Within Reach: The Contribution of Dynamic Viewpoint to the Perception of Remote Environments

DISSertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Taylor Murphy, M.S.

Graduate Program in Industrial and Systems Engineering

The Ohio State University

2017

Dissertation Committee:

Michael Rayo, Advisor

David D. Woods

Alexander Morison
Abstract

Remote sensor platforms, operating as part of a human sensor system, allow practitioners to extend their reach into remote environments normally inaccessible to humans and to substantially change the scale at which they work. Despite the utility these sensor platforms provide to domain practitioners, their operation remains difficult, slow, and error prone. Previous work has claimed these problems stem from the fact that currently available sensor platforms are not designed to work with the human operator as part of a larger perceptual system. As a result, operators struggle to understand the physical layout of the remote environment, and the opportunities for action in that environment. These claims are well founded from perceptual psychology, but have not yet received empirical verification. Additionally, research performed on human sensor system perception has not yet addressed the effects of viewpoint motion on perceptual performance. Both gaps in the human sensor system perception literature have been addressed in the current work. The three experiments performed to address these gaps extended previous research examining operators’ ability to judge the reachability of target objects in a remote environment. The first two experiments found that the human sensor system was well modeled using approaches from perceptual psychology. These results support the claims made in previous work, that sensor platforms operate as part of a larger perceptual system. The third experiment in the current work found that viewpoint motions that
provides new perspectives onto a scene of interest significantly improved participants’ perceptual performance. These results provide the first empirical verification of Roesler (2005) and Morison (2010)’s Perspective Control model. Taken together the results from the current work have implications for the design and testing of future sensor platforms in order to overcome the challenges facing current generation platforms and improve their performance.
Acknowledgments

I would like to thank the members of my committee. Without the help of Alex Morison, Mike Rayo, and Dave Woods this dissertation would not have been possible. Thank you for years of teaching, support, and encouragement. I wouldn’t be where I am today without all of you. I’d also like to thank my lab mates, Kati Walker and Asher Balkin. Thank guys for the sympatric ear and all the help you’ve given me along the way. Thank you to all my friends and family who not only supported me through my PhD, but also were willing to lend their time to help make this a reality.
Vita

May 2005 .................................................. Upper Arlington High School

2009 ............................................................... B.A. Psychology, Miami University

2012-2013 .................................................. Research Fellow, Consortium Research Fellows Program

2013 ............................................................... M.S. Industrial and Systems Engineering,

The Ohio State University

Publications


Fields of Study

Major Field: Industrial and Systems Engineering

Specialization in Cognitive Systems Engineering

Minor Field: Visual Analytics

Minor Field: Research and Simulation Techniques
# Table of Contents

Abstract................................................................................................................................................. ii  

Acknowledgments..................................................................................................................................... iv  

Table of Contents.................................................................................................................................. vii  

List of Figures......................................................................................................................................... viii  

Chapter 1: Introduction............................................................................................................................ 1  

Chapter 2: Human-Sensor Systems Considered as Perceptual Systems.............................................. 33  

Chapter 3: Viewpoint Motion and Perception of Remote Environments............................................. 59  

Chapter 4: Discussion.............................................................................................................................. 93  

References............................................................................................................................................. 114  

Appendix A: Exit Questionnaire........................................................................................................... 120  

Appendix B: Informed Consent Form .................................................................................................... 124
List of Figures

Figure 1 - Various UUVs and their applications. On the left side of the panel a UUV uses a torque device on a subsea structure. The structure, the yellow object on the left side of the image, is part of an underwater oil and gas field. The upper right image is a UUV using a manipulator arm to grasp a hose. The lower right image is an underwater inspection tool for imaging underwater structures.................. 2

Figure 2 - Urban Search and Rescue (USAR), and Explosive Ordinance Disposal (EOD). The three images on the left side of the figure show inside and outside views of collapsed structures and rescue personnel. The top and bottom images on the right side are of the Talon (top) and PackBot (bottom) EOD platforms.......... 4

Figure 3 – Fixed with and rotary wing UAVs. Beginning with the upper left image and moving clockwise, a miniature ‘quadcopter’ UAV. The Aeryon Scout micro UAV. The X-47B. The USAF Firescout helicopter. The Global Hawk surveillance platform. The MQ-1 Predator ......................................................... 6

Figure 4 - Operator's mediated view of the area of interest. To overcome the difficulty of an often ambiguous view, operators invent workarounds to simplify their tasks. .............................................................................................................................. 9

Figure 5 – The participant’s view from the primary camera onto the sparse, or low, condition environment. This environment was designed to give very little information about the 3D layout of the remote environment. The condition is roughly analogous to operating a platform underwater................................. 40

Figure 6 – The participant’s view onto the medium condition environment. The medium environment adds the ground plane as a visual landmark. This environment was designed to be roughly analogous to operating in a deconstructed environment. These environments have ambiguously scaled landmarks and few consistent shadows. ........................................................................................................ 41

Figure 7 – The participant’s view on to the high condition environment. This environment added shadow information and texture to the target object. The environment provides similar visual information to operating a platform in direct sunlight. The target’s texture and the shadow on the ground plane should provide a
sense of scale to the environment, while the shadow itself is a source of information. ................................................................. 43

Figure 8 – The remote sensor platform used in the virtual environment is based off of the Talon platform (left panel). The virtual platform has four servo motors to move the manipulator arm. The servos are highlighted the left panel and labeled A, B, C, D. The left panel also shows two different camera positions. On the right side of the panel is the primary, or stalk camera. On the left side of the panel is the camera mounted on the manipulator arm used in Chapter 3. Both cameras correspond to their positions on the actual Talon platform. ...................... 45

Figure 9 – The structure of both experiment 1 and 2. Each condition is made up of three phases (Training, Practice, and Data Collection). Each phase has a number of trials. Both Training and Practice have eight trials. In experiment 1 the data collection phase had a variable number of trials, based on the participant’s performance. In experiment 2 the data collection phase had 252 trials total. Note that this diagram does not include the frequent breaks given to participants in both experiments. ................................................................. 48

Figure 10 – The participant’s view of the environment during the training phase. The box in the upper left corner is a graphical interface participants interacted with in order to move the manipulator arm. The goal of each trial was to move the arm to make contact with the ball. The arm is depicted here hitting the ground with the end effector, triggering a warning on the graphical interface. ...................... 49

Figure 11 – The series of UI states for each trial. The initial stimulus mask and its countdown timer, the camera feed with buttons to make a judgement, and the post-stimulus mask. The post-stimulus mask only appeared if participants did not make their judgements within three seconds. The software remained at the post-stimulus mask until participants made a judgement. ...................... 51

Figure 12 – The space of all possible end effector positions split into quadrants. A pool of 20 pre-made manipulator arm configurations (made up of sets of servo positions) was created to set the arm’s position in each trial. A member of the pool was randomly selected in each trial. In order to reasonably sample over the space of all possible end effector positions, the space was split into quadrants. The pool contained five configurations for each quadrant. ...................... 54

Figure 13 – Estimated psychometric functions for the worst fit, median fit, and best fit participants across three experimental conditions. ................................................................. 57

Figure 14 – Distribution of McFadden’s pseudo-R2 across experimental conditions. The green zone represents what McFadden (1978) refers to as “excellent fit.” .......... 58
Figure 15 – All perspectives simultaneously reveal and obscure information (Tittle et al., 2002). Shifting perspective reveals new information about the environment. The upper panel shows a piece of street art viewed from a specific perspective. This perspective creates the illusion of a 3D image interacting with the people around it. However even a small change in perspective breaks the illusion, revealing new information.

Figure 16 – The movement timeline for each trial in the ego-centric conditions. The timeline begins with the stimulus mask. At time 0 the stimulus is revealed. From 0 – 0.2 seconds, no movement occurs (labeled static). From 0.2-1.8 seconds the viewpoint pans to the left using the quadratic easing function highlighted in grey. The position of the camera is show at four different times. The quadratic easing function smoothed the movement from the starting orientation to the end.

Figure 17 – The movement timeline for the exo-centric movement type. The timeline is made up of the same segments as the ego-centric movement and the camera’s position is show at four different points in time. The viewpoint’s movement describes an arc centered on the target object at the top of each panel. The velocity of the viewpoint follows the same quadratic easing function shown in Figure 16.

Figure 18 – The timeline for the pseudo-arm viewpoint motion type. The timeline shares the same segments as the previous to movement types. This motion type fixed the camera to the manipulator arm. The arm ended its movement in the same position and at the same orientation as the camera in the exo-centric movement condition.

Figure 19 – The structure of experiment 3. Unlike experiments 1 and 2, the training phase was only administered at the beginning of the experiment session. Each of the eight condition were made up of a Practice and Data Collection phase. Each data collection phase was made of 150 trial. 20 of these trials were catch trials, designed to interfere with the participant recognizing any pattern in the placement of the target object. The other 130 trials contributed towards the psychometric function parameter estimation.

Figure 20 - Least squares means for α across all motion conditions. The horizontal black bar represents the manipulator arm’s maximum reach. * denotes significant differences at p < .05.

Figure 21 - Least squares means for β across all motion types (left panel), and visual environments (right panel). * denotes significant differences at p < .05.
Figure 22 – Least squares means for $\alpha$ across grouped motion conditions. The horizontal black bar represents the manipulator arm’s maximum reach. * denotes significant differences at $p < .05$. ................................................................. 88

Figure 23 - Least squares means for $\beta$ across grouped motion conditions. * denotes significant differences at $p < .05$. ................................................................. 89

Figure 24 – Distribution of reported previous experience with virtual environments (e.g. video games) and robotics. ............................................................................. 90

Figure 25 – Average reach performance for the pseudo-arm and static motion types. Four target objects are shown in the top panel and represent the physical reach threshold, the average perceived threshold for pseudo-arm, the average perceived threshold for static, and the maximum sampling boundary. The bottom four panels show each target object from the standard over-the-shoulder perspective used in all motion types except pseudo-arm. Note that the pinned observations described in the results section of chapter 4 indicated that participant’s perceived thresholds were likely beyond the max distance pictured above........................................................................................................ 102
Chapter 1: Introduction

The past two decades have seen a substantial increase in the use of remotely operated sensor platforms, and as a result understanding and acknowledging their shortcomings has never been more important by stakeholders across many different work domains. Often referred to as teleoperated robots, platforms in this class are usually mobile, feature many different possible sensor configurations, and are driven by human operators frequently located outside line of sight. Remote sensor platforms allow human stakeholders to operate in places normally inaccessible to humans and to shift the scale at which they work, often changing the nature of work (McGuirl, Sarter, & Woods, 2009). Sensor platforms offer tremendous utility to stakeholders in work domains such as explosive ordnance disposal (EOD), urban search and rescue (USAR), internal medicine, agriculture, intelligence gathering, law enforcement, underwater inspection, and underwater drilling. Unmanned aerial vehicles (UAV) alone are projected to become close to a $30 billion industry by 2027 (Abboud, 2017) (see Figure 3).

One of the primary benefits of these platforms is the new viewpoint that they provide operators into a remote environment, offering perspectives previously unavailable or too cost prohibitive. These new viewpoint into the world allows operators, and other stakeholders, to gather information to inform decision making (Casper & Murphy, 2003;
Messina & Jacoff, 2006). In some instances, sensor platforms also give operators the ability to manipulate the remote environment. The ability to manipulate the remote environment opens an even wider set of possible applications for platforms (Balkin et al., 2014). The adoption and capabilities of remote sensor platforms will only continue to grow.

Figure 1 - Various UUVs and their applications. On the left side of the panel a UUV uses a torque device on a subsea structure. The structure, the yellow object on the left side of the image, is part of an underwater oil and gas field. The upper right image is a UUV using a manipulator arm to grasp a hose. The lower right image is an underwater inspection tool for imaging underwater structures.

Despite the clear utility sensor platform provide, operating current generation platforms remains slow, difficult, and error prone (Morison, 2010; Morison, Woods, & Murphy,
Both observations of sensor platforms being operated by experts and research performed in laboratories show current platforms require steep learning curves, long periods to perform their tasks, and produce frequent performance problems during operation (Balkin et al., 2014; Casper & Murphy, 2003; Scholz et al., 2006). These challenges become even more apparent when operators must run platforms outside their own line of sight. In these situations, the operator’s knowledge of the remote environment is entirely mediated by the sensors on the platform. When the platform moves outside line of sight even simple actions like opening a door handle require long periods of time to execute and are highly prone to mistakes. Existing research frequently creates prescriptions to address these challenges, as well as attempts to create metrics or benchmarks with which to measure the performance of human sensor systems (Goodrich & Schultz, 2007). However, many of the candidate metrics focus task outcome (e.g. time on task, errors during task, tasks completed, etc.) and provide little diagnostic information to generate new design ideas. Additionally, of the commonly used performance metrics very few address commonly observed workload bottlenecks experienced when platforms work outside the operator’s line of sight (Morison et al., 2015).

Many of the recommendations from the HRI and HCI literature address either the sensor platform or the human operator, treating the two as isolated components. These recommendations can be loosely split into categories such as more rigorous training, increased sensor field of vision, high resolution sensors, manipulator arms with more degrees of freedom, or automating the platform to eliminate the need for a human
operator. While adopting these recommendations have undoubtedly led to task level performance improvements, many of the same challenges observed in 2001 continue to be as prevalent today as they were then (Morison et al., 2015). Even using today’s state of the art bomb disposal robot, operators continue to have difficulty executing everyday tasks (Guizzo, 2011). New approaches to overcome the challenges involved with Remote Sensor Systems are required.

Figure 2 - Urban Search and Rescue (USAR), and Explosive Ordinance Disposal (EOD). The three images on the left side of the figure show inside and outside views of collapsed structures and rescue personnel. The top and bottom images on the right side are of the Talon (top) and PackBot (bottom) EOD platforms.
One promising approach to improving remote sensor platforms is to change the locus of the research from considering the platform as a stand-alone entity, to exploring the interaction between human and sensor platform. The key to this shift in thinking is to consider the human operator and the remote sensor platform as highly interdependent parts of a single human-sensor system (Morison et al., 2015; Tittle, Woods, Feil, & Roesler, 2004). In this context, interdependence means that the human relies on the sensor platform to accomplish their goals, and the sensor platform is dependent on the human’s ability to understand the remote environment and supply appropriate input, be they direct manual input or more abstract commands to a partially autonomous agent (Klein, Woods, Feltovich, Hoffman, & Bradshaw, 2004). Reclassifying the remote sensor system to being part of a Human-Sensor system shifts the focus of research to the interactions between the components of the system and opens up many new avenues of research to improve task level performance in the field.

One example of this new focus on interaction between components of the larger system is to consider how the operator uses the platform as a tool to extend their own perception into the remote environment. Whether operating the platform manually or simply making decisions for a semi-autonomous platform, the operator relies on their own perceptual capabilities to understand the remote environment and provide appropriate input to the platform. However, extending the operator’s perception into a remote environment is not automatically accomplished simply by providing them a viewpoint into the remote environment. The ability to extend an operator’s perception must be explicitly designed
into the remote sensor platform. According to (Morison, 2010) a well-designed platform will allow the user to shift the viewpoint in order to take new and useful perspectives with little effort.

Figure 3 – Fixed with and rotary wing UAVs. Beginning with the upper left image and moving clockwise, a miniature ‘quadcopter’ UAV. The Aeryon Scout micro UAV. The X-47B. The USAF Fire scout helicopter. The Global Hawk surveillance platform. The MQ-1 Predator.

Leveraging human perception to improve computer interface design has a long history. Some of the first work combining these two dates back to work on utilizing the center-surround found in the human retina (Woods, 1984). This work directed interface designers to create computer interfaces with detailed information in the center, and less detailed contextual information around the periphery. The information on the periphery would act as a guide for users trying to navigate between many different information screens. By building a meaningful organization of the information on individual screens
and the structure of the information screens as set, this visual momentum would allow users to quickly and easily explore large amounts of data.

Leveraging the strengths of the human perceptual system was carried forward into work on remote sensor platforms. The work of Tittle et al. (2002) presented a strong argument for designing remote sensor platforms around the needs of operators’ perceptual systems in order to leverage strong processing capabilities humans display when navigating their immediate environments. These authors observed that the natural perception-action cycle humans use to understand and navigate their immediate surroundings could easily be broken if platforms were not designed to help compensate for decoupling the operator from their natural environment. The consequence of breaking the perception action cycle would shift the type of mental work operators perform from a quick and reliable direct perception task, to an effortful and error prone inference task. Forcing operators to make inferences about the layout of the environment, and the platform’s ability to take action in that environment, would greatly increase task time and create a much heavier cognitive workload for the operators. Current generation remote sensor platforms would appear to be suffering from the exact problems described by (Tittle et al., 2002).

