Simulating Augmented Reality Spatial Accuracy Requirements for Target Acquisition Tasks

John J. Graybeal, Rachel T. T. Nguyen
KINEX, Inc.
Manassas, VA

Todd W. Du Bosq
U.S. Army CCDC C5ISR Center
Night Vision and Electronic Sensors Directorate
Fort Belvoir, VA

ABSTRACT

Augmented reality (AR) technologies are one method of supporting military visual search tasks, such as target identification, recognition, and acquisition. However, whether or not a given AR technology actually improves human performance depends on many factors, including the quality of the display and the quality of the AR information provided to the Soldier. In this paper, we describe current research efforts by the U.S. Army CCDC C5ISR Night Vision and Electronic Sensors Directorate to use simulation to study one aspect of AR information quality: spatial accuracy. Specifically, we examine the level of AR spatial accuracy required to improve human performance as a function of range and the amount of spatial error in the AR symbology. Participants were placed in virtual scenarios and asked to locate and target a single virtual human holding a weapon amongst many unarmed virtual humans. Participants used realistic sensor controls to scan the virtual field of regard and to locate the target. Baseline performance was characterized by having participants locate targets without any AR assistance. In other control trials, participants were guided to the target with perfectly accurate AR symbology, located both on a situational awareness ring and in the operator's field of regard. In imperfect AR trials, participants were guided by AR symbology distorted by fixed amounts of angular error (1°, 2°, 3°, or 4°) between the target, the observer, and the AR symbology. Our results examine the effects of AR spatial accuracy on target acquisition time, comparing imperfect AR to perfect AR and unaided searching. The ultimate goal of our research program is to support product development and virtual prototyping by simulating task-specific and sensor-specific AR accuracy requirements for sensors and head-up displays.

ABOUT THE AUTHORS

John J. Graybeal recently began working as an Engineering Psychologist at the U.S. Army CCDC C5ISR Center Night Vision and Electronic Sensors Directorate (NVESD), where he previously worked as a contractor with KINEX, Inc. He received his Ph.D. in Psychology in 2017 from George Mason University, with a concentration in Cognitive and Behavioral Neuroscience. Since 2017, he has worked in NVESD's Perception Lab, where he evaluates human performance with a wide range of sensor imagery and emerging technologies. He regularly conducts research with Soldiers and other human subjects using visual tasks (e.g., object detection, object identification) with both real and simulated imagery. He is responsible for leading the NVESD Perception Lab's efforts to modernize and refine human testing methodologies, including training practices for vehicle identification skills that support NVESD perception tests. His research also focuses on human performance with augmented reality, and he is responsible for providing design recommendations for and evaluations of augmented reality displays. His other research projects and interests involve sustained attention, situational awareness, effective teaching and training, and studies aimed at improving experimental methodologies.

Rachel T. T. Nguyen is a Psychology doctoral student in the Human Factors and Applied Cognition (HFAC) Program at George Mason University (GMU). She received her M.A. degree in Psychology, HFAC in 2019 from GMU and her B.A. degree in Psychological Science in 2017 from California State University San Marcos. She currently works in the Visual Attention and Cognition Lab at GMU, where she studies visual attention (e.g., pre-attentive processes, object recognition), visual search (e.g. object detection and identification), and perception; she also has research interests in eye movements, augmented and virtual reality, and measurement. Rachel has worked as a Research Assistant at KINEX, INC. with the U.S. Army Night Vision and Electronic Sensors Directorate and is currently a Behavioral Science Intern in the Integrated Adaptive Cyber Defense (IACD) program at the Johns Hopkins University Applied Physics Laboratory, where she works on human factors engineering tasks on trust and automation.

Todd W. Du Bosq is the Field Performance Branch Chief in the Modeling and Simulation Division at the U.S. Army CCDC C5ISR Center Night Vision and Electronics Sensors Directorate. He received his B.S. degree in physics from Stetson University in 2001 and his M.S. and Ph.D. degrees in physics from the University of Central Florida in 2003 and 2007, respectively. Since 2007, Dr. Du Bosq has provided innovative research and guidance to the U.S. Army on the human acquisition performance of infrared and reflective sensors regarding military vehicles, human targets, laser markers, and IEDs for source selections, war games, simulation, training, and system performance trade studies.

Simulating Augmented Reality Spatial Accuracy Requirements for Target Acquisition Tasks

John J. Graybeal, Rachel T. T. Nguyen
KINEX, Inc.
Manassas, VA

Todd W. Du Bosq
U.S. Army CCDC C5ISR Center
Night Vision and Electronic Sensors Directorate
Fort Belvoir, VA

INTRODUCTION

The U.S. Army CCDC C5ISR Center Night Vision and Electronic Sensors Directorate (NVESD) has been a world leader in the development and evaluation of electro-optical and infrared sensors for over sixty years, supporting technology applications ranging from vehicle mounted sensors, to weapon sights, to head-mounted displays. Many technological advancements have made it possible to improve the way we present sensor information to a human operator. As such, augmented reality (AR) technologies have become an important area of research and development for NVESD, as part of the U.S. Army's focus on increasing Soldier lethality.

Augmented reality technologies attempt to enhance human sensory experiences by inserting digital information into the user's experience of the "real world" (Wu, Lee, Chang, & Liang, 2013). While there are many different forms of AR, the present work focuses on the visual overlay of digital information onto the human visual field, either by augmenting a live sensor feed or utilizing a see-through display. There are many ways augmenting a Soldier's visual field might assist with operational tasks. For example, AR symbology marking enemy units might improve target acquisition, AR labels identifying an object might facilitate object recognition, and AR waypoints indicating a navigation route might improve navigation efficiency. While the goal of improving situational awareness and military task performance by giving operators additional information is not particularly novel, the ongoing maturation of see-through and helmet-mounted display technologies, as well as the improvement of a host of supporting technologies (e.g., lightweight computer and graphics processors, information system networks, global positioning systems), has made it progressively possible to provide Soldiers with increasingly complex information in unobtrusive, mobile platforms.

