ARMY RESEARCH LABORATORY

Impact of 360° Sensor Information on Vehicle Commander Performance

by Keryl Cosenzo, Jason Metcalfe, Stephen Gordon, Tony Johnson, Theodric Feng, Brad Brumm, A. Willliam Evans, and Erin Capstick

ARL-TR-5471 **March 2011**

NOTICES

Disclaimers

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5425

ARL-TR-5471 **March 2011**

Impact of 360° Sensor Information on Vehicle Commander Performance

Keryl Cosenzo, Theodric Feng, and A. William Evans
Human Research and Engineering Directorate, ARL

Jason Metcalfe, Stephen Gordon, Tony Johnson, and Erin Capstick
DCS Corporation

Brad Brumm
U.S. Army Tank-Automotive Research, Development and Engineering Center

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
March 2011	Final	October 2009–September 2010

4. TITLE AND SUBTITLE
Impact of 360° Sensor Information on Vehicle Commander Performance

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Keryl Cosenzo, Jason Metcalfe,[*] Stephen Gordon,[*] Tony Johnson,[*] Theodric Feng, Brad Brumm,[†] A. William Evans, and Erin Capstick[*]

5d. PROJECT NUMBER
IMOPAT ATO

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: RDRL-HRS-E
Aberdeen Proving Ground, MD 21005-5425

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-5471

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
[*] DCS Corporation, Alexandria, VA
[†] U.S. Army Tank Automotive Research, Development and Engineering Center, Warren, MI

14. ABSTRACT
Maintenance of local security is essential for the lethality and survivability in modern urban conflicts. Among solutions the U.S. Army is developing is an indirect-vision display (IVD) based sensor system supporting full-spectrum, 360° local area awareness. Unfortunately, such display solutions only address part of the challenge, with remaining issues spawned by the properties of human perceptual-cognitive function. The current study examined the influence of threat properties (e.g., threat type, distance, etc.) on detection performance while participants conducted a patrol through a simulated urban area. Participants scanned a virtual environment comprised of static and dynamic entities and reported those that were deemed potential threats. Results showed that the most influential variables were the characteristics of the targets; threats that appeared far away, behind the vehicle, and for short periods of time were most likely missed. Thus, if an IVD system is to be effective, it will be necessary to improve range performance and optimize the amount of viewing time for 360° imagery. Some results indicated target salience as also important. As such, real-time image processing may ultimately be necessary to account for perceptual-cognitive factors affecting detection and identification performance.

15. SUBJECT TERMS
360° situation awareness, displays, commander

16. SECURITY CLASSIFICATION OF: Unclassified			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Keryl Cosenzo
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** (*Include area code*)
Unclassified	Unclassified	Unclassified	UU	106	410-278-5885

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

The experiment described in this report was conducted to support the Improved Mobility and Operational Performance through Autonomous Technologies Army Technology Objective (IMOPAT ATO). The goal of the IMOPAT ATO is to enhance vehicle and platoon mobility performance and local-area/battlespace situation awareness for the vehicle crew and dismount infantry through the development and integration of: indirect vision-based intelligent manned and unmanned vehicle mobility, advanced crew stations, 360°/90° (horizontal/vertical) situational awareness (SA) systems, crew and dismount scalable interfaces, and neurophysiologically-based and behavior-based Soldier monitoring and workload management technologies. This experiment focused on one facet of the ATO goals, 360°/90° SA.

In modern battlespace concepts, it is believed that available technology does not provide a sufficient field of view (FOV) to enable indirect vision systems to support mobility and situational awareness functions required to meet changing force needs. The indirect vision systems that are necessary for closed-hatch operations do not currently provide full SA of the battlespace for vehicle crew or dismount Soldiers, which is especially problematic in complex, dynamic urban environments. That is, current sensor and information display systems are unable to provide contiguous, simultaneous 360°/90° imagery for vehicle crews. Moreover, the current hemispheric (360°/90°) imaging systems require continuous manual intervention by operators (i.e., scanning through successive camera views) to maintain local area awareness. This proves especially demanding in urban environments where the presence of three-dimensional (3-D) terrain (i.e., threats above ground, such as in tall buildings) increases scanning workload requirements. Additional mitigations and enhancements remain challenged as well. Automatic event/target detection technologies are limited, acoustic gunfire detection is made more difficult by many acoustically reflective surfaces in urban environments, and flash detection requires line of sight. The integrated on-the-move (OTM) moving-target indicator (MTI) cue for detection of pop-up and fleeting targets has not been demonstrated with hemispheric imaging systems in realistic environments.

To overcome the current sensor system limitations, the IMOPAT ATO partners (U.S. Army Tank-Automotive Research, Development and Engineering Center [TARDEC] and Communications and Electronics Research Development and Engineering Center [CERDEC]) are developing and integrating a 360°/90° sensor system for a manned ground vehicle. The partners will position multiple imaging sensors around the vehicle to provide day/night capability for manual scanning and control for vehicle crews. Additional technologies under consideration for integration are gunfire detection/localization, pop-up target cueing, image capture to aid the vehicle occupants in detection of immediate threats, and a digital video recording (DVR) capability to capture video of the threat at the time of detection as well as immediately prior to the event. Technology that will provide the Soldier automatic and manual imagery logging for past event revisit and interrogation during or after missions is also being considered to aid the Soldier in maintaining local SA. Summarily, the vision for this sensor system is that the field of regard will allow the commander to avoid threat-weapons fire, day or night, in complex urban environments and that the crew and squad members can manually detect threats or be cued by the detection algorithms to an immediate threat. Cueing of the vehicle occupants to detection/location of immediate threats is considered essential to vehicle protection and management of crew workload.

The U.S. Army Research Laboratory Human Research and Engineering Directorate's (ARL-HRED) role is to support the development of the display concepts for the 360°/90° sensor system and to conduct experiments to evaluate the effect of the technology and its implementation on operator performance.

The first step to providing enhanced SA to the operator is to provide enhanced visual displays. To this end, CERDEC developed a 360°/90° sensor system. This system is comprised of six cameras affixed at various positions around the vehicle. The sum of the six camera views provides the enhanced view, which is displayed on the vehicle commander's 17-in screen. Unfortunately, there is not sufficient display space to show the full 360°/90° view at one time. Thus, a novel approach to displaying the 360°/90° information is needed. The next step in the development process was the generation of design concepts for the vehicles commanders display. DCS, in support of TARDEC, CERDEC, and ARL, lead this design process. The objective was to develop concepts that enabled the commander to have the greatest field of view at all times with the greatest resolution possible.