Observations of real operators suggest that they struggle to understand physical layout of the remote environment and what actions they can accomplish in that environment with their platform (Guizzo, 2011; T. Murphy & Morison, 2016). When operators are working outside line of sight, they will often take time to develop workarounds to either better
understand the environment or to simplify their task to overcome the lack of understanding. While these workarounds are often successful, they are slow and highly error prone. The difficulty in understanding environmental layout, and creating workarounds, stands in stark contrast to people’s natural perceptual capabilities. When operating in their immediate environment people quickly make sense of their physical surroundings and are able to reliably judgements about what actions they are capable of taking. The disparity in performance between these two cases suggest something about having perception mediated by current generation platforms hobbles operator's’ ability to perceive the remote environment with the same ease they do their immediate environment. These observations would seem to corroborate the predictions of (Tittle et al., 2002), serving as examples of what happens when the design of remote sensor systems break the perception action cycle.

To date several studies have investigated operators’ perceptual capabilities when perception is mediated by remote sensor platforms. All of these studies share a common thread of measuring the perception of affordances in the remote environment as a way gauge the operators’ perceptual capabilities. The concept of an affordance, originally developed by Gibson (1979), describes a stimulus that informs an actor whether or not they can perform an action in their environment. An affordance is a property simultaneously of the environment and the actor. In the present case, the actor is the human-sensor system. The actor’s half of the affordance is measured by the capability of the remote sensor platform but the affordance itself is perceived by the human operator.
Several variations of human sensor system perception studies have been published to date, most notably on the affordances of passability, drivability and reachability. Perception of passability and drivability were compared for operating in line of sight versus operating outside line of sight. In a separate study, the effect of camera height on passability perception was studied. In addition to these studies, Murphy (2013) compared reachability perception across three different sensor configurations.

Figure 4 - Operator's mediated view of the area of interest. To overcome the difficulty of an often ambiguous view, operators invent workarounds to simplify their tasks.

The studies conducted on human-sensor system perception conducted to date have consistently implied that the dynamics of human sensor system perception are appropriately modeled using the traditional psychophysical metric. However previous research has shown that adding technological capabilities to perceptual tasks formerly performed by humans alone can drastically change the dynamics of the system (Sorkin &
Woods, 1985). To date there has been no empirical verification that the models and metrics designed to study human perception can successfully model human sensor system perception. This lack of verification constitutes a gap in the literature on human-sensor systems, and should be addressed if this vein of research is to continue on theoretically firm ground.

A second important gap in this nascent body of literature is the use of viewpoints that remain static relative to the chassis of the platform. The ability to move viewpoint in a smooth and meaningful way is a critical component of human perception (J. J. Gibson, Gibson, & Gibson, 1957), and thus to the human sensor system. Additionally, viewpoint movement is an important tool available on many remote sensor platforms today, that is utilized by platform operators. If one accepts the idea that the human operator’s perception is an important part operating a remote sensor platform, then the lack of research on the effects of viewpoint motion constitutes a critical gap in the HRI literature. Filling this gap is an important roadblock to designing more effective sensor platforms in the future.

**Specific Aims/Contributions**

This dissertation addresses the two literature gaps presented above. Chapter 2 describes the results from two experiments intended to verify that human sensor system perception can be described using the models and metrics of visual perception. Chapter 3 describes the first investigation of viewpoint motion in human-sensors systems. The study in
chapter 3 of this dissertation compared the effect of different viewpoint motion on human-sensor systems’ ability to perceive the remote environment surrounding the sensor-platform. This study fulfills two primary aims by studying the effects of viewpoint motion: providing experimental evidence about the impact of viewpoint motion on operators’ ability to directly perceive the environment, and discover whether the visual richness of the environment impacts direct perception under viewpoint movement. In order to address the literature gaps, all three experiments described in this dissertation measured participants’ perception of an object’s reachability using the remote sensor platform.

To determine participants’ perception of reachability they were presented with a virtual sensor platform, which included a multi-jointed manipulator arm, and a virtual target inside an environment. Participants were then asked to judge whether or not the target was within the manipulator arm’s reach. Participants’ judgement of reachability was probed repeatedly with the target appearing at various locations. The experimental apparatus systematically varied the target’s distance from the base of the robot by feeding the participants’ previous answers into one of several sampling algorithms (Leek, 2001).

Ecological Perception and Affordances

Throughout his work in ranging from the 1950’s to 1979, James Gibson advanced a new ontology of how perceptual systems understand the world around them. Gibson eventually gave this new way of thinking the name Ecological Perception. The term
ecological was used because in Gibson’s formulation the organism sensing the environment and the environment itself were part of an inseparable system. One of the most prolific examples of this mutuality is Gibson’s concept of an affordance. First introduced in his 1966 book The Senses Considered as Perceptual Systems, Gibson would continue to elaborate on this idea, culminating in his final work, The Ecological Approach to Ecological Perception (1979).

In the time since the publication of his final work the concept of the affordance has gain significant traction not only among perceptual psychologists, but also among designers (Norman, 2004) and technologists (R. R. Murphy, 1999). The term today has generated many different definitions both within the Ecological Perception community (Dotov, Nie, & De Wit, 2012) and fields adapting the term for their own applications.

Inspired in part by the Gestalt psychologists who came before him, Gibson’s approach radically departed mainstream models of perception which held that perception was the result of the brain running a series of processes on the information from a retinal image. Instead Gibson advanced the idea the information about an environment is directly perceived by an animal. Thus the information contained in the environment becomes an important object of study (J. J. Gibson & Gibson, 1961). This information that animals pick up are not cues that the perceptual system then processes to create perception, rather the information about the environment are the invariant properties of the environment.
These invariants are formless, timeless, and are observer neutral (Cutting & Cutting, 1982).

Gibson pushed the mutuality of the environment and the animal further, positing that the proper unit of analysis in perception is not an organism in isolation, as in most inferential models, but rather the organism inside it’s natural environment. This idea of system made up of animal and environment culminated in Gibson’s creation of affordances in his (1979) book.

> The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill… I mean by it something that refers to both the environment and the animal in a way that no existing term does.
> 
> – Gibson (1979)

This often-quoted line from Gibson is his most compact description of an affordance. An affordance is a characteristic of both the environment and an organism that determines what actions are available to the organism in that environment. Gibson posited that the aspects of an object are perceived are not its qualities, like its length in feet, but rather the action that object affords the animal (J. J. Gibson, 1979). These affordances are specified by the layout of surfaces of an object, or the environment. For instance, if a horizontal surface is large enough and rigid enough, relative to the animal, it may afford supporting the animal. Importantly the properties of size and rigidity are not perceived as concepts of the object/environment itself, but rather in terms of their relation to the animal. This concept was later illustrated by Warren (1984) in their study on stair climbing.

Participants in this study judged whether or not they could claim staircases of different heights. Regardless of participants’ height, the factor that determined whether stairs were
perceived as climbable was the ratio of the stair’s height to the leg length of the participant’s leg.

These affordances are another form of invariance that inform Gibson’s larger approach. He posits that affordances are invariant, and do not change with the needs of the observer (J. J. Gibson, 1979). This is one departure from the Gestalt from which affordances were derived. These earlier theories used a mailbox example, saying that the mailbox only had a “valence” when an observer needed to use it.

Gibson makes an important point that would later become the subject of debate in the Ecological Perception community.

> “An affordance, as I said, points two ways, to the environment and to the observer. So, does the information to specify an affordance… It says only that the information to specify the utilities of the environment is accompanied by information to specify the observer himself, his body, legs, hands, and mouth.”

- Gibson (1979)

Theorists continue to debate about what exactly counts as an affordance. Some camps claim that an affordance is a property of the environment (Turvey, 1992), others claim it is an emergent property of the environment and animal together (Stoffregen, 2003).

One of the more recent, well accepted schools of thought on affordances comes from the work of Anthony Chemero. Chemero challenges the commonly held view that affordances are a property of the environment. In critiquing the assertion that affordances are not perceived, only acted upon, Chemero (2001) draws a distinction between what he
calls predictional properties and placed features. Predictional properties of a specific object, whereas a placed feature is a property relative to a specific animal. Affordances, Chemero argues, are most often a type of placed feature being a property of the environment and the observer simultaneously. According to Chemero conclusions are based on the false assumption that all optical information is with respect to individual objects when in fact optical information can be individual objects or about the environment as a whole.

Both (Chemero, 2003) and (Stoffregen, 2003) suggest that affordances are aspects of an animal-environment systems, not properties of an environment. Chemero goes on to argue that affordances are relations. Using an example of perceiving the potability of a drink, Chemero posits that what is perceived is the relationship between the observer and the object. Thus, two observers can see the same drink without their perceptions overlapping, each observer maintains their private perceptions. The perception of this relationship is the basis of Chemero’s definition of affordance. He goes on to disagree once again with Shaw & Turvey by claiming that the animal part of the affordance relationship is not an effectivity, as posited by Shaw & Turvey, because effectivities are dispositions. Dispositions, given the right circumstances, are guaranteed to happen. Having the ability to walk under ideal conditions is no guarantee that you won’t trip.

A major dispute in the field of Ecological Perception has surfaced in the last seven years about the specificity of affordances. Chemero (2003) breaks with traditional Gibsonian
theory by proposing that affordances can constrain behavior rather than specify it. Relating affordances to the selection pressures of evolution, Chemero finds it unlikely that animals exclusively rely on specifying information. Specifically, that animals that rely on variables that relate ambiguously to environmental properties will not survive to reproduce. The key to this new view is that affordances as constraining variables can still guide animal behavior. This new understanding of affordances is challenged and extended by Withagen & van der Kamp (2010) in which they claim that Chemero’s work, while an important step, does not sufficiently explain what determines the object of perception. They argue that perceptual information is not specified in the ambient light array, but rather is a relation between that light array and perceptual processes. The authors further claim that their definition of perceptual information is very similar to the concept of an affordance, and is thus similar to Gibson’s assertion that what animals perceive are affordances.

In summary, Ecological Perception posits that affordances are one of the primary sources of information perceptual systems directly perceive about the environment around them. Using this framework, measuring one’s ability to directly perceive an environment can be logically accomplished by investigating one’s ability to perceive affordances. It is on this basis that the literature on the perceptual performance human-sensor systems has been built.
Mediated Perception and Human Sensor Systems

Some of the early work done on seeing the use of remote sensor systems as tools to extend an operator’s perception can be found in Tittle et al. (2002). These authors lay out the difficulties of trying to act in a remote environment through an intermediate agent, what they call the “remote perception situation.” The authors recognize that using remote sensor platforms to perceive an environment can break the natural link between perception. When this link is broken the operators’ ability to perceive the layout of surfaces, and consequently the available affordances, is severely curtailed. The authors conclude that the impoverished view provided by most platform designs can be minimized by augmenting the video available to the operator. The primary suggestion they present is to give the operators control over the viewpoint attached to the platform. This would allow the operator to move the viewpoint in order quickly and easily inspect a scene from multiple points of view. Woods et al. (2004) covers a variety of topics that they believe must be addressed to improve the coordination between human stakeholders and sensor platforms, including the consequences of breaking the perception action link.

It is important to note that these issues do not just apply to stakeholders manually piloting a platform. If a human stakeholder’s goals require them to understand the remote environment, they must be able to leverage their perceptual systems to gain that understanding. Thus, even as platforms gain autonomous capabilities in the future, reducing the need for manual operators, these principles will remain relevant.
Morison et al. (2015) begin to describe a design space for the sensors systems mounted on remote vehicles. They reiterate the loss of ability to move and explore a remote scene fluently when ability to directly perceive the scene is lost. As a result, to begin understanding how to design better systems the authors posit that all human sensor systems are also perceptual systems and that they support, whether intentionally or not, the remote problem holder’s ability to perceive features of the remote environment. As a result, remote sensor systems should be designed around to enable the human-machine system to act as a perceptual system. The authors propose a promising approach to designing a perceptual system: Perspective Control. This new approach to design emphasizes operators’ ability to quickly and easily take new perspectives within a spherical coordinate system centered at a point in the remote world. Importantly, this control method creates smooth transitions between viewpoints in order to take advantage of important perceptual information normally available to animals as they move and explore.

**Designing Remote Sensor Platforms to Extend Perception**

Many unmanned vehicles, or more generally remote sensor systems, in use today are designed from a technology first perspective, resulting in difficult to use platforms that require long periods to accomplish relatively simple tasks (Yanco, Scholtz, & Drury, 2004). One substantive barrier to improving the performance of these systems is improving operators’ ability to understand and act in the remote environment. Using Ecological Perception, we can apply what we know about how animals perceive the
world to the design of remote sensor systems. Designing systems with the operator’s needs at the forefront can better leverage the power of the operator’s perceptual system.

Adopting an ecological view of perception has several implications for the design of remote sensor systems. Firstly, taking an ecological perspective shifts the unit of consideration from the technological platform to the platform and the human stakeholder as a single system. The robotic platform becomes a stand in for the human operator. This shift in focus requires that the interaction between these two system components become one of the central concerns of the design processes. Voshell et al. (2005) call this designing to facilitate human-robot coordination.

Another implication of adopting an ecological perspective on perception is that remote sensors systems can break the natural perception-action link that animals rely on to collect information about the environment around them (Tittle et al., 2002). This break impairs the operators’ ability to quickly, and seemingly effortlessly, understand the layout of surfaces around them. As a result, operators have difficulty understanding what actions are available to them in the remote environment (Morison et al., 2015). The consequences of this difficultly can be seen in (Casper & Murphy, 2002) where operators struggle to understand the scale of the environment. Reconnecting the perception-action link becomes one of the key design challenges of remote-sensor systems. Tittle et al. (2002) call this creating functional presence, the ability for the operator to act as if they
were in the remote environment themselves. This is distinguished from the concept of presence, which is the feeling that one is actually present in a remote environment.

Approaching remote sensor system design from an ecological perspective also implies that the common design methodology of providing individual displays for each sensor is insufficient to create a sense of functional presence. Obliging operators to integrate multiple data streams in order to understand the layout of the environment around the remote sensor platform falls outside of perception and forces them to perform effortful mental work. One proposed solution to this integration is to give operators control of a single, dynamic viewpoint (Morison, 2010; Morison et al., 2015; Tittle et al., 2002). This approach is promising because it not only allows operators to take multiple viewpoints on an area of interest, but the smooth transition between those views provides additional visual information that is critical to perceiving surface layout and thus affordances (Roesler, 2005).

Adopting an ecological approach also has implications for the physical control devices that operators interact with. In Mantel et al. (Mantel, Hoppenot, & Colle, 2012)’s design process guidelines, equal emphasis must be placed on designing the platform-environment relationship and the operator-console relationship. This approach dovetails with Woods (1991)’s Symbol Mapping principle which describes design as having two primary mappings: those between the underlying domain and an interface’s representation, and those between the representation and the user’s perceptual system.
One instantiation of a physical control device proposed and developed by Morison (2010) based on Roesler (2005)’s work on perspective control. This physical control device provides a one-to-one mapping between the physical movement of the controller and the movements of sensors in the remote environment. This one-to-one mapping allows the user to take advantage of their own perceptual system to understand the orientation of the sensor in the remote environment.

The PackBot (iRobot Inc.) was introduced in 2002 as an explosive ordinance disposal (EOD) platform (see Figure 2). An updated version was release in 2007 and by 2010, 3000 platforms had been delivered to the US Army. In 2013, the army secured another contract that would supply it with these platforms indefinitely. More recently, PackBot was the first type of unmanned ground vehicle (UGV) to respond to the Fukushima disaster (“iRobot 510 PackBot”, 2016).

Generally, platforms like the PackBot are used by organizations to transport objects that are potentially explosives or disarm bombs. Fulfilling these high-level goals are tasks such as navigating cluttered terrain, searching complex environments, gripping objects, placing objects. These tasks can be further understood in terms of the types of information required to carry them out. For example, picking up an item might require an operator to understand if the object is graspable, carryable, and reachable.
The PackBot has little or no autonomous capabilities and is entirely controlled by a single operator. In order to pick up or manipulate a target object, the operator has control over a 5 degree of freedom arm and an array of cameras. The cameras, as with many EOD platforms, are mounted to the robotic arm. One camera is mounted on the ‘shoulder’ of the robotic arm, on the joint connecting the arm to the vehicle’s chassis. Panning this camera will move the whole robotic arm left and right, however the camera can tilt independently. Another camera is mounted to the ‘wrist’ of the robotic arm. This camera cannot be panned independently of the arm, and cannot be tilted without effecting the position of the arm’s grippers (i.e. end effector). The last camera is mounted to an additional arm segment beyond the end effector. This camera can be tilted independently relative to the end effector and panned independently.

PackBot is controlled via a ground control station (GCS). As shown in figure 1 each camera is shown to the operator as a direct feed. Camera feeds can be presented either simultaneously or individually. Newer PackBot models also provide a 3D representation of the platform to show its current configuration. Additional parts of the interface show the tilt of the chassis relative to gravity. The user controls the platform through either two joysticks mounted to the GCS, or through a gamepad controller (also featuring two joysticks) (“iRobot 510 PackBot”, 2016).

The PackBot, and other similar UGVs, have provided a tremendous amount of utility to operators in the field. However, they can be slow, difficult to use, and error prone
(Casper & Murphy, 2002; Morison et al., 2015; Yanco et al., 2004). This mixed success paints the picture of a system designed from a technology-first perspective that strain’s operator’s ability to adapt to challenging circumstances.