However, while AR technology is extremely promising, potential risks to human performance also deserve careful attention. For example, while AR technologies have the potential to direct the operator's attention towards critical visual information, the AR system may also capture visual attention in undesirable ways, such that operators become inattentive to other critical information in the "real world" (Dixon et al., 2013; Radu, 2014; Tang, Owen, Biocca, & Mou, 2003). Likewise, while AR has the potential to decrease the cognitive load of operators by reducing the amount of information that must be kept in their memory (i.e., because that information is present on the display), increased cognitive load with AR systems has also been reported in the literature (Dunleavy & Dede, 2014). Thus, AR must be implemented carefully to avoid major risks to military personnel, or such technologies may do more harm than good.

One critical area of this research challenges a common assumption for those developing an AR system: that AR information provided to the Soldiers will be useful and will help improve Soldier performance. NVESD is interested in understanding the "quality" of AR information provided to Soldiers. The quality of AR information has many dimensions, including its relevance, timeliness, perceptual saliency, communicative clarity, and its accuracy. All of these dimensions are worthy of research to define the necessary quality of information needed to improve human performance.

The current research focuses on a single issue related to AR information quality: detriments to human performance caused by inaccurate information. As with any technology, AR systems will not be able to provide operators with perfect information at all times. AR errors may be caused by limitations of the AR and/or display technology itself (e.g., poor spatial accuracy for displayed symbology), or by receiving imperfect information from a human or computer source and subsequently relaying that inaccurate information to the human operator (e.g., an ally designating a civilian as a target by mistake). In these situations, AR errors risk not only reducing any performance benefits the AR system is expected to bestow, but may actively harm performance to such an extent that performance with inaccurate AR is worse than performance with no AR.

The purpose of this investigation is to examine the effects of angular, spatial distortion in AR symbology designed to support target acquisition, as spatial registration of AR symbology remains a significant technical challenge (Brunyé, Moran, Houck, Taylor, & Mahoney, 2016). Specifically, we investigate the effects of angular error between the true target and displaced AR symbology indicating a target at multiple ranges. Our primary research questions are 1) how much angular AR accuracy is needed for AR information to improve human performance, 2) how much AR error is necessary to harm human performance, and 3) how much AR angular error is necessary to decrease performance below that of perfect AR guidance. To study the effects of AR errors as they relate to target acquisition, our simulation uses the Night Vision Image Generator (NVIG) software to simulate the sensor feed of an infrared sensor. A realistic sensor controller enabled operators to scan a ring of virtual human targets, searching for a single human holding a weapon.

It is well known that humans can react faster to visual stimuli if they are given spatial cues to inform the human of where a visual stimulus is likely to occur (Posner & Cohen, 1984). Our AR simulation influences human target acquisition performance by both designating targets on a situational awareness ring and by placing a cue above the target in the operator's field of view (see Figure 1). The targeting cue in the situational awareness ring assists the user in gauging which direction and how far to rotate the sensor, based on the distance between the targeting symbology and the sensor's current azimuth heading; this symbol on the situational awareness ring also acts as a pre-cue that orients a person's attention to the symbology appearing over the target in the operator's field of view (Egly, Driver, & Rafal, 1994). The targeting symbol in the field of view subsequently orients the operator's attention to the designated target. While this type of AR symbology is not unique, little data exists quantifying the benefits of such AR symbology, or how spatially accurate such symbology needs to be in order to improve human performance. This is the second round of data collected from our NVIG target acquisition simulations, and the new data reflect several important simulation upgrades following the original pilot experiment (Graybeal & Du Bosq, 2018). Changes to the simulation and their impact on the data are discussed, and we also explore whether the initial experiment training is sufficient given a slightly more complicated target acquisition task.

METHODOLOGY

Adaptation of Existing Simulation Capabilities

Although imagery generated by NVIG has been used in past experiments to support evaluations of human performance with a sensor at NVESD (e.g., Graybeal, Du Bosq, & Nguyen, 2019), such experiments usually relied on static images presented on a computer monitor. Thus, one of the Perception Laboratory's goals has been to create more immersive scenarios for objective tests of human performance that might better represent the sensor's operational use.

In order to create the interactive scenarios used in this test, we leveraged NVIG's existing ability to represent a virtual world from a sensor view-port that could be controlled by a human operator (i.e., ability to rotate a sensor 360° in an immersive virtual world). A new software modification was required and developed in order to present the operator with a series of scenarios sequentially (i.e., when a given experimental trial was completed and the user gave a response, the next scenario needed to load automatically). This modification was necessary to support the large number of trials required by this experiment to generate adequate measurement sampling and statistical power.

Another NVIG capability that we initially lacked was a more efficient method of loading entities into precise locations within NVIG to build immersive scenarios. This was also a simple software development task, and we developed the capability to read in a spreadsheet of coordinates and entity names, so that scenarios could quickly be generated by modifying coordinates in a spreadsheet rather than adding entities manually to a scenario. This was required for careful control of entity location, and also to facilitate placing a large number of entities into a large number of scenarios.

