Augmented Reality: Understanding Human Performance with Imperfect Systems by Using Virtual Simulations

John Graybeal, Ph.D.
KINEX, Inc.
P.O. Box 3107, Manassas, VA 20108
United States of America

Todd Du Bosq, Ph.D.
U.S. Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate
10221 Burbeck Road, Fort Belvoir, VA 22060
United States of America

ABSTRACT

Augmented reality (AR) displays are an emerging technology that provides computer-generated critical information to Soldiers, oriented towards enhancing Soldier decision-making capabilities and situational awareness. However, while AR technologies have great potential for enhancing Soldier effectiveness and safety, poorly designed AR technologies have the potential to distract, confuse, or mislead operators if they provide inaccurate information. A greater understanding of how accurate specific AR systems must be to improve human performance on various military tasks is still needed, as providing even slightly inaccurate information to a soldier may at times be worse than providing no information at all. Fortunately, simulations of AR technologies can be used to study aspects of its design and subsequent human performance to answer questions about AR accuracy requirements and other related human factors questions. These simulations can begin long before the first hardware prototype is finished, accelerating the technology's progress towards maturity. In this paper, we describe ongoing efforts by the U.S. Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate to simulate and understand human performance with AR and present some preliminary data from a target acquisition simulation. Our current research focuses on accuracy requirements for AR displays, and we discuss NVESD's plan to leverage our simulation capabilities and our Perception Laboratory to test and define sensor- and task-specific AR accuracy requirements for electro-optical and infrared sensor applications.

1.0 INTRODUCTION

The U.S. Army RDECOM CERDEC 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. Digital technology has made it possible to improve the way we present sensor information to a human operator, and as such, augmented reality (AR) technologies have become an important area of research and development for NVESD.

Augmented reality technologies attempt to enhance human sensory experiences by inserting digital information into the user's experience of the "real world," [1]. 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" [2-4]. 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 [5]. Thus, AR must be implemented carefully to avoid major risks to military personnel, or such technologies may do more harm than good.

The purpose of this paper is to present NVESD's ongoing research simulating AR use in a virtual environment, to present preliminary pilot data from one of the simulations, and to share how user feedback has been used to iteratively improve these simulations. We hope to demonstrate the importance of carefully studying human performance with AR and to demonstrate how simulations can be conducted early in the research process to inform device design constraints.

1.1 AR "Red Team" Research Program

One of NVESD's current research efforts focuses on a single issue related to AR: detriments to human performance caused by inaccurate information. It is unreasonable to expect AR systems to provide perfectly accurate information to human operators at all times. AR errors may be caused by limitations of the AR and/or display technology itself (e.g., poor geospatial accuracy for displayed symbology), other technology passing inaccurate information to an AR system (e.g., an automatic target recognition system misclassifying an object as a threat), or other humans passing incorrect information to the AR system (e.g., an ally designating an inappropriate target). In all 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.

NVESD has historically leveraged its electro-optical and infrared modelling capabilities and human Perception Laboratory to define the necessary sensor characteristics for military operators to complete specific visual tasks. Likewise, we have recently begun to address similar questions specific to sensors that incorporate AR technologies as we attempt to define what ways and to what extent AR systems need to be accurate in order to improve performance. NVESD's ultimate objectives of the AR "Red Team" research program are 1) to develop simulation capabilities and experiments that contribute to broad and general guidelines regarding AR system requirements and 2) to develop simulations and experiments capable of defining sensor-specific and task-specific AR requirements, the results of which can be leveraged to inform the design of specific technologies under development at NVESD.

The current paper briefly presents three ongoing simulation experiments at NVESD designed to explore the effects of AR mistakes on human performance during different military tasks, as well as the specific methodology and initial pilot data from one of those simulations (i.e., target acquisition).