The PackBot does provide multiple viewpoints into the remote environment with is important to help operators accomplish their goals. However, from an ecological viewpoint providing multiple discreet, mostly stationary, sensors to switch between is far from desirable. Morison (2010) calls the inclusion of multiple cameras on a remote sensor system a workaround, to compensate for operators’ limited ability to perceive through sensors. A biological perceptual system has a single viewpoint that moves smoothly between different vantage points. This system samples many different parts of the environment, not just the immediate point of interest (Voshell et al., 2005). having to purposefully shift one’s attention between different discreet viewpoints is burdensome and requires a great deal of explicit cognitive work. An operator must explicitly reason about which viewpoint would be most advantageous, given their understanding of the remote environment’s layout, the state of the robotic platform, and their current set of goals.

A second concern for these systems is the coupling of the system’s ability to move vantage points and the system’s ability to manipulate objects in pursuit of its goals. Notice that all of the PackBot’s cameras are mounted on a part of the robotic arm. The result of this configuration is that in order to move the platform’s viewpoints to take a
new vantage point, the operator will have to move the manipulator arm, potentially interfering with their ability to manipulate the scene. The exception to the is the PackBot’s camera attached below the end effector, it is the only camera able to change positions without interfering with the end effector. However, this camera is limited to movements up and down.

Operators using the PackBot are limited to only their visual abilities. They have lost the use of any other sense like hearing, touch, proprioception, their sense of balance, etc. The PackBot makes no attempt to compensate for these sense by augmenting they operator’s view (Tittle et al., 2002), it simply presents the raw video feed. This loss of additional senses, and the inability to smoothly transition to different viewpoints present the operator with an impoverished stimulus for the operator. The operators must compensate for this impoverished stimulus by explicitly reasoning about the remote environment and thus the actions available to them. The shift from directly perceiving the environment to explicit reasoning hampers any attempts to create a functional presence.

The controls operators use to drive the platform present at least two significant challenges to functional presence. First is the physical control device operators use. The PackBot has physical controllers: a set of joysticks integrated into the GCS and a gamepad controller, similar to what one would find with a video game console. Both of these controllers rely on a pair of joysticks and a myriad of buttons that can change interaction modes. Each interaction mode remaps some or all functionality of the joysticks. The mapping between
the movements of the physical control device and the result of those movements on the platform are arbitrary (Morison et al., 2015). The mapping can be learned over time; however, any new functionality will have to be learned anew. The second problem with the PackBot’s control is what exactly the controller moves on the robotic arm. The physical control devices’ joysticks manipulate individual servos. Thus, the operator must plan out how the movement of any individual servo will impact the position of the end effector. This saddles the operator a large amount of cognitive work.

Starting from an understanding of Ecological Perception, several fundamental changes can be made to the PackBot’s design that would greatly improve its ability to act as an extender of the operator’s perception. Of primary concern is the link between the PackBot’s ability to move its robotic arm, and its ability to move its viewpoint. One key ability of biological perceptual systems is that they can easily move their viewpoint to sample the world from multiple vantage points. Additionally, these transitions are performed smoothly, producing additional visual information like motion parallax. The current PackBot design does not facilitate easy, smooth viewpoint transitions. There are two promising approaches to alleviating this barrier: Perspective Control and virtual environment synthesis.

The Perspective Control approach, as mentioned above, was created by Roesler (2005) and build upon by Morison (2010). At its core, Perspective Control allows an operator to control viewpoint transitions around a point of interest. The system uses a spherical
coordinate system to describe the movement of the viewpoint, with the point of interest at the center of the coordinate system. This system creates the ability to quickly take different views of the target area and the smooth transitions between them. Additionally, the Perspective Control system moves the locus of control from individual servo motors to the viewpoint itself. Applying this to the PackBot would require a camera mounted on a multiple degree of freedom arm and controlled by software that can calculate the inverse kinematics required to achieve the view specified by the user’s input. The use of movements described by a spherical coordinate system with a constant radius is supported by early research on control metaphors for moving a virtual viewpoint. This early research reported participants using complex combinations of translating and rotating the camera in order to orbit an object of interest while keeping it in view (Ware & Osborne, 1990). This observed behavior confirms the utility of the viewpoint movements proposed by Roesler. However, while only the best performing participants in Ware’s research could exert the required control over the viewpoint to achieve this motion, the use of Roesler’s spherical coordinate system make the same movement of the viewpoint trivially easy for users.

While the new Perspective Control camera could be controlled by the existing hardware, the mapping between that hardware and the movements of the camera remain arbitrary. Morison (2010) proposes a physical control device based on Perspective Control that would create a one-to-one mapping between the movements of the physical controller and the movements of the Perspective Control camera in the spherical coordinate system.
An additional approach to given users control of a single viewpoint that has become popular in laboratory based unmanned ground vehicles is the creation of 3D virtual environments. This approach is used in IHMCs entry in the DARPA Robotics Challenge (M. Johnson et al., 2015). This data fusion approach takes data from several different sensor modalities and fuses them into a 3D virtual environment that users can then explore with a virtual camera. These 3D virtual environments are often build from LIDAR range data and the surfaces are textured from the onboard camera’s data. By building this virtual environment, users can take any view imaginable with the virtual camera. Combining the fused sensor with a scaled virtual representation of the robotic platform could be a powerful tool that allows operators to quickly identify the actions available to them in the remote environment. It is interesting to note that the Perspective Control Technique and the 3D environment technique are not mutually exclusive. In-fact using Perspective Control to constrain the virtual camera’s movement, and the Perspective Controller to move the camera would provide a more natural mapping to the user’s exploration of the environment.

*Measuring Direct Perception in Human Sensor Systems*

As previously mentioned, a small body of literature has grown up surrounding the measurement of human sensor systems’ ability to directly perceive remote environments. While the studies making up this body of literature are conducted for different purposes,
they all derive their methodology from well-known studies of affordance perception in the Ecological Perception literature.

One of the most commonly studied domains that utilize teleoperated robotics is Urban Search and Rescue (USAR). In their landmark report about the usage of robotic platforms in the aftermath of the World Trade Center attack on 9/11, Casper & Murphy (2002) identified many common challenges operators faced in the field. One of these key challenges was the ability to judge whether the robot platform could pass through apertures in the deconstructed building. The ability to perceive the passability of an aperture in a remote environment has become the subject of many studies (Jones, Johnson, & Schmidlin, 2011; K. S. Moore, Gomer, Pagano, & Moore, 2009). A brief comparison of these studies follows.

Moore et al. (2009) provide a first in kind study attempting to quantify differences in operator performance when they have direct line of sight (DLS) to their platform and when they are teleoperating (TO) the platform. Their study measures operator’s ability to perceive the passability of an aperture under the DLS and TO conditions. Two experiments were carried out. In the first experiment, the investigators asked participants to judge the passability of various sized apertures with respect to three differently sized robots under TO and DLS conditions. They found that in the DLS condition participants consistently underestimated the impassability boundary for all platform sizes (i.e. the aperture was much wider than the platforms’ width). The investigators hypothesize that
this is to give the robot an additional margin of safety (Mark et al., 1997). In the TO condition, there appeared to be an effect for robot size. Perception of passability with reference to a small robot was found to be accurate. However, as the size of the robot increased participants began to underestimate the impassability boundary. Interestingly, the authors note that in classical studies of affordance perception participants’ judgements did not vary after the size measurements were converted into unit-less π numbers (Warren, 1984), however under the TO condition, participants’ became much more varied in their answers.

The second experiment conducted by Moore et al. (2009) sought to better understand factors influencing performance in TO by varying the camera sensor’s height and the platforms’ distance from the aperture. The choice of these variables was based on Woods et al. (Tittle et al., 2004)’s comparison of camera height to an operator’s natural eye-height during locomotion. Consistent with experiment 1 participants overestimated the size of the aperture relative to the platform’s width, causing them to judge that the platform could pass through an aperture when it could not. This effect appeared to magnify as the camera height approached the ground and when the platform was further from the aperture.

In their 2011 study of teleoperated systems Jones et al. attempt to parse the perception of passability and the perception of drive-ability with reference to a sensor platform. Additionally, the studied compared these perceptions with operators’ ability to actually
drive the platform through the aperture. In experiment 1 participants drove a small remotely controlled vehicle through apertures of different size. As the vehicle approaches the aperture, participants were asked to judge whether or not the robot could fit through the aperture. The results suggested that participants were accurate at judging the passability of the aperture, correctly judging that the platform was smaller than the aperture. However, they were not often successful at driving the platform through the aperture without hitting the sides. This suggested to investigators that the relevant constraint for successfully driving the vehicle through the aperture was the size of the aperture relative to the robot’s width plus a safety margin. Experiment 2 investigated the perception of this safety boundary in more detail. A second condition was added to in which participants were asked whether they could drive the vehicle through the aperture without hitting the side (labeled drivability). Investigators then compared the perception of passability to drivability. The results showed that participants accurately perceived the passability, however did not accurately perceive the drivability of the aperture. The investigators draw three general conclusions from their work: 1) Operators should base their decision to drive a vehicle through an aperture on a judgement of drivability not passability. 2) Results from experiment 2 suggested that drivability judgments do improve with practice. They call for future research to look at training in more depth. 3) The suggest that platform’s selected for a given mission be at least 20% small than any aperture they may encounter.
Organization of this document

The dissertation is split up into four chapters: Introduction, Fitting the Psychometric Function, Viewpoint Motion and Direct Perception, and Discussion.

Chapter 1 contains the reasoning behind the 3 experiments conducted in this dissertation and contains summaries of previous research done in fields relevant to the experiments, and remote sensor platforms in general. This chapter contains background research from two areas: Gibson’s ecological approach to visual perception, and previous work on classifying and measuring human sensor systems as perceptual systems. The work described in each background section informs the design of the three experiments.

Chapter 2 describes the details of experiments 1 and 2. Both experiments were conducted with the goal of verifying that the performance of human sensor systems match what has been observed in research on human perception. The first section of the chapter discusses the reasoning that lead up to the experiments’ execution. The methods section gives a detailed description on how the two experiments were implemented and identifies the previous works that informed decisions in the experiment design. Lastly the results section details data analysis performed to verify that human sensor system performance does match what has been observed in human perception.

Chapter 3 describes the details of experiment 3. Building on the methodology established in chapter 2, experiment 3 measured the effects of viewpoint motion on the ability of
human sensor systems to directly perceive the remote environment. Chapter 3’s structure mirrors that of chapter 2. The first section outlines the reasoning behind the experiment. The method section details the set up and execution of the experiment, focusing on the methodological differences with chapter 2. Lastly the results section describes the data analysis performed.

Chapter 4 discusses the results of both chapter 2 and chapter 3 and their implications for the design, testing, and fielding of human sensor systems. This section also details potential shortcomings and lessons learned during the design and execution of the three experiments. Lastly this chapter discusses a larger series of studies that will continue this research in the future.
Chapter 2: Human-Sensor Systems Considered as Perceptual Systems

As discussed in Chapter 1 previous work has claimed that human sensor systems are perceptual systems, and thus are subject to the same strengths and weaknesses of humans perceiving their surrounding environment. However, as Sorkin & Woods (1985) showed, the psychophysical performance of a human and a human-machine system are not necessarily the same. The interaction between the human component and the machine component can drastically alter how the system performs. Thus, the claim that Human-Sensor systems are similar to human perceptual systems, made explicitly by (Morison et al., 2015) and implicit by (Fong, 2001; Jones et al., 2011), is certainly justifiable from a theoretical standpoint it has not been empirically investigated. To verify these claims one can show that the perceptual performance of the Human-Machine System is best fit by the statistical models that the psychometric literature uses to describe human perceptual performance. If the performance of both entities is best described by the same model, it is reasonable to conclude that they are the same.

This chapter will present data from two separate investigations of Human-Sensor System performance in order to show that that it is best fit by the same psychometric functions used to describe human performance. Both experiments presented in this chapter sought
to investigate psychophysical performance by measuring Human-Sensor Systems’ ability to directly perceive the reachability of remote objects. These experiments were designed based off of classic studies in ecological perception (Mark et al., 1997; Warren, Warren, Whang, & Whang, 1987). Both experiments seek to establish a stimulus threshold at which the object goes from being perceived as reachable to being perceived as unreachable, by the remote sensor platform’s manipulator arm. Additionally, the experiments attempt to measure the reliability with which the Human Sensor Systems apply this threshold. These values are obtained using standard methods from psychophysics which sample the Human-Sensor System’s perception at many different stimulus levels, in this case distance from the remote sensor platform.

The primary goal of both experiments in the chapter is to determine how well the performance of human-sensor systems can be mapped to the psychometric functions used in psychophysics to model human performance.

*Hypothesis 1* - The performance of Human-Sensor System will best be fit using sigmoid logistic regression with a high goodness-of-fit, regardless of sampling method or visual condition.

The two experiments described in this chapter both follow the same general design, however differ in two important ways. The primary difference is the sampling method used to determine the stimulus level, or target distance, for each trial. The first experiment, hereafter referred to as the staircase experiment, used an adapted version of
the Weighted Transformed Staircase method with fixed step size (Alcalá-quintana & Garcia-pérez, 2007; Levitt, 1971; Wetherill & Levitt, 1965). The second experiment, hereafter referred to as the constant stimuli experiment, used the classic method of constant stimuli (Simpson, 1988; Urban, 1910). The second difference between the two experiments was the number of sessions participation was spread over. Participants in the staircase experiment completed their participation in one session. Participants in the constant stimuli experiment spread their participation over three sessions. The design and execution of both experiments will be detailed in the next section.

Methods

This section will discuss the setup and execution of the two experiments and includes information on participants, experimental design, apparatus, and procedure. The results of these experiments will be discussed in the following section.

Experiments and Sampling Methods

As described above, the two experiments each used different sampling methods to determine target location in each trial. The first collect used a Weighted Transformed Staircase method (García-Pérez, 1998; Levitt, 1971) to sample the space of possible target locations. Experiment 2 used the Method of Constant Stimuli (Simpson, 1988). In the ecological perception literature, these methods are used to determine the threshold between perceiving the ability to successfully take an action or not (See Chapter 1). However, the purpose of these two experiments were to determine the similarity of
Human-Sensor System performance to human performance. As a result, these two methods were used solely as a means to sample the stimulus space and provide a reasonable stopping rule. Participants’ threshold and reliability data were determined by fitting the data points generated from the methods with a psychometric function. Two different sampling methods are used in this chapter because the sampling rules for each method have different implications how the Human-Machine System’s underlying psychometric function is sampled and thus different opportunities of bias. By using two very different methods, it is hoped to provide converging lines of evidence that the performance data is best fit by the psychometric function. The rest of this section will discuss the advantages, drawbacks, and implementation of the three sampling methods utilized in this dissertation.

Experiment 1 utilized a staircase sampling method based on the recommendations of Garcia-Perez (2007). As discussed in chapter 1, Garcia-Pérez recommends specific combinations of up/down rules and step sizes. Experiment 1 implemented staircases with a 3/1 up/down rule and a step size ratio of 0.968. This combination will target the 79.38% point on the psychometric function, and it optimal when applying fixed step-size staircase methods to a yes-no experimental design. The 3/1 up/down rule means that the participant must answer negatively three times before the stimulus level is increased, but must answer positively only once before the stimulus is decreased. The step size determines how far the target will move from trial to trial.
In order to avoid participants learning the simple distance selection rules, experiment 1 implemented staircases in interleaved pairs, alternating which staircase was used from trial to trial. The members of a staircase pair used inverted up/down rules (i.e. 1/3 and 3/1 and step size ratios) (García-Pérez, 2001). The interleaved staircases provided a seemingly more randomized experience to the participant. The choice to invert the up/down rules in one of the pair members was done deliberately so that the two points estimated on the psychometric function would be symmetrical around the 50% threshold value of the function. By averaging the two together, the participant’s reachability threshold value could be estimated non-parametrically. Additionally, having estimated two points on the psychometric function, a rough slope metric can be obtained by taking the difference of the two points.

Experiment 2 implemented the Method of Constant Stimuli (Urban, 1910). This approach preselected a number of stimulus levels to be sampled multiple times. This type of sampling should result in the least biased sampling, as points are evenly spread and sampled an equal number of times. While the Method of Constant Stimuli is a highly reliable approach, it is one of the least efficient sampling methods. As a result, experiment 2 split the data collection into three session for each participant in order to avoid participants losing focus. Each session corresponded with one experimental condition.
The parameters of chosen for the experiment 2 were the result of trying to balance three primary constraints: maximizing the sampling resolution, obtaining sufficient samples at each stimulus level, and minimize the risk of participants becoming bored. Each condition sampled the participant 252 times over 28 stimulus levels, or target distances in this case, resulting in 9 samples per level. The range of sampling was determined by balancing tradeoffs between a desire for higher sampling resolution, the number of trials that could be administered, and information from both pilot data and experiment 1. The final range of distances selected was from .5 arm lengths to 2.0 arm lengths. The minimum distance was selected to ensure that the target would be visible under all conditions. The furthest distance was selected based on the likelihood that participants would judge the target reachable in experiment 1 and the pilot data.