Scene Generation

We created a series of virtual scenarios where participants had to search for, detect, and acquire a human target who was holding an AK-47 rifle; each scene contained only one target. Virtual humans were arranged in a partial ring (i.e., arc) around the sensor's location, so that each virtual human was equidistant from the sensor. The ring of potential targets covered a total area of 60° (30° on either side of the sensor's initial orientation), regardless of target range. In order to ensure our experimental design was sensitive to the effects of AR error, potential human targets were placed closely together at fixed intervals. This ensured AR spatial errors would force the operator to conduct visual searching in order to find the target. In other words, the AR system never made mistakes where the correct target, and only the correct target, would appear in the sensor's field of view despite angular error being present. The virtual humans were inserted into a flat, open terrain, such that scenes were devoid of buildings, vegetation, and other visual clutter (see Figure 1). Participants had a maximum of 90 seconds to find each target; this time limit was used to prevent participants from spending an excessive amount of time on any given trial and to prevent operator fatigue.

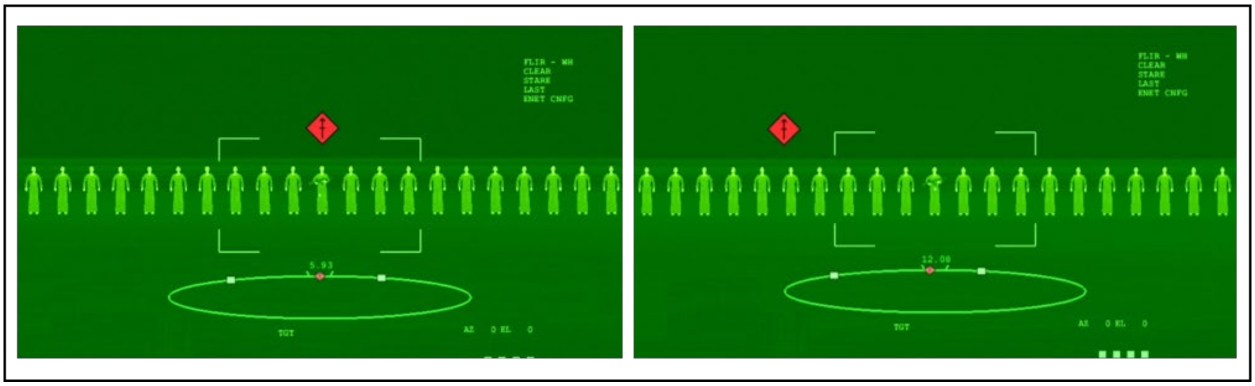

Figure 1: Sample scenes from the target acquisition simulation. Each scene contained an arc of virtual humans, and participants must align the center targeting reticle with the person holding the weapon. In the left image, the AR symbology correctly designates the target. In the right image, the AR symbology contains angular error, and is displaced slightly from the true target. Note that the AR symbology both appears on the screen above the target and on the sensor's situational awareness ring.

Experimental Design and Hypotheses

We studied the effects of several independent variables on target acquisition performance using a within-subjects design. First, we studied the effects of AR accuracy. Participants experienced six categorical levels of the AR performance: no AR (i.e., a control condition where participants had to complete the task unaided by any AR system), perfect AR (i.e., no angular displacement between the AR symbology and the true target), and four levels of imperfect AR, consisting of 1°, 2°, 3°, and 4° of angular error between the AR symbology and the true target. We hypothesized that greater amounts of AR error would increasingly impair target acquisition performance.

Second, we explored the effects of distance between the sensor and the target (i.e., range). We studied the effects of three ranges: a "Close" range where the target was easily visible without engaging the sensor's optical zoom, an "Intermediate" range where the target was visible without engaging the optical zoom but optical zoom greatly aided target acquisition, and a "Distant" range where the target was not detectable without engaging optical zoom. We hypothesized that target acquisition would take longer at extended ranges. In particular, we wanted to explore the interaction between range and AR information; for example, while 1° of AR angular error might not impair performance at a "Close" range, it might cause sufficient harm at more distant ranges.

Target locations were counterbalanced across AR error conditions and ranges, so that targets appeared in equivalent, but not identical, locations. This was accomplished by creating eight location "windows" (e.g., 4.5°-7.5°), but randomly placing the target within that window, so that target locations could not be learned by participants using the sensor's azimuth heading. The target location windows were centered at 6°, 9°, 12°, and 15° to the left and right of the sensor's starting origin. A target was placed at each of these target windows, for each range, for each AR condition.

Consequently, the total number of target acquisition trials per participant was 144 (6 AR conditions by 3 Ranges by 8 target window locations = 144). These trials were subdivided into eight blocks of 18 trials so that participants could periodically take breaks; each block was counterbalanced to contain three trials each of the six AR conditions. Each participant took each block, and each trial within a block, in a randomized order.

Sensor Controls and Targeting Reticle

We used highly realistic sensor grips, developed previously at NVESD, as the human/computer interface. The simulation controller was mounted to a stationary desktop in front of a computer monitor; pushing on the grips, either to the left/right or up/down, caused the sensor to rotate at a speed proportional to the strength of the push. The controller's sensitivity was set low enough to allow operators to easily acquire the targets at the most distant range (i.e., too much sensitivity makes it difficult to acquire small, distant targets).

Soldiers could engage the optical zoom of the sensor and used a "Laser Range Finder" button to designate targets. A "Menu" button was used to control a simple dialogue box that appeared after a Soldier designated a target, allowing the operator to "Confirm" or "Cancel" the designated target. A third button was used as a "speed boost" that enabled operators to rotate the sensor's field of view faster. This feature was added, along with a change to make the sensor rotate at a speed proportional to the sensor's field of view (which changes with the level of optical zoom), after participants in the initial pilot implementation complained that the sensor rotated too slowly for close targets (Graybeal & Du Bosq, 2018).