1.1.1 Target Acquisition

Our first simulation under the AR Red Team research program investigates the effects of AR spatial accuracy on target acquisition time. Specifically, we are studying 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 AR angular error is necessary to decrease performance below that of perfect AR guidance, and 2) how much error is necessary to decrease performance below that of unaided performance (i.e., performance with no AR guidance). To study the effects of AR errors as they relate to target acquisition, our simulation uses the Night Vision Image Generator (NV-IG) software to simulate the sensor feed of a third generation Long Range Acquisition Sensor (LRAS3). Sensor grips simulating the controls of the LRAS3 allow operators to scan a ring of virtual human targets, searching for a single human holding a weapon. Dependent variables of interest are target acquisition time and target acquisition accuracy. This experiment is the primary focus of the paper and will be described in greater detail (see section 2.0).

1.1.2 Vehicle Identification

Vehicle identification with thermal imagery has been a major area of past research for the NVESD Perception Laboratory [6-7], and our second ongoing simulation is designed to simulate AR systems assisting soldiers with thermal vehicle identification. Our primary research question is how much AR classification accuracy is needed at various ranges to improve or detract from human performance. In this simulation, Soldiers are asked to identify images of thermal vehicles generated in the NV-IG simulation software, while being aided by a simulated AR system. Each image has a vehicle identity displayed above the target vehicle, simulating the assistance of an AR system. Images are presented to Soldiers sequentially in groups, with each group containing an inherent accuracy level (e.g., 100%, 75%, 50%) of the simulated AR labels. Further, we are investigating the effects of these variables in both time-limited (i.e., five seconds to make a decision) and time-unlimited circumstances. Dependent variables of interest are vehicle identification accuracy and response time.

Figure 1: Sample imagery from the thermal vehicle identification simulation. The yellow simulated AR label correctly identifies a T-72 tank (left) while the simulated AR label misidentifies the same tank in a second image (right).

1.1.3 Navigation

Our third simulation aims to assess the impact of inaccurate AR route waypoints on navigation performance. Our initial, primary research question investigates whether subtle or severe AR mistakes are more damaging to navigation performance. If followed, severe AR errors may be more damaging to performance, but they may also be more easily detected and disregarded by operators. Throughout the simulation, Soldiers are tasked with navigating to a specific building via a specified route in a virtual city generated in NV-IG. Soldiers are instructed not to deviate from specified "safe" route, even though it may not be the most efficient path. The route is displayed on a static map of the area on a second monitor screen, which participants can rotate and reorient based on their current heading. Dependent variables of interest are time to reach the target destination, time spent outside the designated route, the number of incorrect turns made, and path efficiency.

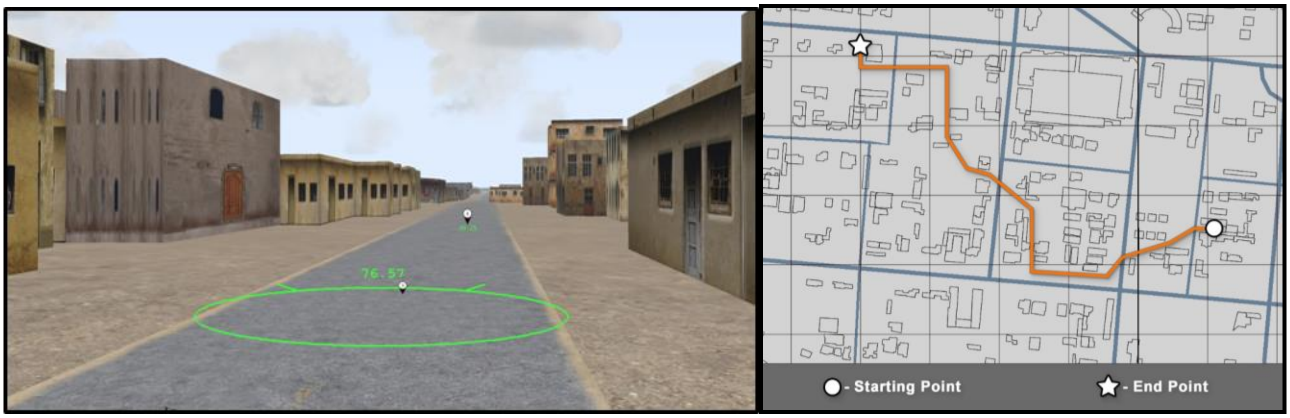

Figure 2: Sample imagery from the AR navigation simulation. Participants follow AR waypoints through a virtual city (left) while attempting to navigate a designated, safe pathway indicated on a map (right).