**Participants**

12 participants were recruited to take part in experiment 1 and an additional 12 were recruited to participate in experiment 2. All participants were aged between 18-46, and were recruited via email or posters at the Ohio State University. Participants in the first experiment spent between 90 minutes and two hours in the lab during their participation. Participants in the second experiment attended three 45-60 minute data collection sessions at the lab within the space of a week. The Constant Stimuli approach required many more trials than the Staircase method, and as a result it was decided to spread data collection several sessions to minimize the chances of participants becoming distracted.
Experiment Design

Both experiments used a 3x1 within subjects design. The independent variable was the visual environment in which the participant made the reachability judgement. Participants were exposed to a visually sparse environment, a moderately rich visual environment, and a rich visual environment. All other aspects of the scene were either static or randomly varied (see the Apparatus section of this chapter for further description). The order in which these conditions were presented was counterbalanced in order to eliminate any learning effects that might confound the results.

Experimental Conditions

Each experimental condition was designed to offer a different amount of visual information for the participant to perceive when making their judgement of target reachability. The conditions are referred to sparse, moderate, and rich. This visual information included in each of the three conditions was chosen to approximate the types of visual information available in different environments remote sensor platforms are used in.

The sparse condition (Figure 5) contained a small amount of visual information, including target size, target position within the camera’s field of view, and shading. The target object had a solid color with no texture applied. This condition was created to loosely mimic operating a sensor platform underwater, where there is no consistent horizon, indirect lighting and few useful cues to the scale of objects. Outside the normal
experience of human operators, this condition would be more difficult for humans to work in in-person. With the participants’ perception limited to the fixed sensor, this condition was predicted to elicit the worst threshold accuracy and worst judgement reliability.

Figure 5 – The participant’s view from the primary camera onto the sparse, or low, condition environment. This environment was designed to give very little information about the 3D layout of the remote environment. The condition is roughly analogous to operating a platform underwater.

The moderate condition took the sparse environment and added a ground plane. The ground plane served as a landmark to help judge the height of the target (Figure 6). However, without shadow information the ground plane would only be somewhat helpful in judging reachability. One potential confound with the ground plane was the texture put
on it. If the texture contained obvious patterns, participants could compare the screen position of the target to the pattern and use this as a kludge for judging reachability. To counter this confound, a random black and white noise texture was applied to the ground plane. Additionally, from trial to trial the orientation of the texture was randomly changed. The visual environment in the moderate condition resembled a deconstructed environment like a collapsed building (see Figure 2). In these environments, there is little reliable or useful shadow information for operators to pick up on. Additionally, surfaces surrounding the remote platform are extremely irregular and provide few, if any, visual landmarks to help judge the relative size or distance of objects.

Figure 6 – The participant’s view onto the medium condition environment. The medium environment adds the ground plane as a visual landmark. This environment was designed to be roughly analogous to operating in a deconstructed environment. These environments have ambiguously scaled landmarks and few consistent shadows.
The high condition added two sources of visual information: a texture to the target and a strong, directional light source. The sparse and moderate visual conditions contained a diffuse light, which cast no shadows, appearing to be source-less. The rich condition maintained the overall luminance value, but added shadows being cast from a specific point above the platform’s chassis. With this change, the target and the manipulator arm cast shadows on the ground plane (Figure 7). Under some rare circumstances the manipulator arm cast a shadow on the target. The size of the shadow on the ground plane would act as a very strong piece of information with which to make the reachability judgement. The rich visual condition was designed to mimicked operating a robotic platform outdoors in direct sunlight. While a real environment would also contain many objects of either familiar or ambiguous size, the virtual environment was kept simple in order to control for potential shortcuts to judging reachability. The decision to keep the virtual environment sparse was also based on the unknown effect of having a human operator’s viewpoint only 2-3 feet from the ground. Eye height is a prominent source of visual information (Mark, 1987; Wraga, 1999). Systematically altering effective eye height may have a biasing effect on how the participant perceives the scale and distance of objects. Some evidence of this effect can be found in previous studies of human-sensor system perception of passability (K. S. Moore et al., 2009).
Figure 7 – The participant’s view on to the high condition environment. This environment added shadow information and texture to the target object. The environment provides similar visual information to operating a platform in direct sunlight. The target’s texture and the show on the ground plane should provide a sense of scale to the environment, while the shadow itself is a source of information.

Apparatus

Both experiments were performed at computer workstations, using MacBook Pro laptops. The laptop’s screen measured 19 inches on the diagonal, and used a resolution of 1920x1080. Participants produced input using a standard three-button mouse. Participants were run in an area isolating them from outside noise and distraction.

The experiment apparatus was created using a 3D videogame authoring engine called Unity 3D published by Unity Technologies (https://unity3d.com/). The Unity 3D engine
provides a suite of capabilities needed to simulate and render 3D environments. The program created within Unity handled rendering of the 3D environment, the layout of the environment, the logic of the user interface, the timing of the stimulus exposure, the user’s input, the various sampling algorithms, and the recording of test data. The logic of the software apparatus was designed to be simple. The simple program and the high-end performance of the laptop ensured that participants always viewed the simulation at 60 frames per second or higher.

The software apparatus consisted of two primary components, the virtual environment and the user interface. The virtual environment stood in for a physical remote environment, and simulated a camera feed into that environment. The user interface dictated the flow of the data collection session and allowed users to provide input.

The virtual environment consisted of a blue background, a flat ground plane, spherical target object, a directional light, a camera sensor, and a sensor platform. The sensor platform used was based off of the Talon platform built by QinetiQ North America (https://www.qinetiq-na.com/products/unmanned-systems/talon/). The platform included a manipulator arm, used in a bomb disposal type load out of the Talon, with four controllable servo-motors (see Figure 8). The position of the primary camera was determined based off of a common type of camera used in many different sensor platforms, including the talon. The camera was positioned on a stalk above, behind, and slightly to the left of the manipulator arm. The orientation of the camera was such that the
chassis of the robot was visible at the bottom of the feed as a visual landmark helping the participant understand the position of the camera in relation to the platform’s chassis. Note that on most platforms that utilize this camera perspective, the camera can pan and tilt to help the user look at the environment around the remote platform.

Figure 8 – The remote sensor platform used in the virtual environment is based off of the Talon platform (left panel). The virtual platform has four servo motors to move the manipulator arm. The servos are highlighted the left panel and labeled A, B, C, D. The left panel also shows two different camera positions. On the right side of the panel is the primary, or stalk camera. On the left side of the panel is the camera mounted on the manipulator arm used in Chapter 3. Both cameras correspond to their positions on the actual Talon platform.

Many aspects of the virtual environment were randomly varied from trial to trial in order to minimize any additional sources of visual information that would help participants perceive the distance of the target object. These aspects included target height above the ground plane, target diameter, target color, orientation of the direct light source, ground plane texture scale and orientation, and the configuration of the mechanical arm. The
mechanical arm required additional constraints because it provided an important landmark from which to judge reachability. In order to eliminate the arm as a confound it was important not only to keep it in view for each trial, but to also have it fill roughly the same percentage of the sensor feed. To achieve this consistency 20 arm configurations were predefined prior to the study, which were then selected at random at the beginning of each trial.

Procedure

Both experiments followed a similar structure. The general structure of the experiments were nested components. The experiments were comprised of conditions, each condition was comprised of phases, each phase was comprised of trials. See Figure 9 for a summary each experiment’s structure. Both experiments were comprised of 3 experimental conditions: low, medium, and high. Each of those three conditions were comprised of a training phase, a practice phase, and a data collection phase. The first two phase phases were designed to familiarize the participant with the experimental apparatus, learn about the manipulator arm’s capabilities, and have a chance to practice the experimental task. The following will describe the procedure for the staircase method, bear in mind that the only procedural change in experiment 2 was that each condition was split into its own session.

Once a participant arrived, they were asked to sign an informed consent document and were given a 10-minute verbal briefing. This briefing explained several topics and had
two purposes: bring all participants to the same level of understanding before beginning the study, and provide sufficient context to keep participants motivated while performing the tedious experimental task. The verbal briefing began by explaining the purpose of the study. Participants were taught what a sensor platform is, why they’re useful, what their major drawbacks are, and why human perception might be important. The experimenter covered, at a very high level, human perception and affordances. Participants were told that the purpose of the study was to measure their ability to perceive a remote environment when their perception was completely mediated by the remote sensor platform. After setting up the background of the experiment, the briefing covered the structure of the experiment and the different phases the participants would experience. At the end of the briefing participants were given two final instructions. First participants were reassured that any target appearing inside the maximum reach of the manipulator arm were to be considered reachable. This was added in response to several pilot participants believing that if the target appeared too close to the platform’s chassis, the arm didn’t have the degrees of freedom to reach it. The second, and most important, instruction was to answer every trial as quickly as possible without actually guessing. Guessing was explicitly defined as ignoring the sensor feed from the platform and randomly picking one of the two response buttons. The experimenter emphasized that the purpose of the experiment was to measure their ability to “just know” what the layout of the environment was without explicitly reasoning about it. The experimenter told participants to give their “first-impression” or their “gut-instinct,” phrasing the instruction several different ways was intended to help overcome any linguistic barriers
between participants and experimenter. After the verbal briefing, the computer workstation was started and the rest of the session was guided by the software.

Figure 9 – The structure of both experiment 1 and 2. Each condition is made up of three phases (Training, Practice, and Data Collection). Each phase has a number of trials. Both Training and Practice have eight trials. In experiment 1 the data collection phase had a variable number of trials, based on the participant’s performance. In experiment 2 the data collection phase had 252 trials total. Note that this diagram does not include the frequent breaks given to participants in both experiments.

The first activity participants engaged in was the familiarization phase of the first condition. The familiarization phase was designed to help participants become more familiar with the capabilities of the manipulator arm and to see what the target looked like at various distances, in various visual environments. During this phase, the
participant took manual control of the manipulator arm and were tasked with using the arm to hit the target ball resting on the ground. By forcing participants to essentially perform a problem-solving task (i.e. discover the configuration of servo positions that would result in the ball intersecting the target), participants were forced to reflect on how the arm worked and what it was capable of. This learning was necessary to ensure that even participants with no background in virtual environments or robotics had a basic understanding of the arm’s operation and thus its ability to reach for objects.

Figure 10 – The participant’s view of the environment during the training phase. The box in the upper left corner is a graphical interface participants interacted with in order to move the manipulator arm. The goal of each trial was to move the arm to make contact with the ball. The arm is depicted here hitting the ground with the end effector, triggering a warning on the graphical interface.

The familiarization phase consisted of eight trials. In each trial the target was presented at a new distance, and a new servo was ‘unlocked’. The same set of trials (i.e. target
distances and manipulator arm configurations) were presented across conditions and across participants. The target distances, and the order in which they were presented, were chosen so that the task of hitting the ball with the arm would get more and more difficult across the first four trials. On trial one participants only had manual control over the first servo, the servo at the base of the arm and that moved the servo side to side. The target was positioned so that actuating the single servo would result in hitting the ball (see Figure 10). On the second trial the participant had control over the first and second servo. The target was positioned so that hitting it would require actuating both servos. The next two trials unlocked servos three and four, giving the participant control over all four servos. After the first four trials, the target object was placed at various distances equally distributed around the area within reach of the manipulator arm.

After completing the familiarization phase, all participants should have a common understanding of what the manipulator arm was capable of reaching. While this is an important step, participants also needed to become familiar and comfortable with the primary experimental judgement task. Pilot data suggested that if participants were not given some time to practice the task, that the initial data collected would be confounded with a sharp learning curve as they adjusted to the new task. As a result, in each condition participants completed a practice phase before data collection began.
Figure 11 – The series of UI states for each trial. The initial stimulus mask and its countdown timer, the camera feed with buttons to make a judgement, and the post-stimulus mask. The post-stimulus mask only appeared if participants did not make their judgements within three seconds. The software remained at the post-stimulus mask until participants made a judgement.

The training phase mimicked the data collection task, asking participants to judge whether or not a spherical target object was within reach of the platform’s manipulator arm. Each trial followed the same sequence: a stimulus mask with countdown, timer stimulus presentation, and post-stimulus mask (see Figure 11). The stimulus mask was a grey screen featuring large white numbers in the center of the screen. The numbers counted down from 2 to 1 then the mask disappeared, showing the feed from the remote platform. The stimulus presentation was shown to participants for no more than three seconds. This value was used to minimize the chances that participants tried to explicitly reason out the distances of the object, rather than reporting their perception (Heft, 1993). When the camera feed became visible, the participant was also presented with two white buttons at the bottom of the screen. The buttons were labeled Reachable and Unreachable. To answer the question, participants simply used the mouse to click the
appropriate button. After pressing one of the buttons, the software paused and gave the participants feedback about whether they were correct. Feedback was given through a small text box with the work correct or incorrect displayed inside it. Additionally, the box was color coded red or green based on the feedback. The feedback, along with the sensor feed from the camera, was kept on screen for two seconds so the participant could look at the image again before the next trial began. If participants failed to press one of the two buttons with the three seconds of stimulus presentation the stimulus was removed, leaving the screen black except for the two response buttons. The trial ended once one of the buttons were pressed. After pressing a button the stimulus was put back on the screen for two seconds, after which the next trial began.

Participants were given eight practice trials. The target positions and arm configurations remained the same across conditions and across participants to ensure all participants received the same level of training. An important factor in the training was ensuring that participants were exposed to the entire range of possible target positions and arm configurations in order to avoid biasing their judgements during data collection. As a result, the eight target positions ranged from the closest possible distance out to approximately 2 times the length of the manipulator arm. Distances were more tightly clustered around the arm’s reachability boundary, as those should be more difficult to judge than distances very close to, or very far away from, the boundary.
In addition to ensuring participants were exposed to all possible target distances, an effort was made to expose participants to many different possible manipulator arm configurations (i.e. set of servo positions that dictate the arm’s positioning). By only showing participants a particular subset of possible arm configurations, the experiment would run the risk of biasing participants’ judgements. To avoid this biasing, a set of 20 pre-made arm configurations were created and used throughout both experiments. Ensuring a representative sampling from all possible arm configurations could be a complex issue, as the arm has 4 degrees of freedom and every end effector position can be achieved using several different servo configurations. This problem was simplified by splitting the range of possible end effector positions into four quadrants (see Figure 12) and choosing five end effector positions in each quadrant. 20 different arm configurations were made by using these end effector positions and finding two different servo configurations that would achieve each position. Eight of these arm configurations were selected to be used during the training phase. These eight configurations were chosen from all four quadrants.

After completing the condition’s training phase, participants began the data collection phase. The two major differences between the two phases were the removal of feedback during data, and the manner in which the virtual environment changed from trial to trial. During data collection, many aspects of the virtual environment were randomized to prevent participants from using workarounds, or kludged, to judge the target’s distance (see the apparatus section for a detailed description of the randomization).
Figure 12 – The space of all possible end effector positions split into quadrants. A pool of 20 pre-made manipulator arm configurations (made up of sets of servo positions) was created to set the arm’s position in each trial. A member of the pool was randomly selected in each trial. In order to reasonably sample over the space of all possible end effector positions, the space was split into quadrants. The pool contained five configurations for each quadrant.

Participants were given many breaks over the course of each session. These breaks were optional and participants were given a countdown timer showing suggested break duration. The verbal briefing made it clear to the participants that breaks were optional. During the experiment 1, the primary breaks occurred after each experimental condition. These primary breaks gave participants up to 10 minutes to check their phones, walk around, go to the bathroom, etc. In both experiments, participants were given shorter breaks after each phase, lasting up to two minutes. Lastly during the data collection phase, in both experiments, participants were given 2 minute breaks after a set number of trials. Generally, these data collection breaks occurred every 15 minutes.
After completing the study session, participants were given a chance to ask questions to the experimenter. After this informal debriefing, participants signed a payment receipt form and were given their compensation for participating.

Results

Both experiments sampled participants’ perception of reachability at various distances from the robot’s chassis. The resulting data was a binary outcome, the reachable/not-reachable judgement, paired with a continuous factor, the distance of the target from the platform. A Maximum-Likelihood technique was used to fit each data set with a logistic function with the form:

\[ P_{\text{logistic}} = \gamma + \frac{1 - \gamma - \lambda}{1 + e^{-\beta(x-\alpha)}} \]

\( \alpha \) represents the participant’s threshold value, or the point on the curve at which they have an equal chance judging the target reachable or unreachable. \( \beta \) is the slope of the function, or the reliability with which the participant applied their threshold. \( \lambda \) is the upper asymptote of the function and represents the lapse-rate, or the probability of an incorrect answer. The \( \gamma \) parameter is lower asymptote, or the guess-rate, and represents the probability of a correct answer when the stimulus is not present. This last parameter is a constant set based on experimental design. The current experiment has a yes/no design and thus \( \gamma \) is set to 0. The result of the maximum likelihood technique was a complete psychometric function for each participant in each condition.
The goodness of fit for each psychometric function was determined using a pseudo-$r^2$ metric (McFadden, 1974). The $r^2$ metric used to report goodness-of-fit for logistic functions is different than what is used for linear functions. While McFadden’s $r^2$ does range from 0 to 1, like the linear $r^2$, its calculation is quite different. As a result, the values reported by McFadden’s $r^2$ are significantly lower than what one would expect from the linear version. Figure 13 shows a matrix of data sets from experiment 2. Each data set is an individual participants’ performance in one experimental condition. Each panel shows the proportion of ‘reachable’ judgements to ‘unreachable’ judgements as black dots and the fitted psychometric function for that data. The three rows of the matrix represent the three experimental conditions. The three columns of the matrix show the data set with the worst $r^2$ value, the median $r^2$ value, and the best $r^2$ value for each condition. Figure 13 has been included to give the reader a sense for the relationship between McFadden’s $r^2$ metric and the grouping of the data.

