996). Direction repulsion in motion transparency. *Visual Neuroscience, 13*, 187-197.

Hubel, D.H. (1989), *Eye, Brain and Vision*, New York: Scientific American Library

Ivanenko Y.P., Grasso R., Israël I., & Berthoz A. (1997). The contribution of otoliths and semiscircular canals to the perception of two-dimensional passive whole body motion in humans. *The Journal of Physiology* 502(1):223-33.

Johansson, G. (1970) On theories for visual space perception. *Scandinavian Journal of Psychology* 11(1), 67-74 DOI: 10.1111/j.1467-9450.1970.tb00719.x

Kelly, J.W., Loomis, J.M., & Beall, A.C. (2005). The importance of perceived relative motion in the control of posture. *Experimental Brain Research,* 161, 285-292.

Lappe, M., Bremmer, F. & van den Berg, A.V. (1999a). Perception of self-motion from visual flow. *Trends in Cognitive Sciences*, 3:329-336.

Lappe, M., Bremmer, F. & van den Berg, A.V. (1999b). Going against the flow: reply to Harris and Rogers. *Trends in Cognitive Sciences*, 3:449-450.

Lee, D. N. & Lishman, R. (1977). Visual control of locomotion. *Scandinavian Journal of Psychology*, 18, 224-230. retrieved June 2008 http://www.perception-in-action.ed.ac.uk/PDF s/Lee%201977%20Vision%20and%20Balance.pdf

Li, L., Sweet, B. T., & Stone, L. S. (2006). Humans can perceive heading without visual path information. *Journal of Vision*, *6*(9):2, 874-881, http://journalofvision.org/6/9/2/, doi:10.1167/6.9.2.

Li, L. & Warren, W.H. (2004) Path perception during rotation: Influence of instructions, depth range, and dot density. *Vision Research*, 44, 1879-1889

Li, L. & Warren, W.H. (2000). Perception of heading during rotation: Sufficiency of dense motion parallax and reference objects. *Vision Research, 40,* 3873–3894

Loomis, J.M. & Beall, A.C. (1998). Visually-controlled locomotion: Its dependence on optic flow, 3-D space perception, and cognition. *Ecological Psychology*, 10, 271-285.

Loomis, J.M., Beall, A.C., Macuga, K.L., Kelly, J. W. & Smith, R. S. (2006). Visual control of action without retinal optic flow. *Psychological Science*, 17, 214-221.

Manning, S.D., Rash, C.E., LeDuc, P.A., Noback, R.K., & McKeon, J. (2004). The Role of Human Causal Factors in US Army Unmanned Aerial Vehicle Accidents. U.S. Army Aeromedical Research Lab report #2004-11. retrieved January 2008 http://www.usaarl.army.mil/TechReports/2004-11.PDF

Mather, G. (1998). Introduction to Motion Perception. Retrieved January 2009 www.biols.susx.ac.uk/home/George_Mather/Motion

Mittelstaedt, H. (1999). The role of the otoliths in perception of the vertical and in path integration. *Annals of the New York Academy of Sciences*, 871, 334-344.

Milner, A.D. & Goodale, M.A. (1995) *The visual brain in action.* Oxford: Oxford University Press

Murphy, R.R. (2004). Human-Robot Interaction in Rescue Robotics. *IEEE SMC Part C*, 34(2), 138–153.

Murphy, R.R & Burke, J.L. (2005). Up from the rubble: Lessons learned about HRI from search and rescue. In *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, Orlando, FL, Sep. 26-30 2005. Human Factors and Ergonomics Society.

Muth, E.R., Walker, A.D., & Fiorello, M. (2006). Effects of Uncoupled Motion on Performance. *Human Factors* (48, 3), pp. 600-607(8) DOI: 10.1518/001872006778606750

Osborn, K.(2007). Army records first UAV kills. *Air Force Times* retrieved from http://www.airforcetimes.com/news/2007/09/army_UAV_kills_070915w/ on 23 February 2008.

Pack, C. & Mingolla, E. (1998). Global induced motion and visual stability in an optic flow illusion. *Vision Research, 38*(20), 3083-3093.

Ramachandran, V., & Cavanaugh, P. (1987). Motion capture anisotropy. *Vision Research, 27*, 97-106.

Royden, C.S., Banks, M.S., & Crowell, J.A. (1992) The perception of heading during eye movements. *Nature* 360: 583-585.

Royden, C.S. & Conti, D.M. (2003) A Model Using MT-like Motion-opponent Operators Explains an Illusory Transformation in the Optic Flow Field. Vision Research, 43: 2811 - 2826

Royden, C.S., Crowell, J.A., & Banks, M.S. (1994) Estimating heading during eye movements. *Vision Research*, 34, 3197-3214

Royden, C.S. & Hildreth, E.C. (1999) Differential effects of shared attention on perception of heading and 3D object motion. *Perception & Psychophysics*, 61: 120-133.