1.2 Fidelity and Experimental Validity in an AR Experimental Context

In each of these three simulations, we aim to realistically replicate the way an operator would receive and use information from an AR system to complete a task. The fidelity of any simulation, defined as "the extent to which the virtual environment emulates the real world," [8], critically affects its usefulness. Fidelity refers to far more than the visual representation in a simulation (i.e., the extent to which the simulation imagery is aesthetically pleasing or realistic), as fidelity is a multidimensional construct [9]. In addition to the imagery, aspects of a simulation that correspond to reality may include presented stimuli and behavioural responses, operator muscle movements, interface controls, scenario context, etc. Not all of these aspects are equally important for a given simulation, as the purpose of the simulation determines which aspects are the most critical. For experiments simulating the cognitive use of AR information, the most important component of the simulation is the content of and the manner in which the AR information is presented to the user; experiment participants must use the AR information the same way they would in reality in order for the experiment to be valid. By maximizing the cognitive fidelity of an experiment, we can manipulate the AR information presented to participants in simple simulations and observe effects on performance likely to generalize outside the laboratory. In addition to presenting our methodology and preliminary pilot results of our target acquisition simulation, we discuss the ongoing, iterative process of improving the cognitive fidelity of the simulation based on user experiences.

2.0 METHODOLOGY

2.1 Participants

10 U.S. Army Soldiers were recruited through Headquarters, Department of the Army. The Soldier's arrived for a one-week stay, participating in thermal vehicle identification training and other perception experiments in addition to the augmented reality simulation presented here. Soldiers' ages ranged from 19 to 36 years old (M = 28, SD = 6). Likewise, time spent in service of the military varied widely between participants (M = 7.7, SD = 5.9). All research procedures were carried out under a protocol for human subjects research approved by the U.S. Army Medical Research and Material Command Institutional Review Board.

2.2 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; each scene contained only one target. Virtual humans were arranged in a partial ring (i.e., arc) around the sensors location, so that each virtual human was equidistant from the sensor. The ring of potential targets covered a total area of 30° (15° 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, human targets were placed close to each other, exactly one meter apart. This ensured that when the AR symbology was inaccurately displaced, the operator would need to conduct visual work 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. Participants had a maximum of 60 seconds to find each target.

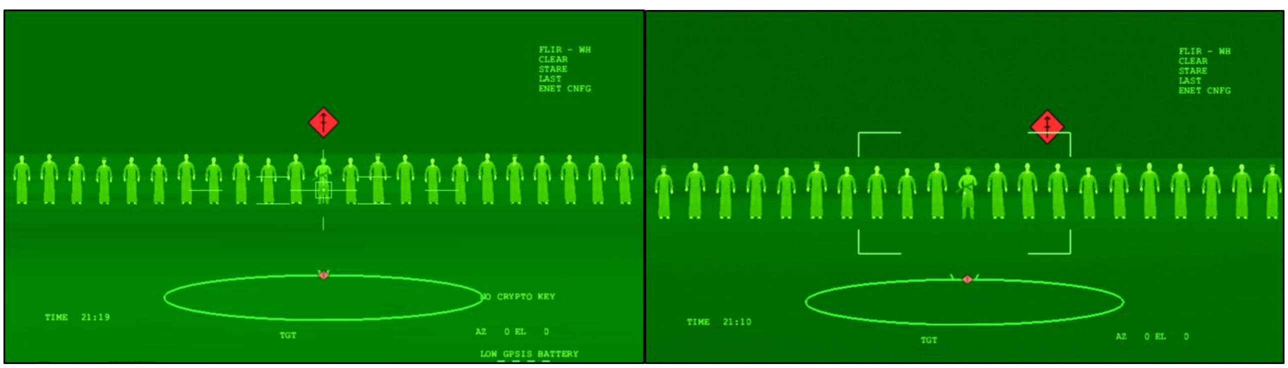

Figure 3: Sample scenes from the target acquisition simulation. The scenes display an arc of virtual humans, and participants must align the center targeting reticle on the person holding the weapon. In the left image, the AR symbology correctly designates the target. In the right image, the AR symbology contains angular error, and is displaced slightly from the true target. Note that the AR symbology both appears on screen above the target and on the sensor's situational awareness ring.