The majority of the psychometric functions fit to participants’ performance resulted in $r^2$ values above 0.2, which McFadden (1978) called “excellent fit.” Figure 14 shows the distribution of fit values across all 72 psychometric functions making up experiments 2 and 3. Each condition in both experiments is represented by a box and whisker plot. The plot shows in 4 of the 6 conditions all psychometric functions achieved a fit at or above McFadden’s range of excellent fit. The two conditions with functions that fell below excellent fit were the low and medium conditions of experiment 2, which utilized the method of constant stimuli to sample the range of target distances. The low condition
included 6 psychometric function that fell below the excellent fit threshold \(r^2 = .16, .15, .15, .11, .19, .16\). The medium condition included 2 functions that fell below the threshold \(r^2 = .04, .14\).

Figure 13 – Estimated psychometric functions for the worst fit, median fit, and best fit participants across three experimental conditions.

An independent-sample t-test shows no significant difference between the two groups of psychometric function fits. A one-way ANOVA revealed a significant difference in fit
values between the three conditions using the staircase method, $F(2,33) = 3.84, p = .031$. A pairwise comparison with a Bonferroni correction shows the rich condition having a significantly better fit than the sparse condition ($p = .0279$). A one-way ANOVA showed significant differences between the three conditions using the constant stimuli sampling method $F(2,33) = 7.74, p = .0017$. A pairwise comparison with a Bonferroni correction shows that the rich condition was significantly better fit than the sparse condition.

Figure 14 – Distribution of McFadden’s pseudo-R2 across experimental conditions. The green zone represents what McFadden (1978) refers to as “excellent fit.”
Chapter 3: Viewpoint Motion and Perception of Remote Environments

Chapter 2 has confirmed that Human-Sensor Systems perform as perceptual system, however little is yet known about how the interaction between the remote sensor platform and the human operator affects the system’s ability to perceive the environment. This chapter will report on how movement of the sensor platform’s viewpoint effects the human operator’s ability to directly perceive the remote environment. As discussed in Chapter 1, viewpoint movement is a core component of perceptual systems. A smooth and continuously changing viewpoint generates visual information critical to directly perceiving the surrounding environment. The ability for a human operator to change the remote sensor system’s viewpoint should be a critical factor in the joint system’s ability to directly perceive the remote environment, and ultimately task-level performance. Toward this end, a third experiment was conducted to measure the effects different types of viewpoint motion can have on the system’s ability to directly perceive the remote environment.

As discussed in Chapter 1, perceptual systems rely on many types of information available in light, what Gibson called the optic array. However, some of the most powerful information about the arrangement of 3D space around the system comes from
a viewpoint motion relative to the environment. Information from viewpoint movement comes from optic flow and motion parallax. Studies in the psychophysics literature have shown that information from those sources alone are sufficient to allow participants to perceive the 3D layout of an environment. The contribution of viewpoint motion to direct perception is so strong that if remote sensor platforms are to successfully extend the operator’s perception into the remote environment, the platform’s viewpoint must be designed to maximize the visual information generated from its movement.

Figure 15 – All perspectives simultaneously reveal and obscure information (David D. Woods, 2002). Shifting perspective reveals new information about the environment. The upper panel shows a piece of street art viewed from a specific perspective. This perspective creates the illusion of a 3D image interacting with the people around it. However even a small change in perspective breaks the illusion, revealing new information.
The second contribution of viewpoint movement to human-sensor systems is the ability to take a new perspective on the scene of interest. Moving one’s viewpoint between multiple perspectives is a critical part of understanding one’s surroundings, “The view from any point of observation simultaneously reveals and obscures aspects of the scene of interest” (David D. Woods, 2002; see Figure 15). Consequently, an ability critical to perceptual systems is to be able to take new perspectives on a scene of interest, inspecting it from different angles. In this context, the term scene of interest refers to a subsection of the surrounding environment that holds useful information for an agent attempting to achieve a goal (Roesler, 2005). As both Roesler and Morison (2010) advocated, if a remote sensor platform is to effectively extend the human operator’s perception in the remote environment it must give the operator the ability to take new and useful perspectives on scenes of interest around the platform.

Given the power of viewpoint movement, experiment 3 was designed with two aims:

- Determine how types viewpoint motion affect the reliability and threshold accuracy of a human-sensor system’s perception by measuring the relative strength of the two contributions from a moving viewpoint.

  Hypothesis 1a – Viewpoint movements that provides a new perspective onto the scene of interest will facilitate more reliable judgements and a more accurate reachability thresholds than those that remain in place and static sensors.
Hypothesis 1b – The viewpoint motion that provides a continuous view of the target object, and its relation to the platform’s chassis, during a transition between perspectives will facilitate more reliable judgments and more accurate reachability thresholds than all other movement types.

- Determine the extent to which the visual richness of the remote environment impacts perceptual performance under viewpoint motion.

Hypothesis 2 – Reachability judgements will be more reliable and reachability thresholds more accurate in the rich visual condition versus the low visual condition across all four motion types.

To fulfill the first aim of experiment 3, four viewpoint movement types were chosen based on the type of contribution they make to perception. The four movement types were labeled static, ego-centric, pseudo-arm, and exo-centric. The static movement type provided neither rotation nor a translation of the camera and was included as a control condition. The ego-centric motion, a panning rotation of the viewpoint, was chosen because it should not contribute either visual information from motion nor provide a new perspective. The pseudo-arm motion, a viewpoint limited to the degrees of freedom offered by the manipulator arm, was chosen because the target object is not visible during the transition between perspectives and thus it provides a new perspective but no additional visual information from motion. The third motion type was chosen based on Roesler (2005)’s Perspective Control approach and was labeled exo-centric. The exo-centric motion ‘orbits’ the viewpoint around a point in the remote environment (see Figure 17). Thus, the exo-centric movement provides both a shift in viewpoints and a
continuous view of the target area during the transition. Each movement will be discussed in more detail in the following section.

The second aim of experiment 3 was to investigate the impact of the remote environment’s visual richness. To fulfill this aim, all four motion types were sampled in two different visual environments. The rich and sparse visual environments from experiments 1 and 2 were used in experiment 3 to make this comparison. The medium condition was omitted to keep the number of experimental conditions within practical means.

Methods

Experiment 3 used an adapted version of the methods used in Chapter 2. The experimental apparatus from Chapter 2 was modified to use a new sampling method and create a dynamic viewpoint, but was otherwise left unaltered. Participants were presented with the same target object and asked to judge the reachability of the target. Both the training and familiarization phases from Chapter 2 remained the same.

The primary difference between the method used in experiments 1 and 2, and the method used in experiment 3 were the addition of the motion types, and the sampling method used to set the target distance from trial to trial. Due to the similarities between the three methods, this section will only describe elements that have been changed for experiment 3. This will omit discussion of the experimental apparatus and detailed description of the
visual conditions. For description of those elements, see the Experiment Conditions and Apparatus sections of Chapter 2. This section is organized into information about the participants, the experimental conditions, and the procedure.

**Updated Maximum Likelihood method**

As mentioned previously, the experiment 3 used Shen & Richards (2012) Updated Maximum Likelihood method to set the target location from trial to trial. The UML method was implemented through a MATLAB (version R2017a) plugin created by the authors of the method (Shen, Dai, & Richards, 2015). This plugin includes the core functionality needed to employ the UML algorithm. The UML method requires several parameters to be setup by experimenters. The psychometric function threshold, psychometric function slope, psychometric function lapse-rate, and the guess-rate ($\alpha$, $\beta$, $\lambda$, and $\gamma$ respectively) each had associated parameters to be tuned for the most reliable estimation. $\gamma$ is a static value that allows experimenters to set how many alternative choices are presented to participants simultaneously (e.g. in a 3-alternative forced choice task chance performance levels would be 33.33%). This value has the effect of scaling the minimum value of participants’ response. Because this study was sampling well above and below the 50% threshold, this value was set to 0. This highlights a notable difference from standard auditory and visual threshold testing, which use a different scaling of stimulus. When testing a standard auditory or visual stimulus, chance levels of performance are expected when the stimulus is near 0. However, when studying reachability chance performance is expected when the target object approaches the
boundary between what is reachable and unreachable. This relative scale can elicit response rates ranging from 0% to 100%. This change results in the \( \gamma \) being set to 0%, rather than 50%.

Each of the three psychometric function parameters to be estimated by the UML method required several parameters to be set by experimenters before data collection. Each psychometric function parameter, (threshold - \( \alpha \), reliability - \( \beta \), and lapse-rate - \( \lambda \)), are represented by a probability distribution. Setting up the UML function requires that prior estimates of these three distributions be defined as a starting place for the algorithm. Additionally, each psychometric function parameter requires a range of possible values be defined to constrain the resulting estimates. The algorithm also requires an initial stimulus level, in this case target distance, and a maximum stimulus level. Lastly the experimenter must choose how many trials to execute before terminating.

Where possible, the UML algorithm was set up using known values from experiments 1 and 2 to increase the convergence rate of the algorithm. The probability distribution for the threshold parameter was set using the mean and standard deviation from the previous experiments. Additionally, the range of possible threshold values was set based on the maximum and minimum values found in chapter 2. Similarly, the values for the reliability parameter were adapted from the previous experiments. The distance of 2.15 times the length of the manipulator arm from experiment 2 was used to constrain the distances the algorithm could sample. The initial stimulus value was a random value.
ranging from the minimum to the maximum target distances at the beginning of each experimental condition. Lastly the stopping condition of the algorithm was set at 130 trials based on previous research showing reliable convergence for all three psychometric function parameters (Hatzfeld & Kupnik, 2016). The remaining parameters, whose values could not be determined based on previous data, were set using the recommendations of Shen et al. (2015). In addition to the setup of the UML algorithm, experiment 3 added 20 catch trials to each condition in order to minimize the chances that participants discovered any pattern in the placement of the target from trial to trial.

**Participants**

This study recruited 30 participants from the general student population of the Ohio State University. Participants were recruited using posters displayed in common areas throughout the main campus, and through announcements in several senior-level classes. The subject pool consisted of 3 female participants and 21 male participants. Participants ranged in age from 18-45, with a mean of 25 years old (s = 5.46 years). All participants had normal or corrected to normal vision. For their participation in the roughly two-hour data collection session, participants were paid $30.

**Experiment Design**

The third data-collect used a 4x2 within subjects experimental design. The two independent variables were viewpoint motion type and visual environment. The viewpoint motion variable had four levels: ego-centric, pseudo-arm, exo-centric, and
static. The visual environment variable had two levels: sparse environment, and rich environment. As with the previous data-collects, a within subject design was used to minimize effects from individual differences, and to reduce the number of participants required. Each participant was exposed to a total of 8 experimental conditions, measuring the performance of all four motion types in both visual conditions.

The presentation order of the experimental conditions was randomized in order to minimize any ordering effects. The exception to the randomization was the first of the eight conditions presented to each participant. The first condition presented to each participant was randomly selected, without replacement, to ensure that all eight conditions were presented first an equal number of times. This was done to minimize the chance of systematic bias towards either an individual motion condition or one of the visual conditions, as conditions earlier in the presentation order are more likely to effect task learning.

Viewpoint Motion Type
As mentioned previously, four types of motion were selected for experiment 3 based on two criteria: the contributions they made to the operator’s perception and their availability on current generation remote sensor platforms. Two of the motions, labeled static and ego-centric, simulated standard robotic capabilities. The third motion, labeled exo-centric, simulated an idealized motion path. The fourth motion, which was labeled pseudo-arm, represented a tradeoff. The viewpoint moved as if it were attached to the
platform’s manipulator arm, which made it more like arm mounted cameras on commercially viable platforms. However, in order to facilitate comparisons with the other views, and despite the fact that commercially available arms could not create a comparable motion, its beginning and end points perspectives were made identical to the exo-centric motion.

The static viewpoint motion type had the viewpoint remain motionless during each trial. The static type represented the control against which to compare the other motion types. The camera sensor in this condition was positioned above, behind, and slightly to the left of the platform’s manipulator arm, giving the participant an ‘over the shoulder’ view of the manipulator arm. The camera was oriented looking straight ahead, parallel to the forward axis of the platform. This positioning was based on the mast camera used on the Talon platform (see Figure 8). Because this motion type neither afforded the participant a new perspective on the scene of interest, and because it produced no visual information from movement, participants using this motion type were predicted to have poor performance.
Figure 16 – The movement timeline for each trial in the ego-centric conditions. The timeline begins with the stimulus mask. At time 0 the stimulus is revealed. From 0 – 0.2 seconds, no movement occurs (labeled static). From 0.2-1.8 seconds the viewpoint pans to the left using the quadratic easing function highlighted in grey. The position of the camera is shown at four different times. The quadratic easing function smoothed the movement from the starting orientation to the end.

The second viewpoint motion type was a simple horizontal rotation, or pan, to the left.

The Perspective Control approach calls this type of motion ego-centric or an inside-out view (Morison, 2010). This motion type was included because it mimics common functionality on remote sensor platforms. However, while this type of movement does serve an important function, it does not generate visual information from its rotation. Nor does ego-centric movement translate the viewpoint to a new perspective on the scene of
interest. The viewpoint in the ego-centric condition was mounted in the same positions as
the viewpoint in the static condition. The viewpoint’s orientation began looking directly
down the z axis of virtual environment, the same orientation as the sensor platform itself.
Over the course of 1.6 seconds, the camera panned to the left a total of 30 degrees. The
acceleration of the viewpoint’s rotation was not instantaneous, as it was thought that
would be more disorienting, but was controlled with a quadratic easing function (see
Figure 16). As a result of the easing function, the rotation of the viewpoint began very
slowly, accelerating until it reached its maximum velocity at the middle of the pan
movement. The rotation then decelerated, rotating very slowly during the last several
degrees of the motion. The quadratic easing function was chosen across all viewpoint
motion types because it maximized the amount of time with a slow-moving viewpoint.

The third viewpoint motion type was what the Perspective Control approach calls exo-
centric motion (Morison, 2010). This type of motion both translates and rotates the
viewpoint relative to the target, moving smoothly from one perspective to another while
keeping the target in view. An exo-centric movement translates the viewpoint along the
outside of a sphere, centered on an origin point in the environment (see Figure 17). For
experiment 3 the origin point position was placed at the manipulator arm’s maximum
reach (46.62 units) along the depth axis. The origin point’s position in altitude was set to
18.09 Unity3D units above the ground plane (0.33x arm length), the same altitude as
manipulator arm’s mounting point on the platform. The origin point’s position was
chosen to be a landmark to help participants judge reachability. This landmark only
become apparent as the viewpoint moves. Due to motion parallax, the space around origin point, in this case the maximum reachability boundary, appears to move very little as the viewpoint translates compared to the space further away. Thus, the closer the target is placed to the actual reachability boundary the less it appears to move on the screen as the viewpoint translates.

When using the exo-centric movement type, the viewpoint started in the same position as in the static and ego-centric types, however its orientation was slightly different, looking directly at the point of rotation in the remote environment. The viewpoint smoothly moved to a second perspective over 1.6 seconds. Using the spherical coordinate system, the movement rotated the viewpoint 26 degrees around the origin point, along the \( \phi \) axis (see Figure 17). Describing the movement in a Euclidian space, the viewpoint translated both forward and to the left of its starting point, rotating to the right to keep the origin point in the center of the screen.

The position of the ending perspective was chosen by balancing two constraints. The first constraint was the assumption that a theta rotation of 90 degrees would result in near perfect threshold accuracy as the viewpoint would be looking perpendicular to the direction the arm was facing. The second constraint was what would constitute a reasonable amount of rotation in a physical, fielded sensor platform. Given the physical environments these platforms are normally deployed in, it is reasonable to assume that the viewpoint would be unable move much further than the edge of the platform’s
chassis. Given the proportions of the Talon sensor platform, and the placement of the origin point, this is equal to roughly 25 degrees of theta rotation. As a result of these two constraints, a theta rotation of 26 degrees was chosen for this motion. The position and orientation of the viewpoint in this perspective tries to maximize the change in view angle to the target while remaining realistic to the real-world circumstances.