A different targeting reticle was displayed depending on whether or not the optical zoom was enabled; each reticle included a single dot at the very center of the screen. Participants were instructed to align that targeting dot with the virtual target holding the weapon.

Figure 2: Perception Laboratory facilities and experimental configuration. The Perception Laboratory has 10 workstations for simultaneous testing (left). Controller grips were mounted to the desk and positioned in front of large, high definition 4K computer monitors (right).

Participants

Eighteen U.S. Army Soldiers were recruited through Headquarters, Department of the Army. The Soldiers arrived during two different sessions for a one-week stay each, participating in thermal vehicle identification training and other perception experiments in addition to the augmented reality simulation presented here (Graybeal, Monfort, Du Bosq, & Familoni, 2018). Soldiers' ages ranged from 20 to 46 years old (M = 30.3, SD = 7.4). Likewise, time spent in service of the military varied widely between participants (M = 7.8, SD = 4.1). All research procedures were carried out under a protocol for human subjects research approved by the U.S. Army Medical Research and Materiel Command Institutional Review Board.

Procedure

Participants were first given a group PowerPoint presentation explaining the AR simulation instructions and sensor controls. Participants were instructed to acquire the targets as quickly as possible. They were also told that an AR system would attempt to help them during the target acquisition task, but that it would not always function perfectly.

Participants then participated in training scenarios to learn the sensor controls and to practice acquiring targets. The training consisted of three trials at each of the three ranges for each of three AR conditions: No AR, Perfect AR, and AR with 4° angular error (27 trials total). These three AR conditions were selected because they covered the full range of AR performance. Once participants completed the training, they began the experiment. While participants could always take a break between any of the eight blocks of trials, they were asked to take a ten minute break halfway through the experiment to alleviate fatigue. The instructions, training, and experiment collectively took approximately two hours.

Data Analysis

Data analysis was conducted using the *R Project* statistical analysis software. Hierarchical linear regression models (Bates et al., 2015) were used to analyze human performance data (target acquisition time and target acquisition accuracy), using Satterthwaite's method of approximating degrees of freedom for the calculation of t and p values (Satterthwaite, 1946). Nested-model comparisons were used to produce interpretable main effects (due to the presence of categorical variables with more than two levels in the primary regression analyses). Two regressions were planned per dependent variable to answer our primary research questions: the first comparing all AR conditions to No AR assistance and the second comparing all imperfect AR conditions to Perfect AR assistance. Additional post-hoc regressions, subsetting the data at various ranges, were conducted to further test these hypotheses at specific ranges; the Bonferroni correction was applied to both the dual regression approach (α =.025) and post-hoc analyses (α =.008) to control the rate of Type 1 inference errors.

As target acquisition accuracy is a binary variable, logistic regression was used to analyse it. Target acquisition accuracy was calculated purely in terms of angular error between the true target, the sensor, and the target designation pathway through three-dimensional space indicated by the participant; vertical accuracy was ignored. A response was scored as correct if the designated path through three-dimensional space was no more than half of a meter away from the target (i.e., the designated path was closer to the target than any other virtual human).

RESULTS

Target Acquisition Time

A nested model comparison revealed a significant main effect of range on target acquisition time $X^2(1, N = 18) = 422.97$, $p < .001$, such that target acquisition time increased at longer ranges, as predicted ("Close": M = 12.39 s, SD = 6.03 s; "Intermediate": M = 22.52 s, SD = 15.75 s; "Distant": M = 27.96 s, SD = 18.73 s). Likewise, a nested model comparison revealed a significant main effect of AR condition $X^2(5, N = 18) = 639.00$, $p < .001$. Target acquisition times were fastest with perfect AR, increased with increasing amounts of angular error, and were slowest with No AR (see Table 1). Likewise, variance in target acquisition times increased with increasing amounts of angular error, with No AR representing the least consistent acquisition times and Perfect AR representing the most consistent acquisition times.

A hierarchical linear regression model revealed that, compared to No AR, all AR information significantly improved target acquisition times (all p-values < .001). Further, compared to the No AR condition, the *increases* in target acquisition time as *range increased* were significantly smaller with all AR conditions (all p-values < .001). In other words, AR information protected against the impairments in target acquisition normally seen with increased range, although the magnitude of this protection decreased as angular error increased. Further post-hoc regressions, subsetting the data by range, indicated that all AR conditions were a significant improvement over No AR at all ranges (all p-values < .001).