2.3 Experimental Design and Hypotheses

We studied the effects of several independent variables on target acquisition performance. 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 with longer 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 the "Close" range, it might cause sufficient harm at the "Distant" range.

Targets were placed at fixed locations so that the relatively slow sensor rotation speed would not confound target acquisition time across AR conditions. A total of eight target locations were used, with targets being located at 3°, 4.5°, 6°, and 7.5° to the left and right of the sensor's starting origin. We removed the sensor's azimuth heading to prevent participants from learning these locations, and piloted the experiment to ensure the target locations could not be memorized. A target was located at each of these locations, 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 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.

2.4 LRAS3 Controls and Sensor Targeting Reticle

We used highly realistic LRAS3 controller grips, developed previously at NVESD, as the human/computer interface. Component buttons and button layout on the simulation controller are nearly identical to the real LRAS3 controls. The primary divergence between the actual sensor and the simulation controller is the method of rotating the sensor's field of view (FOV). The LRAS3 sensor is typically tripod- or vehicle-mounted and is physically rotated by the sensor grips. In contrast, the simulation controller is mounted to a stationary desktop in front of a computer monitor; pushing on the grips, either to the left/right or up/down, causes the sensor to rotate at a speed proportional to the strength of the push. As no absolute controller sensitivity exists for a sensor that is physically rotated, the controller's sensitivity was set low enough to allow operators to easily acquire the "Distant" targets (i.e., too much sensitivity makes it difficult to acquire small, distant targets).

In general, the controls for the LRAS3 simulation were greatly simplified compared to the actual controller to facilitate rapid training. Soldiers could engage the optical zoom of the sensor and used the "Laser Range Finder" button to designate targets. The "Menu" button was repurposed to control a simple dialogue box that appeared after a Soldier designated a target, allowing them to "Confirm" or "Cancel" the designated target. All other LRAS3 buttons were disabled.

The LRAS3 displays a different targeting reticle depending on whether or not the optical zoom is enabled; we modified each to include a single dot at the very center of the screen. Participants were instructed to align that targeting dot with the virtual target they wished to designate.

Figure 4: Perception Laboratory facilities and experimental configuration. The Perception Laboratory has 10 workstations for simultaneous testing (left). LRAS3 controller grips were mounted to the desk and positioned in front of large, high definition 4K computer monitors (right).

2.5 Procedure

Participants were first given a group PowerPoint presentation explaining the AR simulation instructions and LRAS3 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 following 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 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 90 minutes. At the completion of the study, participants were debriefed and several took the opportunity to provide informal feedback regarding the simulation.

2.6 Data Analysis

Data analysis was conducted using the R Project statistical analysis software. Hierarchical linear regression models [10] 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 [11]. 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 Perfect AR and the second comparing all AR to No AR assistance. Additional post-hoc regressions, subsetting the data at various ranges, were conducted to further test 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 closer to the true target than any other virtual human.

3.0 RESULTS

3.1 Target Acquisition Time

A nested model comparison revealed a significant main effect of range on target acquisition time $X^2(1, N = 10) = 247.02$, $p < .001$, such that target acquisition time increased at longer ranges, as predicted ("Close": $M = 11.31$s, SD = 6.48s; "Intermediate": $M = 14.48$s, SD = 7.54s; "Distant": $M = 20.83$s, SD = 12.97s). Likewise, a nested model comparison revealed a significant main effect of AR condition $X^2(5, N = 10) = 97.42$, $p < .001$. Target acquisition times were fastest with perfect AR and 1° angular error, 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.