Four design concepts were generated. All of the concepts incorporated a central sensor view that provided 64/48° field of view (see figure 1). This view could be changed by selecting a different camera. The banner provided a wider horizontal field of view (hFOV = 180°) for the operator with poorer resolution (see figure 2). The design concept was thus: through a banner an operator could gain an overall awareness of the area and then use the central sensor to further investigate an area of interest.

Figure 1. Central sensor concept.

Figure 2. Banner concept TC.

The objective of this research was to evaluate the effect of a 360° sensor system, similar to that developed by TARDEC and CERDEC, on the vehicle commander's ability to maintain local area awareness and disseminate the SA information to his crew. In addition to this primary objective, cognitive aspects of the operator's ability to maintain SA were also investigated. To this end, the influence of target features and environmental characteristics on the operator's ability to maintain local SA was objectively measured by various aspects of threat detection performance (detection rates, reaction time, and accuracy). Through this experiment, we planned to determine what combination of displays and interface tools constitutes the optimal interface design for 360° vision as it relates to local situational awareness. Results can be further used to identify performance bottlenecks that may require additional technology beyond a 360° field of view.

2. Method

2.1 Participants

Seventeen male individuals (*n* = 7 military and 10 civilian) participated in the experiment, which was conducted at TARDEC systems integration laboratory facilities in Warren, MI. Participants were recruited from the local population in Warren. Each participant read and signed a volunteer agreement affidavit (appendix A) and completed a demographics questionnaire (appendix B) prior to beginning the experiment.

2.2 Questionnaires

2.2.1 Demographic and Computer Experience Questionnaire

The demographic and computer experience questionnaire (appendix B) was an 11-item survey that requested information regarding age, vision and hearing, military service, and computer experience. This questionnaire was scored after the experiment was completed and was used to gain basic demographic information on the participants for potential use as covariates in subsequent data analyses.

2.2.2 The National Aeronautics and Space Administration (NASA) Task Load Index (NASA-TLX)

The NASA-TLX (Hart and Staveland, 1998) is a multidimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales (mental demand, physical demand, temporal demand, own performance, effort, and frustration) and was administered after each mission. The NASA-TLX is included in appendix C.

2.2.3 Usability Questionnaire and Exit Interview

The usability questionnaire was a 53-item questionnaire that assessed the ease of learning the different interface configurations and how each configuration affected field of view and performance. The exit interview was comprised of a set of 19 questions that were asked by the experimenter. Using the open ended questions, an experimenter queried the participant about the strengths and weaknesses of the configurations, their impact on mission performance, and suggestions for improvement. The usability questionnaire and exit interview were administered at the end of the experiment to obtain user feedback on the 360° system and the associated interface configurations; both are provided in appendix D.

2.3 Experimental Environment

The layout for the experimental environment and test bed is shown in figure 3. For this experiment, the actual Common Crew Station (CCS), which has been developed by TARDEC and CERDEC, was not available; therefore, it was decided to conduct the study using a surrogate system. The chosen surrogate for this work was an Alienware laptop computer, which provided the same screen size (17 in) and resolution (1920 × 1200) as the CCS, but did not provide a touch screen interface. Therefore, for this experiment, participants were required to use a mouse to interact with the system when using all interface configurations.

The Alienware laptop functioned as the Warfighter Machine Interface (WMI) and provided both sensor displays and controls. Sensor display was handled using the system's internal graphics processor, while sensor control was enabled through an interactive graphic positioned in the upper-right corner of the central concept sensor (figure 4).

The Embedded Simulation System (ESS) connected to the WMI and provided two major services: communication with the rest of the system and the vehicle dynamics model. The communication provided by the ESS was crucial for synchronized information interchange, while the vehicle dynamics models were used to maintain an absolute position of the vehicle within the simulated environment.

The Intelligent System Behavior Simulator (ISBS) was another process that ran on the Alienware laptop and provided vehicle control and ensured route following. The ISBS drove the vehicle using information received from the ESS and then provided the ESS with the appropriate actuator data to keep the vehicle inside the pre-specified route conformance parameters.

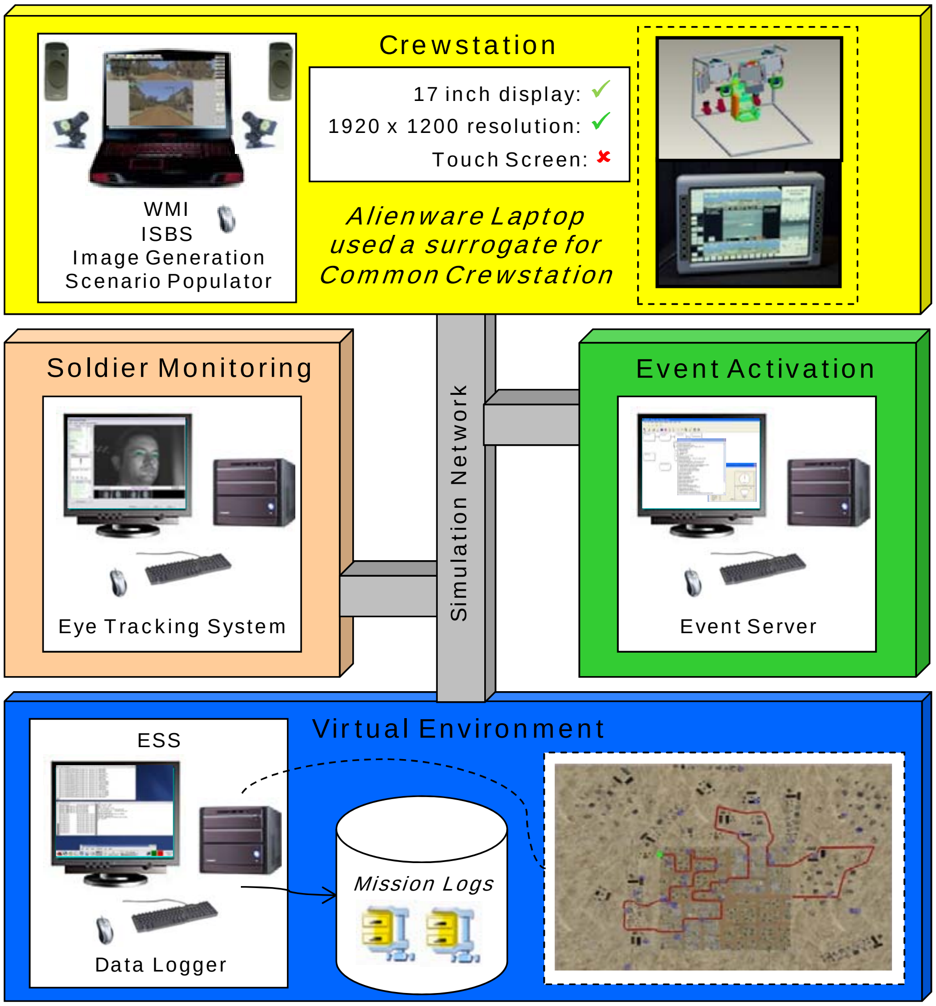

Figure 3. Experimental testbed.