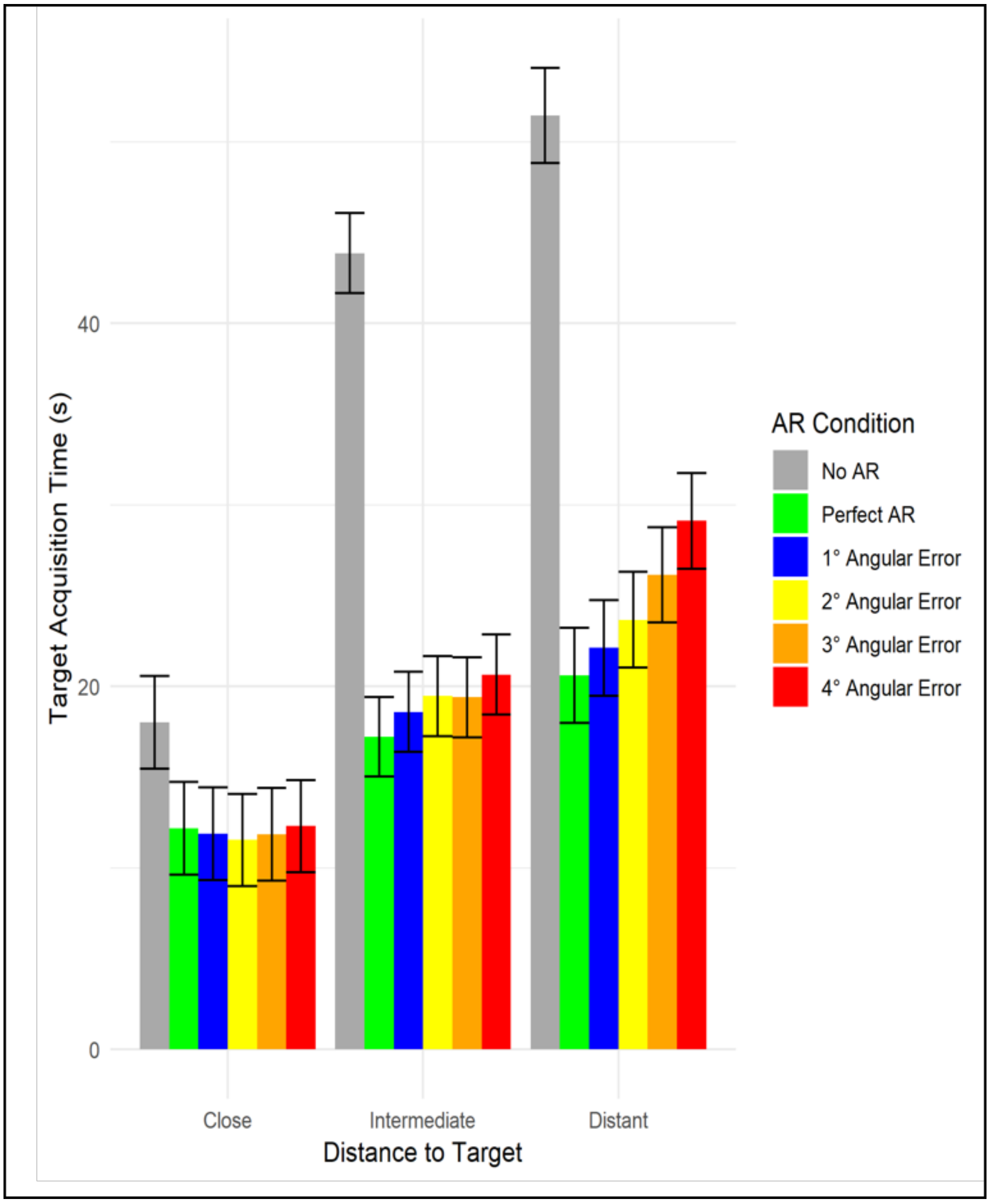

Figure 3: Target acquisition time by range and AR condition. Increased range (i.e., distance to the target) and increased amounts of angular error increased target acquisition times. Error bars represent 95% confidence intervals (based on standard error estimates calculated in the hierarchical linear regression models).

A second hierarchical linear regression model, using Perfect AR as the reference group, revealed that 1° (B = -.69, p = .394) and 2° (B = 1.39, p = .086) of angular error did not significantly differ from perfect AR, but 3° (B = 2.03, p = .012) and 4° (B = 3.86, p < .001) of angular error resulted in significant impairments in target acquisition time. Compared to perfect AR, 3° (B = 2.06, p = .011) and 4° (B = 3.45, p < .001) of angular error also showed significantly larger increases in target acquisition time as range increased. Further post-hoc regressions, subsetting the data by range, indicated that there were no significant differences (relative to perfect AR) with any of the four imperfect AR conditions at the "Close" range. At the "Intermediate" range, only the 4° of angular error resulted in significantly worse performance than perfect AR ($B = 3.10$, $p = .035$). Finally, at the "Distant" range, both 3° ($B = 4.61$, $p = .008$) and 4° ($B = 8.41$, $p < .001$) of angular error again resulted in significantly worse performance.

Table 1. Mean target Acquisition Times by AR Condition

AR Error Condition	Mean (s)	Standard Deviation (s)
Perfect AR	16.07	7.18
1° Angular Error	16.77	8.95
2° Angular Error	17.46	9.20
3° Angular Error	18.06	10.51
4° Angular Error	19.92	12.55
No AR	37.40	26.47

Table 2 shows the difference in mean target acquisition times between imperfect AR and No AR/Perfect AR guidance; it demonstrates how even imperfect AR assistance dramatically improved average target acquisition performance relative to the No AR condition. At the "Close" range, imperfect AR reduced target acquisition time uniformly by approximately 5 to 6 seconds compared to No AR guidance. Both the benefits of AR and the effects of angular error became more impactful at longer ranges, such that target acquisition time was reduced by 30.62 seconds with 1° of angular error at the "Distant" range (compared to No AR guidance), versus a smaller 23.21 second reduction with 4° of angular error. Compared to Perfect AR, there were no meaningful differences between any of the imperfect AR conditions at the "Close" range. At the "Distant" range, 1° of angular error introduced a (statistically non-significant) 1.01 second delay in target acquisition versus a (statistically significant) 8.42 second delay with 4° of angular error.