Rushton, S.K., Harris, J. M., Lloyd, M. R., & Wann, J.P. (1998) Guidance of Locomotion on Foot Uses Perceived Target Location Rather than Optic Flow. *Current Biology*, (22, 8) pp. 1191–1194.

Schneider G.E. (1969) Two visual systems: brain mechanisms for localization and discrimination are dissociated by tectal and cortical lesions. *Science* 163, 895—902. DOI: 10.1126/science.163.3870.895

Stone, L.S. & Perrone, J.A. (1997). Human heading perception during visually simulated curvilinear motion. *Vision Research*, 37, 573-590.

Tvaryanas A.P., Thompson W.T., Constable S.H. (2005). Human factors in remotely piloted aircraft operations: HFACS analysis of 221 mishaps over 10 years. *Aviation, Space, and Environmental Medicine*; (77, 7):724–32.

Ungerleider, L.G. & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale & R.J.W. Mansfield (Eds). *Analysis of Visual Behavior.* Cambridge, MA: MIT Press. pp.549-586

van den Berg, A.V. & Brenner, E. (1994) Why two eyes are better than one for judgements of heading. *Nature 371, 700-702.*

Voshell, M.G., Woods, D.D. & Phillips, F. (2005). Human-Robot Interaction: From Fieldwork to Simulation to Design. Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting. 26--28 September, Orlando FL.

Wallach, H., (1982). Eye movement and motion perception. In A.H. Wertheim, W.A. Wagenaar and H.W. Leibowitz (eds.), *Tutorials on Motion Perception*. Plenum, NY. pp 1-19.

Warren, P.A. & Rushton, S.K. (2007). Perception of object trajectory: Parsing retinal motion into self and object movement components. *Journal of Vision*, 7(11):2, 1-11, http://journalofvision.org/7/11/2/, doi:10.1167/7.11.2.

Warren, W.H., Jr., & Hannon, D.J. (1988). Direction of self-motion is perceived from optical flow. *Nature, 336,* 162–163.

Warren, W.H., Jr., Kay, B.A., Zosh, W.D., Duchon, A.P. & Sahuc S. (2001). Optic flow is used to control human walking. *Nature Neuroscience, 4,* 213–216.

Warren, W.H., Jr., Morris, M.W., & Kalish, M. (1988). Perception of translation heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 646–660.

Watamaniuk , S.N.J. (2005). The predictive power of trajectory motion . Vision research 45(24):2993-3003 doi:10.1016/j.visres.2005.07.027

Wertheim, A.H. (1998). Working in a moving environment, *Ergonomics* (41) 1845–1858.

Wiener, E.L. (1988) Cockpit Automation. In E. L. Wiener & D. C. Nagel (Eds). *Human factors in Aviation*, Pp. 433-461. San Diego, Academic Press.

Williams, K.W. (2004). A summary of unmanned aircraft accident/incident data: Human factors implications. Washington, DC: US Department of Transformation, Federal Aviation Administration, Office of Aerospace Medicine. Technical report Publication No. DOT/FAA/AM-04/24

Williams, K.W. (2006). Human factors implications of unmanned aircraft accidents: Flight control problems. Washington, DC: US Department of Transformation, Federal Aviation Administration, Office of Aerospace Medicine. Technical report Publication No. DOT/FAA/AM-06/8.

APPENDIX A

Experiment 1 Instructions

This experiment has two parts: stills images and moving images. All subjects will view both parts of the experiment. Half will view the still images first and half will view the moving images first. Subjects will be asked to mark the response grid with the letters R (red), B (blue), and Y (yellow) representing the perceived locations of the targets in the image. The grid is a representation of the space they are viewing. The experimenter should mark on the practice grid the subject number. The subject number should follow the convention: 1.1.a where the first number represents the experiment, the second number represents the subject (1-16) and the letter represents the order of presentation (a=stills first; b=moving first). The instruction script should be followed for each subject. Be sure to do the consent form first and to have a sample marked grid in front of the subject while you are going over the instructions. Marks should be made in the open spaces of the grid and not on the grid lines.

Script:

You will see a series of images with three targets in them. The targets are cylinders painted red, yellow, and blue. There are two types of trials: one in which the image will be a movie, the other will be a static picture. There are thirty trials of each type and you will see each image for one second.

(Point to the response grid) This is the response sheet and represents the area that you will see in the images.

(Point to the black circle) This represents the initial viewing position.

(Point to the response grid) After viewing each image you will be asked to mark on this grid where you believe the targets to be.