Table 1: Mean Target Acquisition Times by AR Error Condition

AR Error Condition	Mean (s)	Standard Deviation (s)
Perfect AR	13.37	5.35
1° Angular Error	13.10	6.66
2° Angular Error	13.95	8.33
3° Angular Error	15.14	9.27
4° Angular Error	17.47	12.23
No AR	20.19	14.58

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 perfect AR ($B = -2.34$, $p = .003$). The increases in target acquisition time with increased range observed with the various imperfect AR conditions were statistically equivalent to those observed with No AR. In other words, perfect AR protected against the impairments in target acquisition normally seen with increased range while there was not sufficient evidence that imperfect AR did. Further post-hoc regressions, subsetting the data by range, indicated that all AR conditions were a significant improvement over No AR at close range, (all *p-values* <.001), but that at 4° of angular error was no longer a significant improvement at "Intermediate" ($B = -0.98$, $p = .409$) and "Distant" ranges ($B = -1.83$, $p = .311$) over No AR, while 3° of angular error also lacked evidence of improvement at the "Intermediate" range ($B = -1.89$, $p = .089$) (all *p-values* for 1° and 2° angular error at all ranges were <.002).

A second Hierarchical linear regression model, using Perfect AR as the reference group, revealed that 1° ($B = -.28$, $p = .723$) and 2° ($B = .572$, $p = .464$) of angular error did not significantly differ from perfect AR, but 3° ($B = 1.77$, $p = .024$) and 4° ($B = 4.09$, $p < .001$) of angular error resulted in significant impairments in target acquisition time. Compared to perfect AR, 4° of angular error showed significantly worse increases in target acquisition time as range increased ($B = 3.71$, p<.001). 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 "Close" or "Intermediate" ranges (all $p > .14$), and that only the 4° of angular error resulted in significantly worse performance than perfect AR at the "Distant" range ($B = 5.15$, $p < .001$).

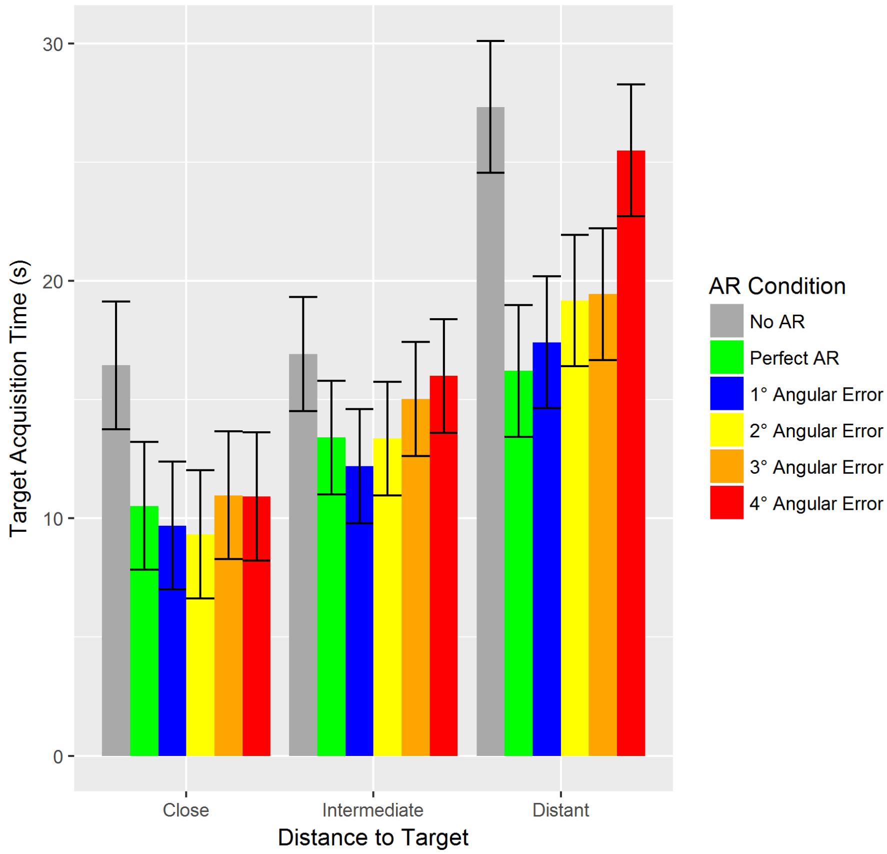

Figure 5. Target acquisition time by range and AR condition. Increased range (i.e., distance to the target) and increased amounts of angular error increased target acquisition times. Error bars represent 95% confidence intervals (based on standard error estimates calculated in the hierarchical linear regression models).