Figure 4. Sensor view with sensor control.

The Event Server controlled execution of all events by receiving vehicle location information from the ESS and using pre-defined trip lines as triggers for event onsets. The trip lines were used to indicate where the participant was in the current scenario and the location of each trip line was specified prior to the experiment using a custom-designed interactive tool, Scenarion. Once an event was triggered, it was sent either to the scenario populator or to the sound player, dependent on whether the event called for image generation or audio commands. In cases where an audio command was needed, a pre-recorded audio file (.wav) appropriate to the event was triggered. All events were recorded and time-stamped in event log files for later use during data reduction and analysis.

The Scenario Populator ran on the Alienware laptop and received events from the Event Server and output distributed interactive simulation (DIS) packets representing a set of entities moving around the database, mostly along routes predefined using Scenarion. These DIS packets were interpreted by the ESS, which passed them back to the internal graphics running the WMI. The DIS packets were also recorded by the DIS Recorder for post-processing.

During the experiment, eye tracker data were gathered for each participant using the Smart Eye Pro system (Smart Eye AB; Göteborg, SE), a commercial eye tracking system that provided infrared (IR) cameras, IR emitters, and software tools for researchers. The Smart Eye system allowed completely non-contact operation in order to provide observation of a participant's natural eye and head movement behavior at adequate spatial resolution (~0.5°). Eye and head movements, along with measurement reliability data, were logged in real time and synchronized with performance data from the other systems. The system was individually calibrated for each participant prior to training, a process that was completed in the approximate 10–15-min window while participants were briefed about experiment protocols. Once calibrated, the Smart Eye system was used to gather data for the entire experiment at a rate of 60 Hz. For these experiments, two IR cameras (and emitters) were used, which were positioned just to the left and right of the experimental display screen.

During the experiment, six log files were generated by the overall system. The log files included: (1) event times and descriptive tags from the event server, (2) user screen interactions from the WMI, (3) entity positions and movements from the DIS Recorder, (4) vehicle state from the ESS, and (5) eye position data from the Smart Eye. At the end of each mission a final tool, the line-of-sight (LOS) checker was used to read the vehicle state log and the DIS recorder log in order to determine the times and locations at which the vehicle had LOS to each of the entities as well as to provide information about in which of the six 360° vision sensors each LOS was present. This created the sixth and final artifact of the experiment, the LOS log.

2.4 Procedure

The overall flow of the experiment and the procedure, along with a general timeline, is shown in figure 5. At any time during the experiment, the participant was allowed to take a break.

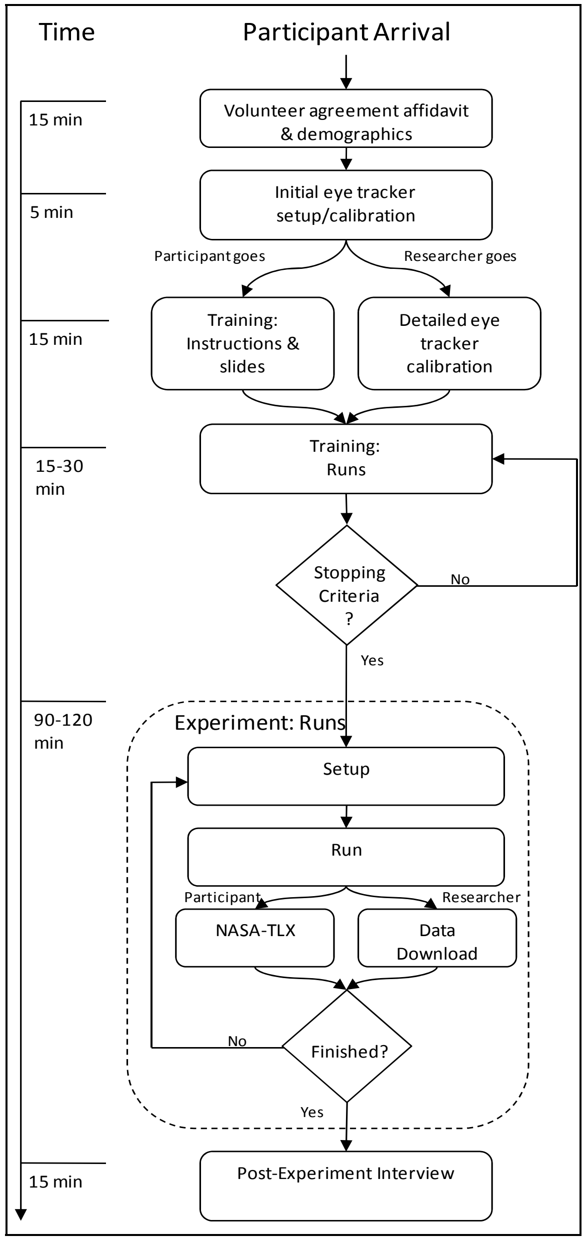

Figure 5. Experiment procedure flow diagram.

Prior to the start of the experiment, the participants were briefed on the purpose and procedures of the experiment and were read the volunteer agreement affidavit (appendix A). They were given the required brief regarding confidentiality as indicated on DA Form 5303-R. In anticipation of possible concerns regarding personal answers on some of the questionnaires, the investigators also described the deliberate actions taken when handling research data. In order to ensure that individual data were not reported or revealed to anyone, each form was reviewed upon receipt by one of the investigators. If any identifying information appeared on the questionnaires (such as name, social security number, birth date, etc.), the investigators deleted the identifying information and replaced it with a neutral code number. This code number became the participant identification number used in data files.

When recruiting military research participants, concerns about actual willingness to volunteer for a given study may arise if higher-ranking personnel suggest participation. To avoid the possibility that volunteers may have been compelled to participate through their chain of command, participants were reminded that they could refuse or withdraw from the study at any time without penalty. Participants were given an opportunity to communicate with investigators "off the record" and were provided several opportunities for refusing or withdrawing in a private manner. Participants who agreed to take part in the study signed the volunteer agreement affidavit. The participants then completed the demographic and computer experience questionnaire.