So you will put an R on a grid space to indicate the position of the red target, a B for the blue target and a Y for the yellow target. Marks should be put in the open spaces of the grid and not on the grid lines. You will use one response sheet per trial and the number on the top left of the sheet corresponds to the trial numbers. You will see each image only once and you cannot go back and review the image. To move forward to the next image, press the right arrow or enter key. You will do a practice run to make sure you understand the instructions.

** Make sure the pointer arrow is off to the side of the screen.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1																						1
2																						2
3																						3
4																						4
5																						5
6																						6
7																						7
8																						8
9																						9
10																						10
11																						11
12																						12
13																						13
14																						14
15																						15
16																						16
17																						17
18																						18
19																						19
20																						20
21																						21
22																						22
23																						23
24																						24
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	

APPENDIX B

Successfully navigating first gate	+2
Successfully navigating second gate	+3
Contacting final target	+2
Knocking over a cylinder other than final target	-1
Finishing under 1:30	+3
Finishing between 1:31 and 2:20	+2
Finishing between 2:21 and 3:00	+1
Finishing between 3:01 and 4:00	0
Missing a gate	-2
Incomplete trial	-4
BONUS: Contact final target w/o knocking over	+3

yellow plain

1	First Gate	Second Gate	Contact Target	Under 1:30	1:31 to 2:20	2:21 to 3:00	3:01 to 4:00	Bonus	Total
+	✓	✓	–	1:05				–	7
-	✓								

−4
+4

think aloud red striped

2	First Gate	Second Gate	Contact Target	Under 1:30	1:31 to 2:20	2:21 to 3:00	3:01 to 4:00	Bonus	Total
+	✓	✓	–	·	1:46			–	7
-									

Sub. states he didn't know he was supposed to actually contact final target

think aloud red plain

3	First Gate	Second Gate	Contact Target	Under 1:30	1:31 to 2:20	2:21 to 3:00	3:01 to 4:00	Bonus	Total
+	✓	✓	✓	44				✓	13
-									

Tests of Between-Subjects Effects

Measure: tortuosity
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power(a)
Intercept	387.377	1	387.377	348.723	.000	348.723	1.000
condition	12.894	1	12.894	11.608	.027	11.608	.723
Error	4.443	4	1.111				

a Computed using alpha = .05

condition

Measure: tortuosity

condition	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
yoked	1.199	.111	.891	1.508
fixed	1.735	.111	1.426	2.043

This analysis was conducted in SPSS which uses one degree of freedom for the intercept assuming that the line does not pass through 0,0.

$F(1, 4) = 11.608$, $p = .027$

APPENDIX C

The three targets used in the spatial arrangement experiment formed a natural triangle. In early attempts to characterize the responses to the stimuli, it seemed fitting to use some form of description of triangles. Ways of describing triangles include describing the sides (equilateral, isosceles, and scalene), the internal angles (right, acute, and obtuse), or the centers. All of the triangles formed by our stimuli could be described as scalene and obtuse leaving the center of the triangle as the most likely candidate for description.

Three different methods were initially chosen: orthocenter, incenter, and centroid. Each of these methods is a measure of centrality.

1. Orthocenter: An altitude of a triangle is a straight line through a vertex and perpendicular to (i.e. forming a right angle with) the opposite side. This opposite side is called the *base* of the altitude, and the point where the altitude intersects the base (or its extension) is called the *foot* of the altitude. The length of the altitude is the distance between the

base and the vertex. The three altitudes intersect in a single point, called the orthocenter of the triangle. The orthocenter lies inside the triangle if and only if the triangle is acute.

2. Incenter: An angle bisector of a triangle is a straight line through a vertex which cuts the corresponding angle in half. The three angle bisectors intersect in a single point, the incenter.
3. Centroid: A median of a triangle is a straight line through a vertex and the midpoint of the opposite side, and divides the triangle into two equal areas. The three medians intersect in a single point, the triangle's centroid.

Each of these was tried on a small sampling of the available data and each was rejected. The primary reason for rejecting these methods was that none of them was computed in the context of the coordinates of our response grid and required conversions to be useful. Beyond that, they each yielded a slightly different result regarding the location of the center. Eliminating the orthocenter seemed like a reasonable choice because it usually fell outside of the triangle. The incenter and centroid yielded slightly different results from each other, but there was no clear reason for choosing one over the other. In an attempt to evaluate these two options they were compared to average X (lateral) and Z (depth) coordinates which were much easier to calculate and did not appear to result in significantly different results (though this was not tested statistically). Using the averages also did away with the conversions that had to be done with the other methods which lessened the likelihood of mathematical errors. The final decision by the researcher was to use the X and Z coordinates.