Figure 17 – The movement timeline for the exo-centric movement type. The timeline is made up of the same segments as the ego-centric movement and the camera’s position is show at four different points in time. The viewpoint’s movement describes an arc centered on the target object at the top of each panel.
The fourth viewpoint motion type was labeled pseudo-arm. This movement represents a hybrid between the exo-centric movement and the movement of an arm mounted camera on commercially available platforms. As the name implies, the viewpoint is attached to the sensor platform’s manipulator arm, just below the wrist joint, as on many commercially available platforms. As a result of its mounting point, the viewpoint’s position and orientation changes when any of the arm’s servos actuate. While a strict comparison between commercially available movements and the idealized movement of the exo-centric condition would be preferable, this hybrid approach was needed to make an arm mounted movement experimentally viable. Manipulator arm movement on commercially available platforms is so slow and deliberative that direct perception of the affordances in the environment is impossible except when the arm is not moving. Arm movements can take minutes to accomplish and require operators use a process of trial and error to achieve a desired perspective. Thus, pseudo-arm movement was designed to fit into the short stimulus exposure time required by the experiment. The pseudo-arm viewpoint type, like the exo-centric type, provides both translation and rotation, moving the viewpoint to a new perspective at the end of the animation. Over the course of 1.6 seconds the viewpoint moved from an initial perspective, located slightly in-front of the platform’s chassis and at the same altitude as the manipulator arm’s base, to the same position and orientation as the end of the exo-centric motion.
Although the pseudo-arm motion type was closer to the exo-centric movement than movement of a commercially available manipulator arm, it featured three major differences from the exo-centric motion. The first difference was that the exo-centric motion kept the target in view of the sensor during the entire movement, while the pseudo-arm motion did not. The pseudo-arm motion was constrained by the manipulator arm’s servo movements. In many common UGVs used by real practitioners, the control of the manipulator arm is limited to manipulating individual arm servos one at a time. Beyond the difficulty of envisioning and executing complex movements, this system results in jerky and complex viewpoint motion that does not keep the scene of interest in view of the sensor. The second between the two motion conditions was that because the viewpoint is attached to the manipulator arm, the participant could no longer see the manipulator arm and its position relative to the target object from the second perspective. This ability to see both the manipulator arm and target in the same view was one of the major strengths of the exo-centric condition. The last difference between the movement types was the starting position. The exo-centric viewpoint began its movement in the same ‘over-the-shoulder’ view as the static and ego-centric movements, while the pseudo-arm movement began much further forward, and lower.
Figure 18 – The timeline for the pseudo-arm viewpoint motion type. The timeline shares the same segments as the previous two movement types. This motion type fixed the camera to the manipulator arm. The arm ended its movement in the same position and at the same orientation as the camera in the exo-centric movement condition.

As previously mentioned, the pseudo-arm movement was designed as a compromise between the slow, complex movement experienced with current platforms and the need to simply the movement to fit it into the 1.6 second stimulus exposure used in the other conditions. The resulting movement differed from commercially available movements.
The differences were primarily caused by the change from individual servo movements to compound movements involving several servos moving simultaneously. The pseudo-arm movement was broken into three compound movements: an initial rotation, linear movement, and final rotation (see Figure 18). The initial rotation panned the viewpoint to the left, similar to the ego-centric movement, in order to allow the arm to reach to the left of the chassis (necessary to position the viewpoint in the same position as in the exo-centric movement). The linear movement was the result of the servos moving to place the viewpoint in the same position as in the exo-centric movement. The final rotation involved the rotation of several servos to achieve the same orientation as the exo-centric condition.

The use of the compound movements created two major differences between commercially available manipulator arm movement and the pseudo-arm movement. The viewpoint’s path is smoother, and much faster than what would result from most currently available platforms. This difference must be considered when comparing the results of this study to what current platforms are capable of. The second caveat is that two additional degrees of freedom had to be added to the manipulator arm beyond what the vast majority of UGVs are built with. These degrees of freedom were added in order to achieve the same final perspective as the exo-centric motion. Without the extra degrees of freedom the final rotation, which oriented the viewpoint to look back at the target, would not be possible. The extra degrees of freedom, and the speed of the movement, and
the smoothness of the movement were all added in order to make comparisons between the pseudo-arm movement and exo-centric motion equitable.

One additional movement was added to the three conditions that moved the viewpoint in order to prevent participants from using a work around to judge reachability. In both the exo-centric and pseudo-arm movements, the final position of the viewpoint resulted in a perspective located to the left side of the target and platform’s chassis. From this perspective, participants could easily measure the position of the target along the horizontal axis of the computer screen in order to approximate the target’s distance from the platform. This work-around was made possible because the ending position and orientation of the viewpoint was the same for every trial during both movement types. As a result, a small random offset around the vertical axis was added to the final orientation of the viewpoint for each trial. This offset eliminated the work around, making as the horizontal position of the target on the screen no longer correlate exactly with its position in the 3D environment. The magnitude of the offset could range from 0-14 degrees to the right of the final orientation, with a random magnitude being chosen every trial (see figure). The As a result, a target located at the arm’s exact reach boundary could be located anywhere within the central \( \frac{1}{3} \) of the computer screen. While the work around only affected the exo-centric and static movements, the offset was added to the ego-centric movement as well in order to prevent any biasing due to the random nature of the offset.
Visual Environment

Experiment 3 utilized two of the three visual conditions featured in the Staircase and Constant Stimuli data-collects, the high and the low. The Medium condition was removed because experiments 1 and 2 found no significant differences in performance between the medium and the high, nor between the medium and the low condition. In addition to performance of the medium visual environment in the previous data collects practical constraints (e.g. time per trial, number of trials per condition, number of viewpoint movements to investigate, maximum time of single collection session) required one of the visual conditions to be dropped in order to keep the maximum time individuals spent in the lab to a reasonable level. Based on reports from participants in the first two data collects, it was decided that participants who were required to participate for more than two hours ran an unacceptable risk of losing their focus on the tedious reachability judgement task.

Procedure

The procedure for the third data-collect was a modified version of the one used in chapter 2. These modifications were made based on lessons learned during chapter 2 and due to the large number of experimental conditions. As in the previous data collects, each participant signed informed consent documents, and were given a verbal briefing before starting. The experiment briefing covered the same information as the briefings in the previous data collects, adding a description of the four motion types to be used in the data-collect. In response to several participants expressing misconceptions during pilot
testing, the experimental briefing also stressed that targets placed extremely close to robot were still considered reachable. Previously some participants had mistakenly believed that the manipulator arm did not have the dexterity to objects so close to the platform’s chassis.

Figure 19 – The structure of experiment 3. Unlike experiments 1 and 2, the training phase was only administered at the beginning of the experiment session. Each of the eight condition were made up of a Practice and Data Collection phase. Each data collection phase was made of 150 trial. 20 of these trials were catch trials, designed to interfere with the participant recognizing any pattern in the placement of the target object. The other 130 trials contributed towards the psychometric function parameter estimation.

Due to the large number of experimental condition compared to previous data collects, the order in which participants experienced different experiment phases had to be modified from the previous experiments. During sessions in experiment 3 each
participant experienced the familiarization phase only once (see Figure 19). As described in Chapter 2, the familiarization phase gave participants manual control of the manipulator arm, via a set of graphical controls, and asked participants to move the arm in order to hit the target sphere resting on the ground. This task helped familiarize the participant with the capabilities of the manipulator arm, to see it in different configurations, and to force them to reflect on how it operates. After familiarization, participants experienced the eight experimental conditions in a randomized order. For each experimental condition, participants experienced a training phase and data collection phase. Like the previous data collects, both the practice and the data collection asked participants to perform the reachability judgement task, however the practice phase gave them feedback about whether their answer was correct.

For each trial in the practice and data collection phases, the participant was first presented with a two second stimulus mask. The mask was made up of a gray screen and a large white count down in the center. During data collection, this screen also included a green progress bar showing the number of trials remaining in the condition. After two seconds the gray screen was replaced by the feed from the camera and the two buttons representing reachable and unreachable. This stimulus was shown for two seconds before the camera feed was disabled, turning the screen black. The program waited for the participants to press one of the buttons before continuing on to the next trial. While the stimulus was present on screen, any viewpoint motion took 1.6 seconds (see Figure 16). The viewpoint remained motionless for .2 seconds before and after each movement. This
additional time was added to minimize participant disorientation at the beginning of the trial. As discussed in the previous section, the acceleration of each movement was tuned so that the viewpoint remained very slow moving near the beginning and ending of the movement. This was done to maximize the amount of time participants had to view the target object.

Between phases (e.g. familiarization, practice, data collection) and between experimental conditions participants could take optional rest breaks. While on break a 5-minute timer appeared on the screen. During this time participants were encouraged to get up and walk around, look at their phone, or go to the bathroom. The study resumed when they pressed the continue button. Participant had been told that breaks were optional during the experiment briefing, being advised that if they decided to take all break their total time in the lab would be closer to 2.5 hours.

After participants completed all motion path and visual environment combinations, they were asked to fill out a paper questionnaire with 10 questions (the complete form in shown in Appendix 1). This questionnaire was added for experiment 3 and solicited information on the participants’ demographics, performance calibration, understanding of the movement types, level of distraction, movement preferences, and previous experience with both robotics and virtual environments. The responses given on the questionnaire were analyzed for broad themes. Questions 1-4 asked for demographic information including age, gender, major, and school rank. Question 5 consisted of 16 statements for
participants to agree or disagree with. Participants were asked to answer on a scale of 1 (“Definitely false”) through 5 (“Definitely true”). These questions asked questions designed to probe the participant’s calibration to their own performance and to gauge whether the participant may have become less attentive over the course of the study. The calibration probe included statements like “I usually understood where the target was positioned,” “I felt like later trials were easier than earlier trials,” and “I felt frustrated.” The engagement probe made statements like “I felt bored,” “This study lost my attention part way through,” and “I had fun.” With both probes an effort was made to make similar statements but invert the answer (e.g. “I felt bored,” “I had fun”). These question pairs were included so that the agreement between the statements could be assessed. Questions 5 and 6 asked participants to rank their preferences in both the visual condition and the motion type. Question 7 asked participants to describe the movement the viewpoint made during the pseudo-arm condition. This question was added to check whether the participants actually understood the movement they were seeing. Questions 8 and 9 asked participants to rate how much previous experience they had with virtual environments and robotics. Participants were asked to answer on a scale of 1 (“No experience”) to 5 (“Daily experience”). Additionally, participants were asked to provide more detail about their previous experience. Question 10 briefly explained psychometric functions and their parameters, including several figures for reference. Participants were asked to draw their own psychometric function. This question was meant as a gauge of the participants’ calibration. After completing the questionnaire, participants signed a payment receipt and received their compensation.
Results

Mirroring the approach from chapter 2, each participant’s data set in each condition was fit with a psychometric function using a maximum likelihood technique. The resulting function described the perceived threshold between reachable and unreachable targets, the reliability of the reachability judgements ($\alpha$ and $\beta$ respectively). The $\alpha$ parameter represents the intercept of a participant’s psychometric function and the $\beta$ parameter represents the slope of that function.

Using the statistical program JMP Pro 12, the psychometric parameter data was analyzed using a mixed effects model, using Restricted Maximum Likelihood to estimate model parameters, in order to look for significant differences in participant performance (West, Welch, & Galecki, 2014). The mixed effects model had the form:

$$ Y_{ijk} = \rho_i + E_j + M_k + (EM)_{ij} + (\rho E)_{ij} + (\rho M)_{ik} + \varepsilon_{ijk} $$

The model included three within subjects fixed effects: the visual environment ($E$), the motion type ($M$), and the cross between the two ($EM$). The model also included three random effects: participants ($\rho$), participants crossed with visual environment ($\rho E$), and participants crossed with motion type ($\rho M$). Lastly the model included an error term ($\varepsilon$). The use of the mixed effects model was chosen because it does not assume sphericity or normality of the data. Modeling without these assumptions is important due to both independent variables being within subjects, as observing the same participants across all
conditions produces data correlations across conditions. Tukey’s HSD post-hoc analyses were performed in order to investigate pairwise differences for the explanatory variables. In addition, the mixed effects analysis allows one to explicitly model the individual differences in participants’ performance. This last feature is especially important to the current analysis, as the data contains large individual differences between participants. The mixed model analysis found that individual differences accounted for 44% of the variance in both the alpha and beta parameter estimates. Explicitly modeling participant differences minimizes the need to transform the data in order to minimize these effects.

A large number of outliers were found whose McFadden’s values fell below 0.2. In addition, inspection of individual participant’s data showed unexpected patterns of yes/no answers. The pattern of yes’s and no’s in many of these participant’s data sets suggested that their performance had caused the UML sampling algorithm to “pin” against the maximum or minimum distance within its sampling range. The proportion of incorrect answers at these boundaries (e.g. the proportions of reachable judgements at a distance of 2.13 arm lengths, or the proportion of unreachable judgements at a distance of .54 arm lengths) was strongly negatively correlated with McFadden’s pseudo-$r^2$ value ($r^2 = 0.75$, after two outliers removed). Based on the correlation between these two pieces of data, it was likely that these instances of pinning biased the psychometric parameter estimates. Participants who pinned only did so in one to two of the eight conditions. This pattern suggests that they were attempting to participate in earnest, but their performance in some conditions was so poor that it exceeded the apparatus’ ability to respond. In order to
minimize the effects of the pinning, all observations with a pseudo-$r^2$ below 0.2 were dropped from the analysis.

The $\alpha$ parameter of the psychometric function was evaluated with a mixed effects model ($r^2 = .86$). The analysis found the main effect for motion type significant, (F(3,71.64) = 4.259, p < .008). The main effect of visual environment was not significant, (F(1,28.82) = 2.7686, p = .107). A significant interaction effect between visual environment and motion type was found, F(3,61.1) = 3.142, p = .0315. A Tukey’s HSD was performed on the four motion types. The motion types had least squares means of 56.63 (SE = 2.646), 62.19 (SE = 2.631), 64.23 (SE = 2.321), and 66.279 (SE = 2.51) for pseudo-arm, ego-centric, exo-centric, and static, respectively (see Figure 20). Note that the optimal $\alpha$ parameter in the psychometric function corresponds with the maximum reach of the manipulator arm of 46.62. The Tukey’s HSD showed that the pseudo-arm motion was significantly different than the exo-centric and static condition (p = .05 level). However, pseudo-arm was not significantly different than the ego-centric condition.
The estimated values for the \( \beta \) psychometric parameter were also compared using a mixed effects analysis (\( r^2 = .70 \)). The analysis found main effects for both environmental condition (\( F(1, 25.93) = 27.886, p = .0001 \)) and viewpoint motion type (\( F(3, 63.817) = 3.817, p = 0.0141 \)). The interaction between the two factors was not significant. The two environments had least squares means of \(-0.1617 (SE = 0.011)\) and \(-0.1077 (SE = 0.012)\) for the rich and sparse conditions respectively. The four motion conditions had least squares means of \(-0.1483 (SE = 0.014), -0.1287 (SE = 0.014), -0.1507 (SE = 0.012), \) and \(-0.1109 (SE = 0.013)\), for the pseudo-arm, ego-centric, exo-centric, and static conditions respectively (see Figure 21). A Tukey’s HSD showed that the exo-centric condition was significantly different from the static motion (p = .05 level).
The results of the mixed effects analyses seemed to suggest that the motion conditions that moved the viewpoint in order to provide a second perspective onto the target object might perform better than those that remained in the same location. To investigate this a second set of analyses was conducted to investigate performance differences between the motion conditions that moved the viewpoint and those that did not. This was accomplished by grouping the pseudo-arm and exo-centric motion type data together, and grouping the static and ego-centric data together. A second set of Linear Mixed Effect analyses were run on the resulting data for the $\alpha$ and $\beta$ parameters. The results of this second set of analyses are described below.
Figure 22 – Least squares means for $\alpha$ across grouped motion conditions. The horizontal black bar represents the manipulator arm’s maximum reach. * denotes significant differences at $p < .05$.

The mixed effects analysis run on the $\alpha$ parameter data showed that individual participant differences accounted for 46% of the variance in the $\alpha$ parameter ($r^2 = .30$). The analysis no significant main effects or interactions (see Figure 22).
The mixed effects analysis run on the $\beta$ parameter data showed 43% of the variance was attributable to individual participant differences ($r^2 = .68$). The analysis found main effects for both environment condition, $F(1, 25.65) = 28.67, p = 0.0001$, and the motion group, $F(1, 25.41) = 8.525, p = 0.0072$, but no significant interaction between the two.

The two environments had least squares means of $-0.161 (SE = 0.011)$ and $-0.107 (SE = 0.012)$ for rich and sparse respectively. The two motion type groups had least squares means of $-0.149 (SE = 0.011)$ and $-0.118 (SE = 0.0125)$ for pseudo-arm/exo-centric and ego-centric/static respectively (see Figure 23).

The responses given on the exit questionnaire were analyzed to look for broad themes. According to the exit questionnaire, the majority of participants came from an engineering or computer science background ($n = 19$), however many different backgrounds were represented including accounting, design, political science, law,
psychology, and medicine. Participants also indicated that they had varying levels of experience with robotics or virtual environments, like video games or computer aided design software. Participants were given a five point Likert scale to rate their previous experience with robotics and a second for their experience virtual environment, with 1 representing no experience and 5 daily experience. 13.33% of participants rated their past experience with robotics as a 4 or a 5, which would indicate daily or nearly daily experience with robotics (see Figure 24). When asked about their previous experience with virtual environments (e.g. video games), 33.33% of participants rated themselves as a 4 or 5. Lastly participants reported using software featuring virtual environments on average 4.75 hours per week (s = 6.93).