Next, the purpose of the experiment was described to the participants, and then they were given an overview of the baseline scanning system display and the advanced 360° scanning system displays (described in section 3 of this report). The experimenter reviewed the functionality of the interface, and how to use the scanning systems for local security and target identification. After receiving the overview, the eye tracker was set up for each participant. Following the initial calibration with the eye tracker, the participant was then shown training slides and familiarized with the scanning system while a second researcher performed a detailed calibration of the Smart Eye system. The training slides presented a description of the following:

- Purpose of the study.
- What is 360°/90° Situation Awareness?
- Pictorial and verbal descriptions of each configuration.
- Steps on how to change the camera views of the 360° system for each configuration.
- Pictorial and verbal descriptions of the types of targets to be identified.
- Steps on how to report targets on the interface.

Once the participant completed the training slides and the researcher finished final calibration of the Smart Eye system, the participant was then required to complete two training missions: one with condition A and one with condition D, described later. Participants were required to repeat the training missions if they failed to detect at least 50% of the targets for that mission. This was determined by a subject matter expert on the technology who was present during the training missions. In addition, participants could repeat a training mission if they desired. After being trained on the system, participants then completed four experimental missions.

After the completion of each mission, the participant filled out a NASA-TLX questionnaire while the experimenter collected and saved the mission-specific data logs. Once the fourth

NASA-TLX was completed after the final mission; participants were given the usability questionnaire and the exit interview to assess their overall impressions and preferences with regards to the system. The final exit interview was open-ended, which also allowed the participants an opportunity to provide voluntary feedback that may otherwise have been missed by the questionnaires.

3. Experimental Design

The experiment was a within-subjects design. There was one primary independent variable of interest, display condition, with four levels. The four levels were:

1. Condition A = sensor only: full-screen mode.
2. Condition B = sensor with a single banner: full-screen mode.
3. Condition C = sensor with a single banner: small-portal mode.
4. Condition D = sensor with two banners: full-screen mode.

The different screen configurations for each of the conditions used in this experiment are described next, and a summary of the screen configurations is shown in table 1.

Table 1. Condition breakdown.

Condition	Sensor Window	Top Banner	Bottom Banner	Full-Screen Mode	Small-Portal Mode
A	×	—	—	×	—
B	×	×	—	×	—
C	×	×	—	—	×
D	×	×	×	×	—

Condition A = sensor only: full-screen mode (figure 6).

- Single 64 × 48 FOV sensor: 1024 × 768 resolution.
- User could pan sensor, using a control in the upper-right corner of the sensor view, through six discrete steps to obtain 360° vision.
- No extended view banners (see configurations B–D for banner concepts).
- Full-screen mode: no small portals occupying screen space (see configuration C for small portal mode).

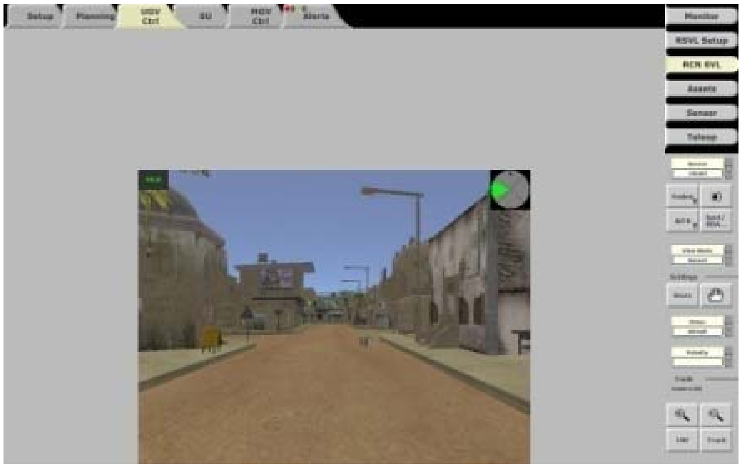

Figure 6. Condition A – sensor only: full-screen mode.

Condition B = sensor + single banner: full-screen mode (figure 7).

- 64 × 48 FOV sensor + 180° hFOV forward facing (top) banner.
- User could pan single sensor (not banner) through six discrete steps to obtain 360^{o} vision.
- Full-screen mode: no small portals occupying screen space.

Figure 7. Condition B – sensor with single banner: full-screen mode.

Condition C = sensor + single banner: small portal mode (figure 8).

- 64 × 48 FOV sensor + 180° hFOV forward facing (top) banner.
- User could pan single sensor (not banner) through six discrete steps to obtain 360^{o} vision.

- Small-portal mode: placeholder for small portals on the left side of the screen provides smaller space for imagery. For the experiment, this area was left blank.

Figure 8. Condition C – sensor with single banner: small-portal mode.

Condition D = sensor + two banners: full screen (figure 9).

- 64 × 48 FOV sensor + 180° hFOV forward-facing (top) banner = 180° hFOV rear-facing (bottom) banner.
- User can pan single sensor (not banners).
- Full-screen mode – no small portals occupying screen space.

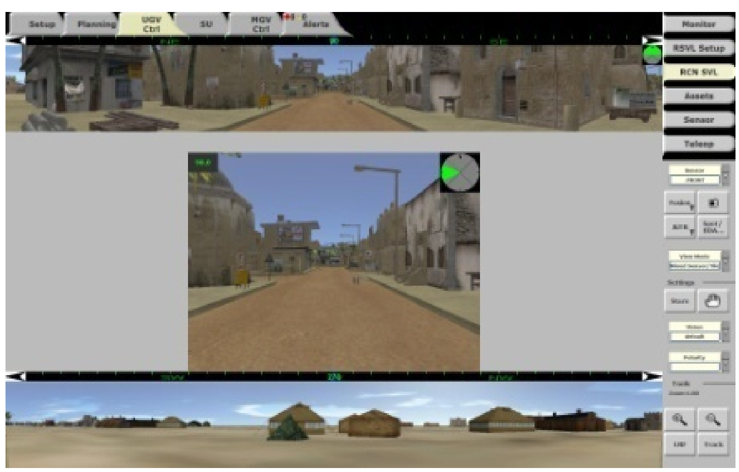

Figure 9. Condition D – sensor with two banners: full-screen mode.

Each level of the independent variable was completed one time, meaning that participants experienced each screen configuration once. In order to prevent confounds due to learning and/or familiarization effects, four separate, but statistically similar mission scenarios were created. Assignment of each display condition to a mission scenario was counterbalanced and the order of condition presentations was randomized across participants (see table 2).

Table 2. Participant-condition-mission pairings.