3.2 Target Acquisition Accuracy

Excluding trials where participants were unable to designate a target within the 60 second time limit, accuracy was extremely high for all participants, at each of the three ranges: 100% at the "Close" and "Intermediate" ranges, and 99.56% at the "Distant" range. As such, target acquisition accuracy in our experiment almost exclusively reflects the ability to identify the target *within the 60-second time limit*, rather than the ability to accurately designate the target in general (i.e., mistakenly designating incorrect targets or errors in correctly aligning the targeting reticle).

A nested model comparison revealed a significant main effect of range on target accuracy $X^2(1, N = 10) = 54.15$, $p < .001$; accuracy was perfect (i.e., 100%) at "Close" and "Intermediate" ranges, while slightly lower (94.99%) at the "Distant" range. Likewise, a nested model comparison revealed a significant main effect of AR error condition $X^2(5, N = 10) = 20.57$, $p < .001$; accuracy at the "Distant" range was lowest with No AR (86.1%), highest with perfect AR (100%), and increasing amounts of angular error decreased accuracy (see Table 2).

Table 2: 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	98.8	4.0
2° Angular Error	96.3	6.0
3° Angular Error	95.0	12.1
4° Angular Error	93.8	8.8
No AR	86.1	15.3

Performance on trials with Perfect AR was invariant (i.e., perfect), causing computational issues for the planned logistical regression techniques, so performance on these trials was omitted. A hierarchical linear regression model, subsetting the data to analyse trials at the "Distant" range and using No AR trials as the reference group, 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, logic dictates that these differences can be viewed as "significant," 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.

3.3 Qualitative User Feedback

While not a formal outcome of interest, participants expressed two concerns regarding the simulation during informal conversations following the experiment: 1) the sensor rotation speed was "too slow" at the "Close" range and 2) they disliked the deletion of the sensor's azimuth heading, as they reported it could be disorienting to search through a large ring of potential targets without one (see section 2.1). We address the implications of this feedback in the Discussion section.

4.0 DISCUSSION

Our experimental approach and results represent an early exploration into the ability to simulate the effects of AR inaccuracies on military task performance, in this case visual search and target acquisition. They provide an experimental template for formally investigating the effects of AR information on human performance before a physical prototype is even complete, and demonstrate our intended approach to systematically simulating AR information for a variety of applications related to electro-optical and infrared sensors and head-mounted displays. In addition to contributing to a general understanding of AR and human performance, our goal is to leverage our simulation capabilities to eventually define sensor- and task-specific AR capabilities required to improve human performance. This work marks a significant step towards that goal, as we have linked our AR target acquisition testbed to our existing simulation capabilities in NV-IG; it is straightforward to change the properties of the sensor being studied, as well as the location and nature of displayed targets within NV-IG.

While the simulation presented here is still being improved, partially in response to participant feedback as discussed below, we were able to show statistically significant changes in target acquisition time based on the observer's distance to the target and on the quality of the AR information provided to the Soldiers with a sample size of only 10 participants. Our results further indicate that incremental degradations in the accuracy of AR information incrementally affect target acquisition performance and that greater AR accuracy is needed in order to improve performance as range increases. These conclusions concur with NVESD's previous AR visual search simulations using static images, as opposed to the more interactive search environment presented to participants in this study [12].