Table 2: Difference in Target Acquisition Times Between Imperfect AR and Control Conditions

	Compared to No AR			Compared to Perfect AR		
	"Close"	"Intermediate"	"Distant"	"Close"	"Intermediate"	"Distant"
1° Angular Error	-5.58 s	-25.64 s	-30.62 s	-0.12 s	1.25 s	1.01 s
2° Angular Error	-6.01 s	-24.81 s	-28.95 s	-0.55 s	2.08 s	2.68 s
3° Angular Error	-5.87 s	-25.02 s	-27.04 s	-0.41 s	1.87 s	4.59 s
4° Angular Error	-5.37 s	-23.79 s	-23.21 s	0.09 s	3.10 s	8.42 s

Target Acquisition Accuracy

Excluding trials where participants were unable to designate a target within the 90 second time limit, accuracy was extremely high for all participants, at each of the three ranges: 100% at the "Close" range, 99.62% at the "Intermediate" range, and 99.74% at the "Distant" range. As such, target acquisition accuracy in our experiment almost exclusively reflects the ability to identify the target within the 90-second time limit, rather than the ability to accurately designate the target in general (i.e., errors in correctly aligning the targeting reticle or mistakenly designating incorrect targets).

A nested model comparison, including trials where participants failed to designate a target in 90 seconds, revealed a significant main effect of range on target accuracy $X^2(1, N = 18) = 33.66$, $p < .001$; accuracy was perfect (i.e., 100%) at the "Close" range, while slightly and progressively lower at the "Intermediate" range (98.08%) and the "Distant" range (96.15%). Likewise, a nested model comparison revealed a significant main effect of AR error condition $X^2(5, N = 18) = 123.85$, $p < .001$; accuracy at the "Distant" range was lowest with No AR (80.581%), highest with perfect AR (100%), and increasing amounts of angular error decreased accuracy (see Table 3).

A hierarchical linear regression model, subsetting the data to analyse only the trials at the "Distant" range was selected, as performance for many AR conditions was perfect at both the "Close" and the "Intermediate" ranges (perfect, invariant performance causes computational problems for regression techniques). Performance on trials with Perfect AR was still perfect at the "Distant" range, so these trials were also excluded. Using No AR trials as the reference group, the regression revealed significant improvements in target acquisition accuracy for all imperfect AR groups (all *p-values* < .001). As performance was highest in the Perfect AR condition, we can infer differences between the No AR and Perfect AR trials are also unlikely to be due to mere chance, despite their exclusion from the model. A second regression, comparing performance on imperfect AR trials to Perfect AR was not conducted due to the invariance in the Perfect AR condition.

Table 3: Mean Accuracy Scores by AR Condition at the "Distant" Range

AR Error Condition	Mean (%)	Standard Deviation (%)
Perfect AR	100.0	0.0
1° Angular Error	99.31	2.9
2° Angular Error	99.31	2.9
3° Angular Error	98.61	4.0
4° Angular Error	99.31	2.9
No AR	80.85	18.3

Analysis of Experiment Training

If the training at the beginning of the study was insufficient for participants to learn to acquire targets as quickly as possible, we might observe progressively faster target acquisition times over the course of the experiment. Analysis of the average time to acquire a target during each of the eight sequential experimental block revealed that target acquisition times decreased throughout the first four blocks and also during the last four blocks, but reset during the transition from the fourth to the fifth block (see Figure 4). This likely reflects an effect of the mandatory break participants took halfway through the experiment. A nested model comparison revealed a significant effect of adding a variable to the model that accounted for the number of blocks since the start of the experiment or the break at the halfway point: $X^2(1, N = 18) = 13.96$, $p < .001$.

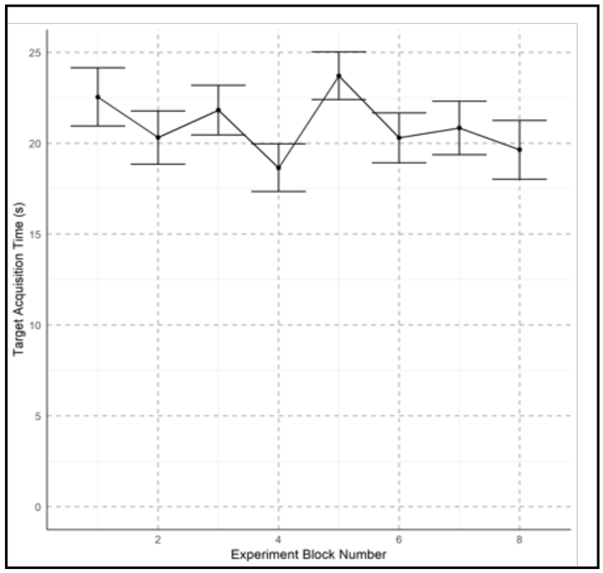

Figure 4: Analysis of average target acquisition time throughout the experiment.

DISCUSSION

Our experimental simulation demonstrates a clear method for studying target acquisition performance with imperfect AR symbology in order to better understand AR sensor requirements. With a relatively small sample size, we were able to demonstrate significant gains in human performance compared to completing the task without AR assistance; we were also able to distinguish when AR errors were becoming severe enough to impact human performance. Lastly, we were able to confirm two basic principles of AR symbology for target acquisition: 1) increasing amounts of AR error lead to progressive detriments in target acquisition performance and 2) that AR accuracy requirements increase as the task becomes more difficult.

Our research program joins a growing body of literature examining the impact of AR errors on human performance (e.g., Brunyé, Moran, Houck, Taylor & Mahoney, 2016; Graybeal & Du Bosq, 2018; Monfort, Graybeal, & de Visser, & Du Bosq, 2018). Degradations in human performance were confirmed by formal hypothesis testing, but even when impairments were not statistically significant, the data followed a clear pattern of progressive impairments due to increasing AR error. This is not a surprising outcome, but it is an important finding to share with the AR engineering community so that designers and consumers of these systems understand that the quality of the AR information provided to Soldiers affects their task performance. This principle is important, not only because small errors can matter, but also because it demonstrates the need to measure and define AR accuracy requirements.