Participant	Mission 1	Mission 2	Mission 3	Mission 4
1	B	C	A	D
2	C	D	B	A
3	A	B	D	C
4	D	A	C	B
5	B	C	A	D
6	C	D	B	A
7	A	B	D	C
8	D	A	C	B
9	B	C	A	D
10	C	D	B	A
11	A	B	D	C
12	D	A	C	B
13	B	C	A	D
14	C	D	B	A
15	A	B	D	C
16	D	A	C	B
17	B	C	A	D

3.1 Mission Objectives

Each mission was 12 min in duration, with 8 min spent in the urban core and 4 min in outskirts of the city. The objective was to identify targets while either stationary or on the move through the use of the sensor systems on board the simulated vehicle. Prior to each mission, the experimenter performed scenario specific setup and initialization. After the completion of each mission, the experimenter performed data collection (i.e., saving log files) while the participant filled out the NASA-TLX questionnaire.

Once a target was identified the participant had to send a threat report via a report panel interface. For each threat, the participant had to enter the object type (armed human, unarmed human, or improvised explosive devise [IED]) and location (vehicle-relative clock position in integer increments from 1 to 12). The threat report options and interface are shown in figure 10.

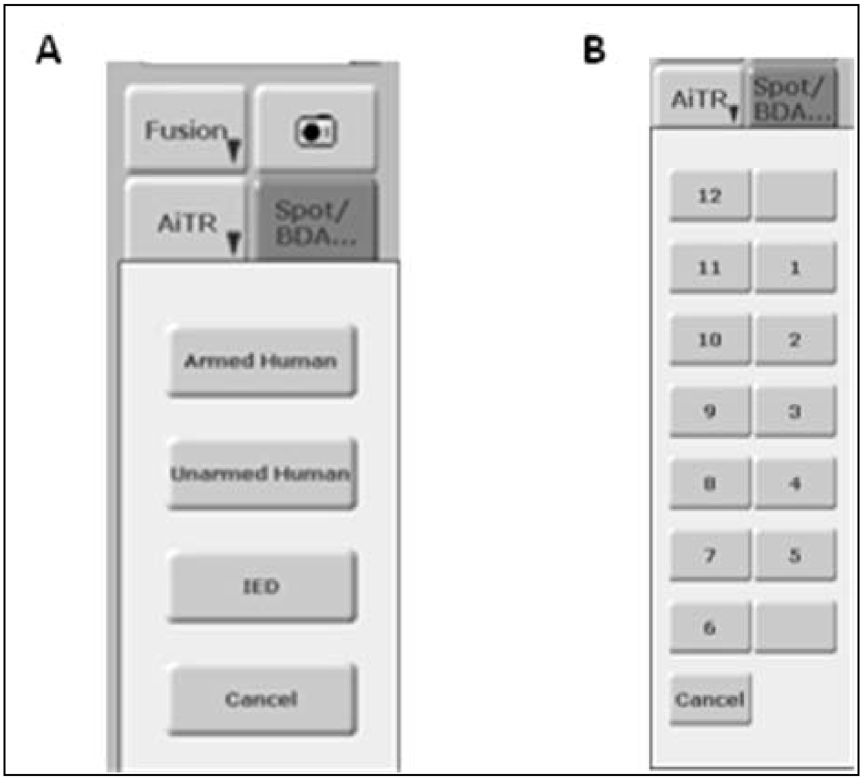

Figure 10. SA report panel for selection of target type (A) and target location (B).

3.2 Targets and Events

The events and targets were developed based on discussion with two subject matter experts on current operations in the Middle East. In addition, significant independent research was conducted involving review of materials from several sources including Soldier blogs from the internet, current periodicals and news sources regarding present-day military activities, U.S. Army photo archives, and formal documents such as U.S. Army field manuals and other such operational/doctrinal materials. Figure 11 shows examples of the presented targets.

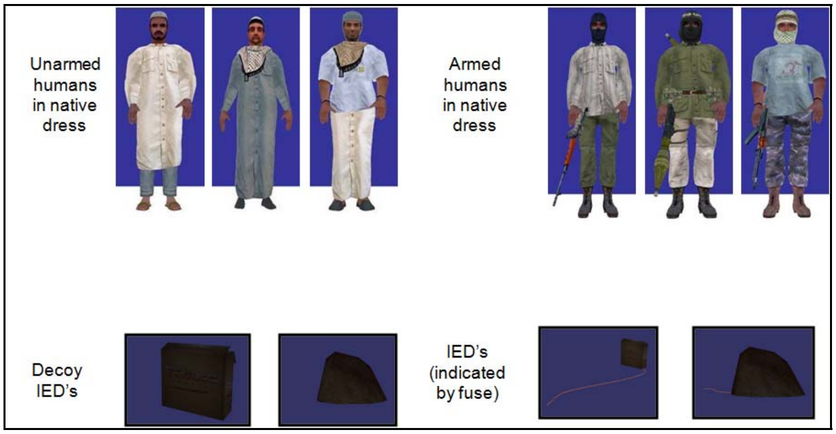

Figure 11. Sample entities used for populating mission scenarios.

Participants were required to report targets that were either armed humans, IEDs, identified high-value targets (HVTs), unarmed individuals performing suspicious behaviors, or unarmed humans detected in the city outskirts (which were designated as free-fire zones). All types were designated as a threat and reported as either an unarmed human, armed human, or IED. Non-threats were not reported. A description of the events and entities as well as the total number of each occurrence within each mission is provided in table 3.

The information in table 3 describes a set of over 100 basic events, with a total of 38 threat events that should have been reported per scenario. However, in the first scenario one of the armed humans never became visible (due to variable scenario dynamics) and thus, in that scenario only, there were 16 (as opposed to 17) armed humans.

Table 3. Summary of entities and events comprising the mission scenarios.

Label	Description	Constituent Entities/Events	No. in Each Scenario
Crowd	Group of 10+ non-threatening humans	Market, protest, children playing, complaints to local sheik, going into mosque, clinic, hospital	1
Hidden IED	Various objects with large wires trailing out, included in both urban core and outskirts	Dirt piles, defunct vehicles, small electronics, etc.; 1 hidden IED was specified as an HVT	8
Decoy IED	Various objects without large wires sticking out	Dirt piles, concrete barriers, concrete piles, defunct vehicles, carpets, etc.	~50
High-value target (HVT)	Targets that are not threatening until radio communiqué warning of danger	Vehicles, people, or objects meeting specific descriptions (i.e., IEDs being made from broken televisions)	3
Vehicle stop	Instances where vehicle motion pauses	Two stops in urban core (1 near suspicious formation of people) and 2 stops in the outskirts	4
Suspicious behavior	Unarmed humans behaving in threatening manner	Coordinated movement of people along multiple axes or individuals staring at vehicle as if spotting for IED detonation	5
Ambush	Group of humans that remained concealed until vehicle was near; no engagement	Armed humans in varying numbers; 1 ambush in urban core and 1 in outskirts	2
Cut off	Blockage of nearest escape or main route	Vehicle, road closed signs, concrete barriers; 1 in urban core and 1 in outskirts	2
Armed human	Humans visibly carrying weapons	Armed humans carrying any of an array of weapons, all were large and visible (RPG launchers, machine guns, etc.)	17
Unarmed human I	Unarmed humans that were considered threats	Unarmed humans behaving in a threatening manner or indicated as HVTs	13
Unarmed human II	Unarmed humans that were not considered threats	Regular humans in native dress that either remained static or were moving along a path not directed in coordinated fashion towards the vehicle (i.e., could not be confused with a suspicious formation)	~10–20

3.3 Data Collection

As previously described, a variety of data were collected during this experiment. The objective measures were based on data contained in table 4.