One of the most important aspects of an AR simulation is the cognitive dimension of fidelity (i.e., people must use the AR and sensor information to complete a task in a way that corresponds to reality). Feedback from our pilot participants suggested that the cognitive fidelity of our experiment could be improved for future simulations. The decision to remove the sensor azimuth heading (in order to prevent participants from learning the location of the targets) changed the way the participants completed the task. Many sensors, including the LRAS3, include an azimuth heading which helps the Soldier maintain situational awareness while scanning for targets. Participants in our simulation reported disorientation as a result of the deleted azimuth heading, perhaps increasing their disorientation compared to what an LRAS3 operator would experience under real operational circumstances. To address this concern, we reintroduced the sensor azimuth heading for future simulations and instead randomly placed the targets within counterbalanced sections of the arc of potential targets; this allowed us to control target location across trials containing different types of AR information while still preventing participants from learning target locations.

Soldiers also commented on the slow rotation speed of the sensor at the closest range (the absolute rotational speed was the same for all ranges, but is perceived as slower at closer ranges). While this speed allowed for proper target acquisition at the "Distant" range, it was annoying to participants and may have changed the cognitive strategy best suited for the task. Indeed, it is possible that the sensor rotated so slowly that participants were able to view each possible target clearly while rotating towards the AR symbology; this would minimize the detriments of minor AR inaccuracies, as participants would gain substantial benefit by simply being pointed in the correct general direction to scan. To address these issues in future simulations, we altered the spin rate as a function of the sensor's FOV so that the sensor rotates faster in general, but slows down when the optical zoom is engaged. This allows for accurate reticle alignment and target acquisition at long ranges while accelerating sensor rotation at closer ranges. Ultimately, we decided the cognitive process we wanted to simulate consisted of the operator receiving spatial information by an AR beacon, orienting quickly to that target, and then struggling to find the target due to AR angular error. To better mimic this cognitive process in our simulation (i.e., increase cognitive fidelity), we added a speed acceleration button for our future simulations. Participants can now orient faster to AR symbology while scanning a larger field of view. This will make the target acquisition times recorded in our future simulations more purely reflective of time spent searching following an AR mistake, rather than a slowly rotating sensor.

In addition to range, other possible variables unexplored in the present study may affect AR reliability requirements. For example, the amount of visual searching required when AR is inaccurate depends on the density of potential targets and the amount visual clutter present (e.g., buildings, vegetation, etc.). Both target density and visual clutter are variables of interest for our future simulations, as AR accuracy requirements may change as a function of both: greater precision may be needed with greater target density and clutter, but errors may be more tolerable with only a few potential targets to scan or in clutter-free environments.

Finally, we should note the statistical analysis presented here could also potentially be improved by utilizing a more complicated analysis technique, such as a hierarchical linear regression modelling approach that predicts aided performance relative to unaided performance in a single regression [12], as the use of multiple independent hypotheses tests and the subsequently employed Bonferroni correction effectively reduces statistical power. A larger sample size for future simulations will also aide in providing clear evidence of detriments caused by AR.

5.0 CONCLUSION

This paper presents our ongoing simulation efforts to study human performance with imperfect AR information. Our current work demonstrates a method of simulating human performance with virtual AR information, which can be applied to a variety of tasks. Our successful collection of preliminary data via simulation marks a major step towards our goal of being able to define sensor- and task-specific AR requirements through simulation. The results of this study demonstrate the effect of angular error on target acquisition performance depends on both the amount of angular error present and the range to the target. As our simulation iteratively improves and we collect data from additional participants, greater measurement specificity will be achieved. NVESD will continue researching human performance with AR through simulation to support the development of electro-optical and infrared sensors.

REFERENCES

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] Graybeal, J. J., Monfort, S. S., & Harkrider, S. (2017). Environmental Cues in Thermal Images Impair Vehicle Identification Training: Simulated Thermal Imagery as a Potential Solution. *Proc. of NATO MSG-149.*

[7] 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.

[8] Alexander, A. L., Brunyé, T., Sidman, J., & Weil, S. A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group, 5*, 1-14.

[9] Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine, 89*(3), 387-392.

[10] 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).

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

[12] Monfort, S. S. (2017*). Performance, trust, and workload in an automation-aided visual search task* (Unpublished doctoral dissertation, George Mason University, Fairfax, VA).