Our simulation data also demonstrate that AR accuracy requirements change as the difficulty of the task increases. As range increased in our study and targets became smaller and more numerous, there was greater opportunity for the AR symbology to remediate human weakness, but deviations from Perfect AR also lead to greater detriments compared to ideal information. In our study, imperfect AR was never worse than No AR; the search task was generally difficult enough (and the errors were small enough) that even when the AR system was off considerably at long ranges, pointing operators in the correct general direction was still beneficial. However, this will not be true of every task (Monfort, Graybeal, & de Visser, & Du Bosq, 2018), and in our previous pilot of this experiment with a smaller field of regard, 4° of angular error at distant ranges yielded performance close to baseline performance with No AR guidance (Graybeal & Du Bosq 2018).

These findings demonstrate the importance of understanding the specific visual task an AR system will support, as AR requirements will change based on specific aspects of the task. NVESD has historically leveraged its electro-optical and infrared modelling capabilities and Perception Laboratory to define the necessary sensor characteristics (e.g., resolution, dynamic range, etc.) for military operators to complete specific visual tasks. In addition to making general contributions to the AR literature, our research program ultimately aims to develop simulation capabilities for empirically evaluating AR reliability requirements for specific sensors for specific tasks. These simulations should be able to set bounding parameters to inform system requirements, aid in the cost/benefit analysis of AR systems, and ultimately lead to more informed acquisition decisions. This research represents a major advancement toward that goal, as sensor parameters, the nature of targets, and target placement can quickly and easily be reconfigured using the NVIG experimental software developed under this effort.

The current simulation improved on several limitations from our initial implementation (Graybeal & Du Bosq 2018). Specifically, we added a "speed boost" (i.e., speed multiplier) button that enabled participants to more quickly orient to target designation symbology. We believe this more accurately represents the cognitive task we are attempting to model, specifically rapid orientation to threats using AR symbology. This change also made it possible to utilize a larger field of regard, as participants were not limited to searching at a single, slow rotation speed. Additionally, we added a sensor azimuth heading and randomly placed targets within controlled quadrants (as opposed to precise consistent locations) so that it was easier for participants to keep track of targets they had previously scanned if they decided to switch directions while searching.

As a result of these changes, improvements to human performance, relative to baseline, were larger than our initial pilot simulation (Graybeal & Du Bosq 2018), and this change is likely due to both the ability of participants to leverage the AR information more efficiently (i.e., by using the "speed boost") and the increased difficulty of the baseline task (i.e., it was more difficult to search a larger field of regard without assistance). In the current study, AR target designation symbology dramatically reduced target acquisition time. Even at the closest range (i.e., during the easiest task), perfect AR symbology reduced target acquisition time by approximately 30%. At the farthest range, Perfect AR reduced target acquisition time by approximately 60%.

Although these changes made the task less tedious and better aligned with the cognitive process we wanted to simulate, NVIG struggled to render all of the potential targets realistically while the sensor rotated rapidly with the "speed boost" engaged. This caused unrealistic and occasionally disorienting motion visualizations that should be improved in future simulations (the sensor's azimuth heading and icons on the situational awareness ring indicating the location of potential targets were not affected). Another aspect of the simulation that could be improved for future iterations is our target placement methodology, as targets were never placed at the most lateral regions in the arc of potential targets. This was done intentionally during the pilot version of the experiment when targets were placed in fixed locations (so that participants could not use the end of the arc of potential targets as a cue to learn probable target locations), but is no longer necessary now that we have developed the ability to randomly place targets within a controlled (and counterbalanced) location window (Graybeal & Du Bosq 2018). While these are both minor task idiosyncrasies, remediating them should improve the quality of future data.

After examining target acquisition times throughout the experiment, there was no clear evidence performance improved substantially over time. This suggests the training was likely sufficient to reach near-optimal performance before the experiment started. Although target acquisition times improved during blocks one through four and again during blocks five through eight, the performance resetting to baseline between blocks four and five during the mandatory break suggests this is only a temporary effect, likely due changes in motivation that allowed participants to acquire targets more quickly the longer they spent acquiring targets in the simulation. Still, for future simulations, it suggests we reconsider how we implement breaks during an experiment, perhaps forcing participants to take more frequent but shorter breaks in order to keep target acquisition performance as consistent as possible.

Our upcoming research will continue to explore AR target acquisition simulations by examining additional variables, such as the size of the field of regard, the sensor field of view, the density of potential targets, and the amount of visual clutter in the scenes; all of these variables have the potential to impact the difficulty of the search task and subsequently the quality of information needed to improve human task performance. This target acquisition study is also part of a larger NVESD effort to study the effects of AR errors on military tasks using military sensors. Other tasks we are currently simulating include AR assistance for thermal vehicle identification (Graybeal, Du Bosq, & Nguyen, 2019) and land navigation (Graybeal & Du Bosq 2018).

Importantly, the current laboratory simulation assumes relatively ideal operating conditions for the human user, whereas AR systems may be frequently utilized when operators must make a decision very quickly, under high stress, or while fatigued from combat or sleep deprivation. For example, time pressure might increase an operator's reliance on visual automation provided by an AR display (Rice & Keller, 2009). Although it is always challenging to simulate operational conditions in the laboratory, future experiments could examine how these variables influence an operator's use of AR.