Table 4. Summary of data artifacts generated during the current experiment.

Source	Description	Sampling Rate	Contents
Crew station log	Data for all WMI button presses	Asynchronous, event-driven	• Simulation time • Unique button identifiers
Sensor view log	LOS detections for each environmental entity, per each of the six sensor views	4 Hz	• Simulation time • Unique entity identifier • Location • Heading • Speed • Sensor from which entity has LOS
Trip line log	Data for all trip line crossings	Asynchronous, event-driven	• Simulation time • Unique trip line identifier
Vehicle status log	Vehicle state data	100 Hz	• Simulation time • Distance travelled • Speed (current) • Acceleration (current) • Location • Orientation • Heading
Smart Eye data log	Eye position data	60 Hz	• Simulation time • Head position (6D) • Eye position (3-D) • Gaze orientation • Gaze-screen intersection coordinates (x, y)

The subjective measures corresponded to the information provided by the NASA-TLX, the usability assessment, and post-experiment exit interview.

3.4 Dependent Variables

All dependent measures were calculated from a reduced, collated, and time-synchronized set of variables extracted from the raw data (i.e., the data logs) set. All data processing and synchronization were handled by a custom written program called the Data Analysis and Reduction Tool (DART). Beginning with merging all data logs into a single binary format, all events and entity characteristics were codified and subsequently collated using a common time stamp (called "simulation time"). Before any dependent variables could be calculated, the merged and collated data for each event had to be associated with each SA report. That is, for each SA report, it had to first be established which of the many entities appearing on the screen was the subject. In cases where there was no scripted threat available to correspond with a report, a false alarm was noted. To facilitate the process of verifying report-entity associations, an automated algorithm was applied and its output was verified manually by trained experimenters. For the sake of simplifying the presentation, the SA report-entity association algorithm is described in detail in appendix E. Following the application of the report-entity association algorithm, the following variables were calculated from the reduced data:

- Task Performance
  - Threat Detection – Was a target detected? Yes or no?
  - Reaction Time – Time between target onset and initiation of the threat report.
  - False Alarms – Number of non threats reported as threats.
  - Accuracy – Calculated based on target type and location accuracy.
    - Overall Accuracy - Calculated as (target type accuracy + clock position accuracy)/2.
    - Target Type Accuracy.
      - 1.0, if completely correct.
      - 0.5, if participant entered armed human and the target was a unarmed human, or vice versa.
      - 0.0, if participant entered armed/unarmed human and the target was an IED, or vice versa, or if it was not a target at all.
    - Clock Position Accuracy.
      - 1.0, if reported clock position was within 1 clock increment of the correct value.
      - 0.0, if reported clock position was more than 1 clock increment from the correct value.
  - Sensor Usage.
    - Proportional Usage – Percent of total mission duration spent looking at each sensor view.
    - Number of Sensor View Changes – Total number of changes between the six sensor views.
- Physiological Measurement
  - Eye-tracking data throughout completion of each mission scenario.

- Subjective Responses
  - Workload measured with the NASA-TLX.
  - Usability assessment and exit interview.

4. Analyses and Summary of Statistical Results

Because of the multivariate nature of the data collected in the current experiment, an analytic method was needed that allowed the use of as much data in a non-reduced format as was possible; that is, to the extent possible, data were examined in raw-reduced format (i.e., per individual threat presentation events) without averaging across threat presentations within different levels of each of the independent variables. For the current experiment, this meant using the success or failure on each threat event presentation, reaction time, and accuracy as the dependent variables. The various threat, participant, and environment characteristics (scenario, display condition, threat type, participant type, etc.) served as potential independent variables. Owing to the binary nature of the dependent variable for threat detection (detected or not), logistic regression was employed as the primary analytic tool. For both reaction time and accuracy, linear mixed model analyses were conducted. For both types of statistical models, the analysis proceeded in two stages, beginning with model building and concluding with the application of the selected model for within-subject analysis of performance.

For logistic regression modeling, the model building stage began by computing estimates on the most complex model, herein referred to as the full model, and then, through the use of partial sums-of-squares *F*-tests (Draper and Smith, 1998), decisions were made as to which effects could be excluded. The basic methodology was executed as follows. For each parameter (independent variable) considered for inclusion in the final model, the logistic regression was re-calculated with a reduced model that excluded its influence and then the residual sums of squares of the reduced model was compared with the residual sums of squares observed with the full model. If the exclusion of the parameter under consideration resulted in a significant loss of explained variance (i.e., if the increase in the residual sums of squares over that observed with the full model was significant as judged by an *F*-test), then it was considered evidence that the parameter should be included in the model. This process was repeated iteratively until all of the terms in the full model were assessed. After the full model was examined in this way, simpler models were assessed in the same manner and this process was repeated until a model was obtained in which all terms were significant. Once a sufficiently simple model was obtained, some conceptually motivated three-way interactions were added to explore potential higher-order relationships among the variables. Because of the sheer number of repeated *F*-tests (36 on the full model for threat detection alone), the criteria for significance was set fairly stringently based on a Bonferroni correction. Specifically, for the first level assessment, the *p* value used as the threshold for inclusion was 0.0014 and as subsequent model building passes were completed on successive models, the inclusion criteria became more stringent. All terms included in the final logistic model were significant to a level of $p < 0.0006$. Of course, for cases where interaction terms were included, all main effects for that term also remained in the model even if they were non-significant. The final logistic regression analysis then applied the simplest selected regression model to the threat detection data while nesting all terms and interactions in the participant variable; effectively creating a within-subjects logistic regression.

For the linear mixed models, a similar model building approach was employed. In essence, the same concept was applied, involving first a computation of the full linear mixed model (with all possible dependent variables and covariates included); the initial mixed linear model was run for all the dependent measures of interest which included condition, threat target type, target range, vehicle mobility, and target environment as predictors. Then, in backwards elimination fashion, subsequent models were assessed with fewer parameters. Instead of using partial sums of squares *F*-tests as with the logistic approach, the linear mixed model analysis simply worked backwards from the full model by eliminating terms that were non-significant. In the case that significant interactions were observed, all associated main effects terms remained in the model (even if they were non-significant). All of the mixed models, participant type (civilian or military), scenarios (1–4), and viewing time (threat presentation time) were entered as covariates and all tests were run as within-subject designs. The reported significant effects were from the final linear mixed model.