Figure 24 – Distribution of reported previous experience with virtual environments (e.g. video games) and robotics.
A stepwise regression was performed on the data to test whether any performance could be attributed to demographic differences. Gender, age, previous robotics experience, and previous virtual environment experience were all included in the stepwise analysis. The analysis found no significant contributions to performance from any of the four demographic classifiers.

A second analysis was done on how well the three estimated psychometric functions fit the underlying data. A mixed effects model was used with sampling type and environment condition as fixed effects, and participants as random effects. This analysis dropped the medium condition from the analysis in order to compare data from experiments 1 and 2 to experiment 3. The resulting model found that individual differences between participants only accounted for 6.8% of the variance ($r^2 = 0.85$). The model found main effects for both sampling type ($F(2,36.51) = 17.43, p = .0001$) and environment ($F(1,23.26) = 59.1281, p = .0001$). The model also found an interaction effect between environment and sampling type ($F(2,32.25) = 17.164, p = .0001$). A Tukey’s HSD found that the Staircase and Constant Stimuli experiments had significantly better fit data than the UML ($p = .05$ level). Constant Stimuli, Staircase, and UML had least squares means of $0.419$ (SE = 0.024), $0.421$ (SE = 0.024), and $0.269$ (SE = 0.018), respectively. An additional Tukey’s HSD on the interaction data found that both rich environments in the constant stimuli sampling method and staircase sampling method were significantly different than all other conditions ($p = .05$ level). However, those two conditions were not significantly different from one another. The Constant Stimuli/Rich
condition had a least squares mean fit of 0.55 (SE = 0.028) and Staircase/Rich condition had a LSM fit of 0.47 (0.028). The Constant Stimuli/Sparse, Staircase/Sparse, UML/Sparse, and UML/Rich had LSM fits of 0.28 (SE = 0.028), 0.36 (SE = 0.028), 0.25 (SE = .025), and .28 (SE = .020), respectively.
Chapter 4: Discussion

The experiments reported in this dissertation provide the first empirical evidence that human sensor systems behave as perceptual systems, and that the ability to take new perspectives on a scene of interest creates significantly better perceptual performance. As early as 2002, platform operators reported difficulty understanding the physical relationship between the remote sensor platform and its surrounding environment when operating outside line of sight (Casper, n.d.). Many research communities have developed metrics to benchmark performance with these systems and offered candidate solutions to improve performance. However some 15 years later evidence suggests that similar problems continue to plague the operation of these systems (Guizzo, 2011; Morison et al., 2015). Many of the metrics proposed to describe the performance while using remote sensor systems provide task level descriptions of how an operator performed while using the platform. Examples include time on task, errors committed, targets found, slips committed, number of interfaces clicks, cognitive workload, etc. While these descriptive metrics thoroughly catalog how well an operator performed, they do not provide strong guidance for new platform designs nor do they embody a strong model of system design. Creating new sensor platform designs to overcome the challenges consistently reported by practitioners in the field will require a new approach to measuring and explaining system performance.
Focusing research efforts on the interaction between human operator and the sensor platform shows promise for creating prescriptive models of how to design sensor platforms. The first of these models considers the human sensor system as a perceptual system and can be traced back to early work describing the need to consider operators’ perceptual capabilities when designing sensor platforms (Tittle et al., 2002). This work itself has roots in early work on graphical computer interfaces and how to support users’ ability to easily explore large amounts of information (Woods, 1984). More recent work has begun to find empirical evidence to support this prescriptive model for platform design by measuring how well these systems can directly perceive remote environments (T. Murphy, 2013; Schmidlin & Jones, 2010). To date the studies have investigated how well the systems can perceive affordances like passability and reachability in the remote environment. Many variations on these studies have been run, and can be split into themes such as operator learning (Schmidlin, 2014), platforms’ physical characteristics, line-of-sight versus teleoperation (K. S. Moore et al., 2009), and perceiving passability versus drivability (Jones et al., 2011). These previous studies form a solid foundation; however, many gaps remain to be addressed in this topic.

This dissertation sought to address two of the gaps in the human sensor system perception literature. The first gap was to verify the applicability of perceptual models to human sensors systems. Several previous works have made the argument that human sensor systems are perceptual systems and operate by the same rules. The empirical
investigations that build off of this argument have not yet demonstrated that the application of psychometric methods to human sensors system is warranted. Chapter 1 of this dissertation sought to address that gap by analyzing how well perceptual judgments, made by human sensor systems, could be fit by standard psychometric functions. This was accomplished by analyzing the data from two similar experiments in which participants judged the reachability of virtual targets.

The second gap in the literature is the effect of viewpoint motion on a human sensor system’s ability to directly perceive a remote environment. To date, all of the work done on human sensor systems has utilized viewpoint that cannot move relative to the robot’s chassis. However, many commercially available platforms rely on both dynamic and static viewpoints. Dynamic viewpoints are important because they allow operators to view the environment from multiple perspectives and are a critical component of ecological perception (J. J. Gibson et al., 1957; J. J. Gibson, Olum, & Rosenblatt, 1955). The experiment in chapter 3 of this dissertation sought to provide a first investigation of viewpoint motion in human sensor systems. This experiment compared types of motion chosen based on their availability on current platforms and what contribution they made to the human sensor system’s ability to perceive reachability. Using a modified version of the experimental methods established in chapter 2 the reliability, reachability threshold, and lapse-rate of participants’ reachability perception was compared across the experimental conditions.
Modeling Human Sensor Systems as Perceptual Systems

The results from experiments 1 and 2 showed that the perceptual performance of the human and machine operating together is well modeled using standard psychophysical measures. This finding bolsters the claim made in previous theoretical work about the human operator and sensor platform operating together as a larger perceptual system. This work claims that the human sensor systems operate, and therefore can be measured, similarly to humans perceiving the environment around them. is under the same constraints and benefits from the same strengths as a human’s own perceptual system. Verification of these claims required that the same models used to describe human perceptual performance be well suited to describing human sensor system performance. The consistently well fit psychometric functions found in the results of experiments 1 and 2 are strong evidence that the human sensor system does in fact act as a perceptual system.

The two sampling methods used in experiments 1 and 2 show a large difference in the range of fit values, described by the McFadden’s $r^2$ metric. The range of fit values in using the staircase sampling method are more consistent than the fit values resulting from using the method of constant stimulus. The only conditions that produced psychometric functions that fell below McFadden’s definition of “excellent fit” ($r^2 < 0.2$) were in experiment 2. There are at least two possible explanations of this difference in ranges, both of which are possibly related. First is the simplest explanation, experiment 1 included many more trials per participant per condition than experiment 2. The staircase
algorithm used a performance based stopping rule, which ended the condition only after certain performance criteria were met. Whereas the method constant stimuli produced a fixed number of trials per participant per condition. In the end, experiment 1 included almost twice the number of trials per participant as experiment 1. This large increase in trials could help explain the tighter range of fit values found in experiment 2. The second explanation of the difference in fit ranges is the sampling method used. The method of constant stimuli and the staircase distribute samples differently. The method of constant stimuli samples the entire range of stimulus levels an equal number of times, whereas the staircase methods attempt to settle on the participant’s perceived threshold. These sampling methods have the potential to under sample or over sample part of the psychometric function. The most likely explanation of the different fit ranges is a mixture of the two proposed solutions. This should inform the design of future studies.

A second interesting trend in the fit data from experiments 1 and 2 are the differences between the three visual environments. The results from both experiments found significant differences between the fit values of the sparse and rich conditions, but no significant differences with the medium condition. The spread of the fit values in the medium environment was much larger than in those in the other two environments. The performance in the medium environment is similar to the results in (T. Murphy & Morison, 2016), which reported the outcome of experiment 1 in terms of the estimated psychometric parameters, rather than the fit of the psychometric function. Murphy et al. (2016) found that the estimated psychometric parameters for the medium environment
were not significantly different than the other two environments but the range of
parameter estimates was much larger. This distribution of performance was unexpected,
as there had been an assumption when designing the environment that increasing the
amount of visual information would improve the participants’ performance. The results
from experiment 1 and 2 do not support this. One possible explanation of the large range
in performance in the medium condition has to do with learning effects in the
participants. The key difference between the low and medium condition is the addition of
the ground plane. The relationship between target and the ground plane is left largely
ambiguous because there is not frame of reference to directly compare the two, given that
the target’s altitude and diameter were randomly changing. It was possible to develop an
understanding of the relationship between the two either by some sort of intuition, or by
exposure to the high condition first. The addition of the shadow in the high condition
directly linked the target and the ground plane. The shadow showed the participant both
the relative size of the target and the target’s position relative to the plane. It is possible
that participant exposed to the high condition before the medium were able to better
predict the relationship between the target and the ground plane. It is reasonable to
believe that the participants who were able to develop this understanding of the
relationship between target and ground plane, either through intuition or by exposure to
the high condition, would have much stronger performance than those who did not.

The second analysis of data fit from chapter 3 showed that the method of constant stimuli
and the staircase had significantly better fits than the UML. However, this is most likely
due to experiment 3 featuring half the trials per condition compared to constant stimuli in experiment 2 and one-quarter of the trials compared to the staircase method in experiment 1. The data sets resulting from the UML sampling method still averaged above the threshold of excellent fit set by McFadden (1978). This second analysis also showed performance data from rich environments resulted in significantly better fit than those from the Sparse condition.

Effects of Viewpoint Motion and Visual Environment

The results indicate that moving the viewpoint to a new, and useful, perspective helped participants have more accurate thresholds and more reliable judgements of the target’s reachability. This finding makes sense as viewpoint motion is a critical part of human perceptual capabilities. If human sensor systems perform as perceptual systems, viewpoint movement should also be a critical component of the human sensor systems perceptual capabilities. The results from chapter three seem to support this conclusion. The second set of mixed effect analyses in the results of chapter 3 looked for differences in the performance of the motions that moved the viewpoint to a new perspective and those that did not. These results constitute the first empirical support of Roesler (2005), and Morison (2010)’s Perspective Control model. Perspective Control lays out one possible viewpoint control metaphor that maximizes the visual information generated from motion and simplifies the process of generating useful perspectives onto a scene of interest. The starting and ending perspectives used in the exo-centric and pseudo-arm conditions were generated using movements described by Perspective Control.
The significant performance differences between the two motion groups are accounted for by the ability to shift perspectives. The perspective offered by the “over the shoulder” view of the platform’s camera is close to the worst perspective for judging the depth of an object in the view of view (see Figure 25). The closer the viewpoint is to being directly behind the manipulator arm, relative to the target, the more the visual scene is ambiguous, up to a stretching or shearing transformation in depth (Todd, 2004). Stated differently, without additional source of information to judge the relative size of an object the perspective directly behind the arm, or nearly behind the arm, provides virtually no information about the depth of an object. With this in mind it is clear that the ego-centric movement, while useful for looking at the environment around the platform (e.g. directly behind the platform), it is not useful in resolving the depth ambiguity of due to the perspective’s location. Almost any change in viewpoint position should begin to provide more information about the depth of an object. Conversely the perspective that should provide the most information with which to judge the reachability of the target is one that looks perpendicular to the line drawn between the platform and the target object, is located halfway between the target and the chassis in depth, and is located to one side of the line. This perspective should resolve all ambiguity about the distance of the object from the platform and show the relationship between the length of the manipulator arm and the target, maximizing reachability judgement performance.
While the results from the second set of mixed effect analyses showed straightforward effects of grouped motion types, the results from the first set of analyses present a more nuanced picture of motion type performance. Pairwise comparisons found that the pseudo-arm motion produced significantly better threshold accuracy than static and ego-centric movement types, and that exo-centric movement produced significantly better reliability than the static condition (Figure 25 shows threshold for the pseudo-arm and static movements). These results are surprising, as adding the visual information from motion to a perspective shift was expected to improve performance.

The lack of conclusive performance differences between the exo-centric and pseudo-arm conditions could stem from at least two possible sources. Firstly, it is possible that the information from motion did not significantly contribute to participants’ ability to perceive reachability. This seems less likely given previous work on phenomena like motion parallax (Rogers, Rogers, Graham, & Graham, 1979) and optic flow (Koenderink, 1986; Lee, 1980). Another possibility for the lack of significant performance differences between exo-centric and pseudo-arm is that the information provided by the perspective shift eclipsed the information provided by the information from motion. The original hypothesis was that both information from motion and a perspective shift would have a cumulative effect, the presence of both would results in proportionally more accurate thresholds and more reliable reachability judgments. However, this would not appear to be the case. Based on the results of the second mixed effect analyses it would appear that the shift in perspective contributed much more to reachability perception than the
information from motion. Based on this finding, it is tempting to conclude that the pseudo-arm cameras utilized by current generation platforms matches the performance one could expect from a platform built around the concept of Perspective control. However, the generalizability of the finding of experiment 3 to field-able platforms is doubtful. The implementation of the arm condition in experiment 3 was primarily based on making fair comparisons to the other movement conditions, rather than maximizing its external validity. The limitation in the design of the pseudo-arm movement is discussed in detail in the next section.

Figure 25 – Average perceived threshold for the pseudo-arm and static motion types. Four target objects are shown in the top panel and represent the physical reach threshold, the average perceived threshold for pseudo-arm, the average perceived threshold for static, and the maximum sampling boundary. The bottom four panels show each target object from the standard over-the-shoulder perspective used in all motion types except pseudo-arm.
One other consideration affecting the lack of conclusive differences between the exo-centric and pseudo-arm condition is the use of a static target object. The static target most likely increased the usefulness of the pseudo-arm condition relative to the exo-centric movement. Participants using the pseudo-arm movement would find it difficult to detect small, or even moderate, target movements as the viewpoint looks away from the target during the movement. The continuous view of the target provided by exo-centric movement should allow participants to easily detect target movements.

The other finding from the first set of mixed effects models was an interaction effect in the threshold accuracy data, between motion type and environment type. The interaction effect means that some combinations of environment type and movement type has a significant effect on the variance of estimated reachability thresholds. The most likely explanation of this interaction effect is that some motion types are more sensitive to the richness of the visual environment. Those motion types that do not provide useful information as a result of moving (e.g. ego-centric and static) should result in worse performance when the visual environment is sparse and improved performance when the environment is rich. Motions that provide additional information, such as visual information from motion or a new perspective, should be less sensitive to outside information. The findings support this claim. The sparse/static condition showed worse performance than the rich/static condition. The sparse/ego-centric condition performed only slightly worse than the rich/ego-centric condition. Both the sparse/exo-centric and
sparse/pseudo-arm conditions performed roughly the same as the corresponding rich conditions.

While experiment 3 did not find significant performance differences between the exo-centric motion and the pseudo-arm motion, the confirmation that the shifts in perspective do lead to better performance seems to support the Perspective Control model. Additionally, the lack of difference between exo-centric motion and the pseudo-arm motion can partially be explained by the design of the pseudo-arm condition (described in detail in the next section). The perspectives for the pseudo-arm motion were selected from those generated by the exo-centric motion in order to make a fairer comparison. One of the core strengths of the Perspective Control model is its ability to generate useful perspectives. The movement described in Perspective Control, moving the viewpoint along the surface of a sphere, constrains the possible perspectives the viewpoint could take. These perspectives are more likely to be useful to the operator. This is in contrast to the current control metaphors, which provide constraints based on manipulator arm servos. This leaves a much larger set of possible views, most of which provide no utility to the user.

A second strength of Perspective Control is the ease in which users can move the viewpoint to useful perspectives. Moving the viewpoint to more useful perspective, or even exploring the space of possible perspectives, using the spherical coordinate system involves changing 3 parameters similar to latitude, longitude, and altitude ($\varphi$, $\theta$, and $r$
respectively) in a geographic coordinate system. This stands in contrast with the control current manipulator arms use, which forces the user to define individual servo positions in order to move the perspective. This process is complex and requires concentration on the user’s part. Some platform designs do offer users inverse kinematic control of the manipulator arm’s end effector. This allows the users to define the position and orientation of the manipulator arm directly, and uses the on-board computer to calculate the required servo positions. While this is most likely a dramatic improvement over individual servo control, it does still require the user to specify several more parameters compared to Perspective Control.