As we continue to improve our methods of simulating sensors and AR content displayed within them, we plan to develop progressively more immersive scenarios. NVESD recently acquired an immersive AR/VR environment for human experimentation. The system consists of a VR headset and a wearable, backpack computer. Integrated into the headset is a pair of visible cameras. The environment consists of plexiglass panels that emit green light. Any objects seen by the camera remain in the VR headset's field of view, but a virtual world is superimposed over the camera's field of view wherever the camera sees the green walls and floor. Our future studies aim to evaluate sensors and the effects of augmenting those sensors with additional information using immersive displays and environments like this one, as depicted in Figure 5.

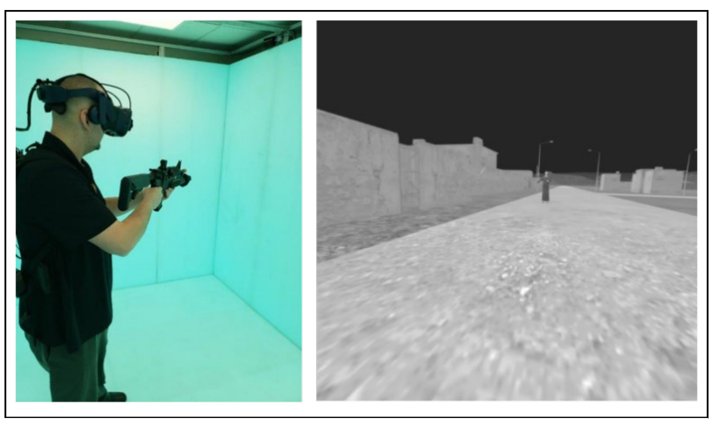

Figure 5: A person standing in NVESD's new immersive VR testing environment (left) and an example scene a participant might see, as rendered in NVIG (right).

Ultimately, besides gaining specific insight into target acquisition performance with AR, we hope that readers will realize the importance of the following principle: for every military task aided by AR, there is a quality threshold for AR information that must be surpassed in order for AR information to provide quantifiable improvements in human performance. Even when human performance is still enhanced by imperfect AR, performance may be degraded significantly by AR errors. Likewise, there is a lower bound for AR informational quality, and presenting AR information of quality beneath this threshold will cause human performance to degrade. We caution against AR designers making implicit assumptions about how accurate an AR system needs to be, as this can be explored empirically through simulation. While our simulation only explores the accuracy component of AR information, many other factors affecting the quality of the AR information exist and are worthy of careful consideration, including whether the AR information is perceivable, intuitive, timely, and relevant.

CONCLUSION

This paper presents a simulation of human target acquisition performance with both perfect and imperfect AR targeting symbology. Unsurprisingly, AR symbology aided target acquisition performance during the search task. Target search and acquisition time increased with range, and incremental AR errors progressively increased target acquisition time. AR information also protected against the impairments in target acquisition normally seen with increased range, although the magnitude of this protection decreased as angular error increased. While AR targeting symbology in this simulation was always a statistically significant improvement compared to a No AR control condition (all p-values $< .001$), larger amounts of angular error induced significant impairments at the more difficult ranges compared to perfect AR targeting symbology. Our results demonstrate that as the task difficulty (i.e., range to target) increases, AR accuracy requirements increase as errors become more damaging. Ultimately, our simulation marks a major step towards our goal of being able to define sensor- and task-specific AR requirements through simulation.

REFERENCES

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., & Singmann, H. (2015). lme4: Linear mixed-effects models using Eigen and S4, 2014. *CRAN, 1*(4).

Brunyé, T. T., Moran, J. M., Houck, L. A., Taylor, H. A., & Mahoney, C. R. (2016). Registration errors in beacon-based navigation guidance systems: Influences on path efficiency and user reliance. *International Journal of Human-Computer Studies*, *96*, 1-11.

Dixon, B. J., Daly, M. J., Chan, H., Vescan, A. D., Witterick, I. J., & Irish, J. C. (2013). Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical endoscopy*, *27*(2), 454-461.

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In M. Spector, M. Merrill, J. Ellen & M. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 735-745). New York, NY: Springer.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*(2), 161.

Graybeal, J. & Du Bosq, T. (2018). Augmented reality: Understanding human performance with imperfect systems by using virtual simulations. *Proc. of NATO MSG-159.*

Graybeal, J. J., Du Bosq, T. W., & Nguyen, R. T. (2019, May). Simulating human vehicle identification performance with infrared imagery and augmented reality assistance. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXX* (Vol. 11001, p. 110010A). International Society for Optics and Photonics.

Graybeal, J. J., Monfort, S. S., Du Bosq, T. W., & Familoni, B. O. (2018, April). Human perception testing methodology for evaluating EO/IR imaging systems. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIX (Vol. 10625, p. 106250T).* International Society for Optics and Photonics.

Monfort, S., Graybeal, J. & de Visser, E. & Du Bosq, T. (2018). Reliability requirements for augmented reality in visual search tasks. *Proc. of Interservice/ Industry Training, Simulation and Education Conference.*

Posner, M. I. & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance, 32*, 531-556.

Radu, I. (2014). Augmented reality in education: a meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, *18*(6), 1533-1543.

Rice, S., & Keller, D. (2009). Automation reliance under time pressure. *Cognitive Technology, 14*(1), 36-44.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components*. Biometrics Bulletin, 2*(6), 110-114.

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003, April). Comparative effectiveness of augmented reality in object assembly. *Paper presented at the meeting of CHI,* Ft. Lauderdale, Florida.

Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, *62*, 41-49.