For sake of simplicity, all results presented herein were based on the final statistical models, those which only contained significant predictors and interactions. For more details regarding the initial full models for each analysis, as well as for more details regarding the predictor variables and their selection, please refer to appendix F (overall operator threat detection analysis).

Based on the results of the model building stage, a final model for the logistic regression included the following terms: participant type, display condition, threat target type, threat location, vehicle mobility, target mobility, range, inter-threat interval, viewing time, condition × location, large type × location, target type × range, location × target mobility and location × viewing time. Table 5 summarizes the results of the final logistic regression model.

Table 5. Summary of the final logistic regression model.

Source	**SS Difference**	**Numerator (Degrees of Freedom)**	**$F(df_{reg}, 2538)$**	**Significance Level**
Participant type	43	1	46.577	0.0000
Condition	128	6	23.248	0.0000
Threat target type	74	6	13.489	0.0000
Threat location	96	8	13.019	0.0000
Vehicle mobility	21	1	23.137	0.0000
Target mobility	59	2	32.183	0.0000
Minimum range after onset	460	3	167.221	0.0000
Inter-threat interval	19	1	20.445	0.0000
Viewing time	262	2	142.743	0.0000
Condition × location	36	3	12.993	0.0000
Threat target type × location	30	2	16.456	0.0000
Threat target type × min range	30	2	16.480	0.0000
Threat location × target mobility	46	1	50.150	0.0000
Threat location × viewing time	31	1	33.923	0.0000

The final statistical model for reaction time included significant terms for: range, vehicle mobility, and threat location as well as several interactions. Shown in table 6, the significant interactions with range included threat target type, threat environment, vehicle mobility, and threat location. Interactions with threat target type, beyond its interaction with range, included threat environment and threat location (note that the interaction with vehicle mobility was non-significant). Additional interactions included threat environment × threat location, vehicle mobility × threat location, and range × threat target type × threat environment.

Table 6. Final linear mixed model results for reaction time.

Source	Numerator (Degrees of Freedom)	Denominator (Degrees of Freedom)	F	Significance Level
Range	2	1213	13.971	0.000
Threat target type	2	1213	1.830	0.161
Threat environment	1	1213	0.313	0.576
Vehicle mobility	1	1213	14.337	0.000
Threat location	1	1213	5.021	0.025
Range × threat target type	4	1213	10.718	0.000
Range × threat environment	2	1213	21.992	0.000
Range × vehicle mobility	2	1213	5.208	0.006
Range × threat location	2	1213	11.878	0.000
Threat target type × threat environment	2	1213	9.643	0.000
Threat target type × vehicle mobility	1	1213	1.894	0.169
Threat target type × threat location	2	1213	3.436	0.032
Threat environment × vehicle mobility	1	1213	1.422	0.233
Threat environment × threat location	1	1213	8.083	0.005
Vehicle mobility × threat location	1	1213	6.131	0.013
Range × threat target type × threat environment	4	1213	6.786	0.000

As shown in table 7, the final model for analysis of accuracy included significant effects for range, threat target type, and vehicle mobility. Two interactions were also observed as significant, including range × threat target type and threat target type × vehicle mobility.

Table 7. Final linear mixed model results for accuracy.

Source	Numerator (Degrees of Freedom)	Denominator (Degrees of Freedom)	F	Significance Level
Range	2	1230	5.350	0.005
Threat target type	2	1230	53.962	0.000
Threat environment	1	1230	0.013	0.908
Vehicle mobility	1	1230	11.117	0.001
Threat location	1	1230	2.893	0.089
Range × threat target type	4	1230	14.546	0.000
Threat target type × vehicle mobility	1	1230	31.049	0.000

Although few effects attributable to the primary independent variable, condition, were significant, an initial discussion of the results will focus on its influence in the context of human performance. Subsequently, data regarding other factors influencing threat detection performance are also presented to allow for a thorough understanding of the domain within which technology may positively influence human performance on a local SA task.

4.1 Effect of Interface Configuration

Overall, the data from the current experiment provided scant evidence of an influence of display condition on threat detection performance. As will be discussed shortly, it appeared as if performance varied based on participant, environment, and target characteristics more than based on the influence of a particular display configuration. One effect, however, provided some insight suggesting that the influence of display condition differed contingent on the location at which a given threat was initially presented. Specifically, the condition × location interaction, plotted in figure 12, appears to have been influenced by variations in threat detection performance to targets presented to the rear of the vehicle.

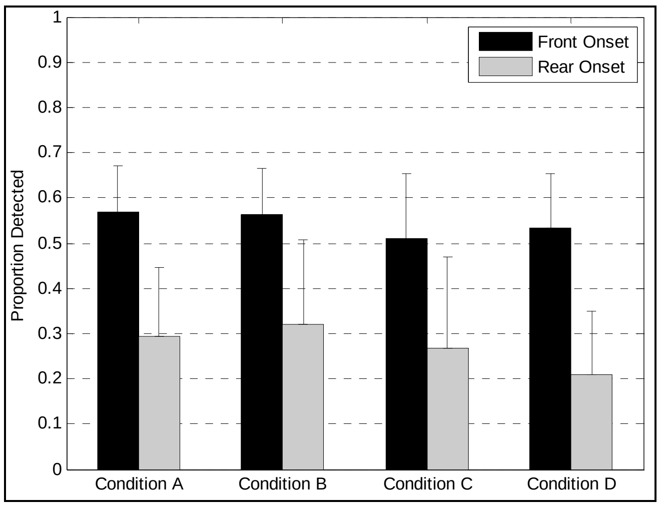

Figure 12. Significant condition × threat location interaction effect for threat detection rate.

The conclusion that detection performance for targets presented to the rear of the vehicle affected the efficacy of display condition, while not overtly apparent from the graphical evidence in figure 12; it is supported by the data presented in table 8. From the data that are shown, it seems that the variation in the amount of performance change due to threat location was driven by how much detection rates dropped when threats were presented to the rear of the vehicle. This trend can be seen in the column labeled Δ% in table 8. The largest decrease in detection performance between targets presented in the front and rear was seen in C and D (31% lower for rear targets), followed by condition A (28% lower). Given that the least change in performance was observed in condition C (24% lower), which was the condition with the worst performance on targets presented in the front, it could be argued that condition B was the best by a small margin. That is, condition B was associated with the second highest forward threat detection performance (nearly equivalent to condition A) and yet it also had the best detection performance when examining targets to the rear.