Limitations

Experiment 3 had several limitations. The primarily limitation was the implementation of the pseudo-arm motion in experiment 3. The pseudo-arm movement was a compromise between several constraints in the study design. The final implementation of the motion has weak external validity and may explain the relatively high performance compared to the exo-centric motion type. The weak external validity was the result of trying to design a fair comparison between the four motion types. The first limitation was the duration of the movement. In designing the four motions, it was clear that all four should have the same stimulus exposure time. Allowing one type of motion to be viewed for longer would run a risk of biasing the outcome. The decision of stimulus exposure time was constrained by the work of Heft (1993), which capped exposure time at three seconds. The other constraint on stimulus exposure time was the attempt to maximize the number
of trials each participant completed, while hedging against them becoming distracted. As a result of these constraints, the stimulus exposure time was set to two seconds and thus all motions must be two seconds or less. While this is a reasonable amount of time for an ego-centric viewpoint to rotate on a remote platform, the complex servo motions required to move the arm on a real platform would take significantly longer. Trying to achieve the starting and ending location on most current generation platforms would take a significant amount of time, even on the order of minutes to align all the servos. During this whole transition, the arm mounted cameras would rarely, if ever, be orientated at the target. Operators would have to actively hold the original image in memory, while also conducting a complex set of movements. The real-world task is significantly different than the quick succession of views presented in this experiment. The decision to use a two second exposure time also required that the arm undergo compound movements, movements of more than one servo at a time. These compound movements significantly smoothed the motion of the viewpoint compared to what operators would experience in real life. This difference in timing, and the resulting changes to the arm-movement could easily explain the pseudo-arm condition performing as well as the exo-centric condition.

It is important to note that operators moving real manipulator arms must perform trial and error guess work in order to move the viewpoint to a new perspective. This is a more difficult task than what was presented to participants in this experiment.

The second limitation of the pseudo-arm condition was the perspective it took at the end of the movement. During the design of the motions it was decided that ending perspective
should be the same between the exo-centric and pseudo-arm. Having different end perspectives would impact participants’ ability to perceive reachability. However, almost no current generation EOD robots used today has the degrees of freedom to take the ending perspective offered by the exo-centric movement. As a result, the Unity3D program had to give the manipulator arm two additional degrees of freedom.

The effect of these two compromises on the pseudo-arm movement had the effect of making it into essentially another exo-centric viewpoint controller (Morison, 2010; Roesler, 2005). As a result, it is difficult to draw empirical conclusions about the performance of pseudo-arm viewpoints compared to exo-centric viewpoints, as was one of the original goals. The positive effect of the compromises was testing the contribution of the visual information to direct perception. Because both pseudo-arm and exo-centric movements end in the same locations, the major difference between the two motions is whether the viewpoint remains oriented towards the target during the movement.

A second limitation of experiment 3 was the implementation of the Updated Maximum Likelihood algorithm. As discussed in chapter 3, the UML algorithm has several parameters that must determined prior to data collection. Included in these are the range of possible stimulus levels, and the possible range of each psychometric function parameter. The UML algorithm for experiment 3 used values based on the data gathered from experiments 1 and 2. The UML algorithm was given a range of possible stimulus level from 25 distance units (0.53 manipulator arm lengths) to 100 distance units (2.14
manipulator arm lengths). This range was a reasonable selection based off of the available data, however the results from experiment 3 show many instances of participants whose performance would result in sampling distances much further away than 100 units. One discrepancy effecting the setup up of the UML algorithm was the parameterization of the psychometric function used. Experiments 1 and 2 used a 2-parameter model whereas experiment 3 used a 4-parameter model. The inclusion of the two additional parameters would results in different estimates of threshold and slope. As a result, the data used to set UML parameters, like the prior probability distributions for the $\alpha$ and $\beta$ parameter, was different what the UML algorithm would estimate. However, this discrepancy should not be of not serious concern, as the creators of the UML algorithm noted that it is robust against poor set up (Shen et al., 2015).

Another potential limitation of this study is that participants likely did not understand the arm motions before data collection. Prior to data collection for each condition, participants were given some practice at the judgement task, using the new viewpoint motion. This practice included feedback and was designed to help the participant become familiar with the viewpoint motion before data collection. However, based on participants’ responses during the exit questionnaire it is unclear whether participants actually understood how the viewpoint was moving in the pseudo-arm condition. Most participants provided descriptions of the pseudo-arm movement that were completely incorrect. If participant did not understand how the viewpoint moved, that lack of understanding would likely effect their judgement performance. In future study designs in
might be advisable to give participants an external view of the platform and show the platform’s viewpoint moving. Showing this external view during the practice phase before data collection in each condition should help participants better interpret what they are seeing while looking through the platform’s viewpoint. Additionally, future study design should consider performance based stopping conditions for the practice phase before data collection. A performance based stopping condition would help to better screen for the worst performing participants and potentially reduce individual differences between participants.

Contributions

The results of this dissertation have several potential benefits to the designing and benchmarking of both current and future sensor platforms. Tracing the design of remote sensor platforms used for EOD and USAR from 2001 to 2017, several key aspects to the onboard sensors systems have changed little since their initial design (see Figure 2). The sensors are in roughly the same positions and the degrees of freedom those sensors have to move are largely unchanged. The insight that human sensor systems act as perceptual systems opens up the possibility for entirely new directions for platform design.

Roesler (Roesler, 2005) and Morison (Morison, 2010) lay out one prescriptive model to guide platform design in their work on Perspective Control. Platforms designed through the lens of Perspective Control will emphasize the ability to quickly and easily take new and meaningful perspectives on the scene of interest. Following the constraints laid out
by Perspective Control, future platform designs should incorporate both the physical
degrees of freedom to achieve this as well as appropriate operator controls that minimize
the cognitive workload. Well-designed sensor platforms built to facilitate perception are
likely to realize improved task level performance, such as reduced operator cognitive
workload, shorter time on task, and fewer errors committed, among other benefits.
However, in order to create a well-designed system, one must be able to differentiate
between designs’ performance.

The classification of human sensor systems as a perceptual system opens up a new class
of diagnostic testing that can be performed to identify performance bottlenecks and create
new solutions to them. As mentioned previously, many current metrics are descriptive,
cataloging the performance of remote sensor platforms at a task level (Goodrich &
Schultz, 2007). These descriptive metrics seem to have had little impact on advancing the
design as sensor platforms continue to suffer from the challenges. Investigation of human
sensor system perception can begin to provide useful feedback to platform designers.

The methodology developed in this work can also be used as a tool to screen candidate
platform operators for their ability to understand remote environments. The results of the
experiments presented in this work showed high individual differences in perceptual
performance. While these differences were accounted for by the mixed effects model, the
raw data could be useful to identify which candidate operators would perform well or
which would perform poorly. In this scenario, less precise performance measurement is
needed to differentiate candidates. As a result, more coarse sampling methods can be used. The staircase sampling method would be ideal for this measurement. Using a pair of interleaved staircases, like the design of experiment 1, would offer quick estimates of the perceived threshold and reliability of judgements for a given participant while avoiding the risk of them learning the pattern of target placement. The staircase method has the advantage of being easy to set up and interpret, however its biggest advantage is that the range over which the algorithm samples does not have to be predetermined. Because the algorithm is not restricted to a sampling range, no assumptions need to be made about the candidate’s expected performance. This will avoid any pinning phenomena like what was described in the results section of chapter 3.

The data from a staircase sampling method could be analyzed in at least two ways. First is the output of the staircase method itself. Individual staircases estimate a single point on a person’s psychometric function. However, this estimate alone does not provide any information on the participant’s judgment reliability, or their threshold placement (see the methods section of chapter 2 for a description of how to use this sampling method). The second data analysis method is to fit the raw data from the staircase sampling strategy with a psychometric function using a Maximum Likelihood regression. This is the method used in the results section of chapter 2. The method has the advantage of directly estimating the parameters of the psychometric function, and can add the lapse-rate parameter to reduce biasing of the threshold and reliability parameters. This analysis
method is more complex but can result in more precise data, as well as a measure of fit for the psychometric function in the form of McFadden’s $r^2$.

**Future Work**

The work presented here opens up several possible directions for future research. The first is to investigate the impact of viewpoint motion on properties other than reachability. Directly perceiving affordances such as passability and drivability require different visual information than reachability. It is possible that directly perceiving other affordances is best achieved with different types of viewpoint motion. Exploring the space of different viewpoint motion with respect to the perception of different affordances could reveal new control metaphors, and help designers create platforms better tailored to specific types of activities.

Testing whether the advantages of the pseudo-arm motion apply to moving compared to stationary targets presents another potential research direction. As mentioned earlier in this chapter, the use of a static target most likely increased the usefulness of the pseudo-arm movement relative to the exo-centric. Changes in target position during a pseudo-arm viewpoint motion would likely be more difficult to detect than during an exo-centric viewpoint motion, as the exo-centric motion rotates the viewpoint to face the target at all times. Explicitly testing this claim could provide useful guidance for platform designers.
Another potential avenue of research is to explore the potential usefulness movements described by perspective control using the spherical coordinate system. While the use of the coordinate system stands on solid theoretical ground, the exact nature of the performance gains it offers remain untested. One interesting study along these lines would be to probe how the magnitude of the viewpoint’s displacement from the initial perspective effects participants’ judgement reliability and threshold accuracy. As described earlier rotating the viewpoint 90° around the φ axis of the coordinate system should result in an ideal view for judging reachability, as long as the viewpoint is far enough away so both the arm and target are in view, but how does performance drop off as that rotation grows smaller? Is there a linear relationship between the maximum rotation and the perceptual performance of the operator? Describing the relationship between parameters of a viewpoint’s motion and the perceptual performance of the platform’s operators could provide designers a powerful tool to help guide the design of future platforms.
References


Casper, J. L. (n.d.). Human-Robot interactions during the robot-assisted urban search and rescue response at the World Trade Center.


114


Morison, A. M. (2010). Perspective Control: Technology to Solve the Multiple Feeds Problem in Sensor Systems. ProQuest Dissertations and Theses, Columbus, OH.


Appendix A: Exit Questionnaire
Exit Questions

1. (Optional) Age ______

2. (Optional) Major ______________________

3. (optional) Year in School (e.g. Sophomore) ________________

4. Please answer the following:

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I felt bored.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Judgement difficulty was different for each movement type.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I believe my answers were accurate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I started drifting off towards the end of each movement type.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I felt frustrated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I understood the environment more quickly later in the study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I felt confused.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I found it difficult to see how far away the target was.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I generally felt good about my performance.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I usually understood where the target was positioned</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I had sufficient training before beginning the data collection.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. This study lost my attention part way through.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. I did not understand the directions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. I was able to understand how the robot arm was positioned.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. I had fun.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. I felt like later trials were easier than earlier trials.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Was one of the visual conditions (i.e. not referring to the types of movement) easier or harder to make judgements in?

   Why was this one easier or harder?

6. Please rank the four types of camera movement (Static, Ego-centric, Arm-Mounted, Exo-centric) from the one you thought made it easiest to judge the distance to the one that made it the hardest:

   1. ______________
   2. ______________
   3. ______________
   4. ______________
7. Please describe, in your own words, how the camera in the Arm-Mounted condition moved (Example: “the camera moved up, then rotated downwards”).

8. How much experience do you have with 3D software like video games, CAD, simulation, etc. (1: No experience – 5: Daily Experience). _____

   About how many hours per week do you use this type of software? _____

   Typically what type of 3D software do you use?


   If you do use robots, what type of systems do you normally interact with?

10. Your performance can be graphed using what researchers call a Psychometric Function. The X axis of this graph is the distance of the target from the robot. The Y axis is the chance of you judging the target Reachable at the given distance. The slope (or spread) of the curve represents how consistent you were in your judgement of the arm’s length from trial to trial and the x value at which you were equally likely to judge something reachable or not reachable (i.e. y=50) represents the accuracy how far you thought the arm could reach.

(Continued on next page)
The picture above represents normal performance. On this curve (the darker line), the value at y = 50 (i.e. the point at which they were equally likely to judge a target reachable or not reachable) is a little offset from the Robot’s actual reach, meaning that the participant thought the arm’s reach was a long than it was. The curve has a moderate slope which means the participant misjudged many targets near what they thought was the arm’s maximum reach.

The curve on the left represents extreme good performance. The slope is very steep, meaning they were very consistent in their judgements, and the x-value at y=50 is very close to the arm’s actual length. The curve on the right represents very poor performance. The slope is very shallow, meaning they were very inconsistent, and the y=50 point is very far from the arm’s actual reach.

Based on the pictures above, and using the figure below, please draw a curve that you think represents your performance.
Appendix B: Informed Consent Form
The Ohio State University Consent to Participate in Research

Study Title: Evaluation of extending perception in human-robot systems
Researcher: Dr. Alexander M Morison, OSU College of Engineering
Sponsor:

This is a consent form for research participation. It contains important information about this study and what to expect if you decide to participate.

Your participation is voluntary.

Please consider the information carefully. Feel free to ask questions before making your decision whether or not to participate. If you decide to participate, you will be asked to sign this form and can receive a copy upon request.

Purpose:

Observations of people using remotely controlled robots has shown that people have difficulty understanding an environment that they can only perceive while looking and acting through a sensor. During one demonstration the expert, human robot operators were unable to consistently pass a cylinder through a similarly sized hole. A new approach to thinking about and designing robotic sensor systems called “Extending Perception” has begun to change the way people view these systems. Extending Perception places emphasis on the robot as a way to extend the human’s natural skill at understanding environments into a remote scene. By doing this, the design of the robot allows the operator to devote time and energy into understanding what they are looking at, rather than trying to pilot the robot. The goal of the research is to verify that Extending Perception is a valid approach to thinking of robotic sensor systems.

The goal of this research is to measure the participants visual understanding of a remote environment, through a simulated robot sensor using a virtual robot simulator (Webots). The research design consists of experiments adapted from perceptual psychology research. Participants will judge things like reachability, passability and the orientation of virtual objects; manipulate virtual objects; and move virtual sensor position all within the virtual environment. The participant will be interacting with the simulation through input devices like a mouse, keyboard and joystick.

Procedures/Tasks:
CONSENT  
Behavioral/Social Science  
IRB Protocol Number:  
IRB Approval date:  
Version:  

If you chose to participate, you will take part in a series of tasks using a game-like environment on a computer. The simulated tasks involve controlling a simulated robot inside a virtual 3-dimensional world. You will be asked to manipulate simulated objects within the virtual environment using a simulated robot arm. In addition, you will be asked to make judgments about spatial relationships such as, between simulated objects, between simulated objects and the simulated robot, and between the simulated robot view and virtual environment. You will perform the tasks using a computer, computer monitor, keyboard, mouse, and 3-dimensional desktop joysticks.

Duration:  
You may leave the study at any time. If you decide to stop participating in the study, there will be no penalty to you, and you will not lose any benefits to which you are otherwise entitled. Your decision will not affect your future relationship with The Ohio State University.

The total time to complete the study is one and a half (1.5) hours.

Risks and Benefits:  
Although the risks associated with participation in this study are minimal, the benefits are potentially significant. The results of this study are applicable to all forms of human-sensor systems, from improving the capability of emergency response robots to improving the performance of robotic-assisted surgery. Moreover, the paradigm shift in human-sensor systems under study would impact scientific understanding of all forms of human-sensor systems from large scale, small numbers such as remotely piloted vehicles to small scale, large numbered devices such as smart phones.

Confidentiality:  
Efforts will be made to keep your study-related information confidential. For the duration of the study personally identifiable information will be stored in a secure area. At the conclusion of the study all data will be de-identified, with all personal information erased. However, there may be circumstances where this information must be released. For example, personal information regarding your participation in this study may be disclosed if required by state law. Also, your records may be reviewed by the following groups (as applicable to the research):

☐ Office for Human Research Protections or other federal, state, or international regulatory agencies;
☐ The Ohio State University Institutional Review Board or Office of Responsible Research Practices;
☐ The sponsor, if any, or agency (including the Food and Drug Administration for FDA-regulated research) supporting the study.

Incentives:  
In appreciation for your participation, we are reimbursing participants for their time at a rate of $20 per hour.

Law considers payments to subjects considered taxable income.
CONSENT
Behavioral/Social Science

IRB Protocol Number:

IRB Approval date:

Version:

Participant Rights:

You may refuse to participate in this study without penalty or loss of benefits to which you are otherwise entitled. If you are a student or employee at Ohio State, your decision will not affect your grades or employment status.

If you choose to participate in the study, you may discontinue participation at any time without penalty or loss of benefits. By signing this form, you do not give up any personal legal rights you may have as a participant in this study.

An Institutional Review Board responsible for human subjects research at The Ohio State University reviewed this research project and found it to be acceptable, according to applicable state and federal regulations and University policies designed to protect the rights and welfare of participants in research.

Contacts and Questions:
For questions, concerns, or complaints about the study you may contact Taylor Murphy; murphy.1018@buckeyemail.osu.edu; (614) 292-1296.

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.

If you are injured as a result of participating in this study or for questions about a study-related injury, you may contact Alex Morison; morison.6@osu.edu.
CONSENT
Behavioral/Social Science

Signing the consent form

I have read (or someone has read to me) this form and I am aware that I am being asked to participate in a research study. I have had the opportunity to ask questions and have had them answered to my satisfaction. I voluntarily agree to participate in this study.

I am not giving up any legal rights by signing this form. I will be given a copy of this form.

Printed name of subject

Signature of subject

Date and time

Printed name of person authorized to consent for subject (when applicable)

Signature of person authorized to consent for subject (when applicable)

AM/PM

Relationship to the subject

Date and time

Investigator/Research Staff

I have explained the research to the participant or his/her representative before requesting the signature(s) above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

Printed name of person obtaining consent

Signature of person obtaining consent

AM/PM

Date and time

Page 4 of 4

Form date: 12/15/05