Table 8. Threat detection differences revealed in a significant condition × threat location interaction.

Condition	Threats in Front			Threats in Rear			Δ%
	Detections	Presentations	%	Detections	Presentations	%	
A	284	500	56.80	40	140	28.57	28.23
B	286	507	56.41	40	132	30.30	26.11
C	257	505	50.89	37	136	27.21	23.68
D	269	505	53.27	30	135	22.22	31.05
Mean			54.34	Mean		27.08	27.27
St. deviation			2.792	St. deviation		3.475	3.133
Coefficient of variation			0.051	Coefficient of variation		0.128	0.115

Unlike threat detection rates, linear mixed model analyses showed no significant difference in reaction time or accuracy performance between display conditions, *p's* > 0.10. Interface condition did not significantly interact with any predictor variables of interest, *p's* > 0.10. Table 9 shows the mean accuracy for each configuration by threat target type. Table 10 shows the mean reaction time organized in the same manner.

Table 9. Mean (±Standard Error [SE]) accuracy performance by display condition.

Interface Configuration	Accuracy Armed Humans Targets	Accuracy Unarmed Human Targets	Accuracy IEDs
A	74.5 (2.65)	89.4 (3.21)	88.7 (4.62)
B	70.4 (3.18)	87.3 (3.36)	89.1 (3.73)
C	78.8 (3.42)	84.9 (4.19)	90.9 (3.60)
D	70.6 (3.44)	82.5 (3.97)	88.6 (3.14)

Table 10. Mean (±SE) reaction time performance by interface configuration.

Interface Configuration	RT Armed Humans Targets	RT Unarmed Human Targets	RT IEDs
A	12.9 (0.68)	10.0 (0.82)	18.6 (1.18)
B	11.24 (0.81)	10.2 (0.86)	16.7 (0.95)
C	12.2 (0.87)	10.8 (1.07)	18.0 (0.92)
D	15.03 (0.88)	11.51 (1.01)	18.9 (0.80)

Though the statistics did not reveal significant patterns for conditions, visual inspection of the means did suggest some interesting patterns that may have been eclipsed by variation due to other factors. The presence of banners in condition B appeared to be associated with shorter reaction times to targets, which was especially evident for armed humans.

In order to gain further insight into how the users interacted with the different display conditions, analyses were conducted on the number of times the operator changed the sensor views (see table 11). Results showed that there was a significant difference between configurations in the number of sensor view changes, F (3, 64) = 10.95, $p < 0.001$. Paired comparisons of the interface configurations revealed that configuration A had significantly more changes than all of the other configurations, p's < 0.001. There was no significant difference between configuration B, C, and D, p's > 2. For example, there were 55% more sensor changes in the condition with only a sensor view (A) than one where the operator had at least one banner (B–D). These data suggest that the presence of banners influenced the efficiency in the utilization of the simulated 360° SA system. That is, display conditions with banners revealed similar detection performance (especially in terms of reaction time and accuracy), but at least half the workload (screen interactions) as the sensor view only condition.

Table 11. Mean (±SE) sensor changes by display condition.

Display Condition	**Total Sensor View Changes**	**Sensor Views per Minute**
A	381.5 (33.7)	28.2
B	171.4 (33.7)	12.6
C	189.2 (33.7)	13.8
D	131.0 (33.7)	9.6

In a similar manner, analyses focused on the proportional use for each sensor view also revealed a display condition effect. Shown in figure 13, this effect was manifested in the proportion of time spent looking to the front sensors versus the rear sensors that varied systematically across display condition. Specifically, it appeared that in both conditions A and D, participants tended not to pan the sensor around to the rear camera views whereas in conditions B and C, they more evenly distributed their viewing time throughout the 360° set of views (though remained biased to look straight forward and backward rather than to the sides). That is, in the conditions with only a forward-oriented banner, participants appeared to use the sensor portal to view the rear of the vehicle during a greater proportion of the missions.

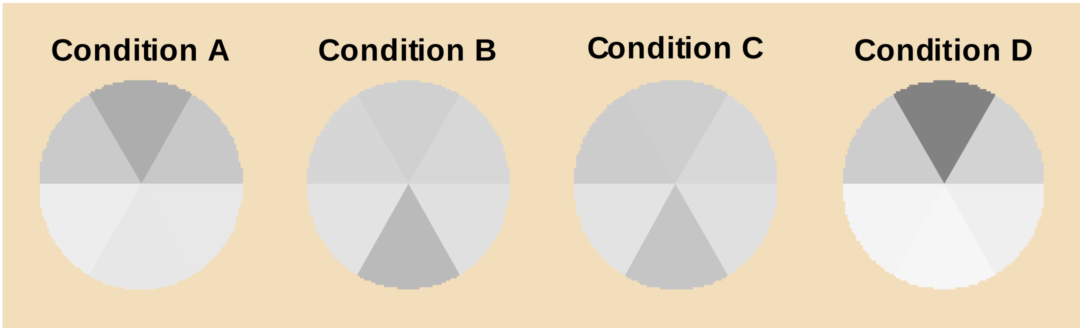

Figure 13. Proportional use of the six sensor portals across the four display conditions. Darker shading indicates a viewing proportion closer to 1.0.

Summarily, there were three important (and related) pieces of information to be gleaned from analysis of the effects in performance due to display condition. First, threat detection performance varied based on whether targets were presented to the front or the rear of the vehicle. More to the point, however, was that certain conditions (B, C) seemed to be associated with a lower rate of performance degradation for rear targets than the other conditions (A, D). In complimentary fashion, the second observation was that participants tended to work harder, as indexed by number of sensor view changes, in condition A as compared with all other Conditions that included the use of banners. Finally, the third effect noted was that participants used their sensor to look to the rear of the vehicle more frequently in the two display conditions in which they were given a forward-oriented banner (B and C). This increased usage of the sensor portal in this manner may, in part, explain why performance degraded less in these two conditions with respect to the viewing of threats presented to the rear of the vehicle.

4.2 Effects of Participant, Environment and Target Characteristics

Beyond the performance effects just outlined, many of the other factors (discussed in appendix F) appeared to exert strong influences over the various aspects of threat detection performance that were studied. It is estimated that these effects are important because: (1) they proved stronger indicators of performance than the primary independent variable (condition) and (2) they may have indicated important areas in which future technology design and development needs to compensate for potential deficiencies in human perceptual-cognitive performance.

As far as baseline threat detection capabilities (whether an operator can distinguish between a real threat and a decoy), it seems the strongest performance indicators other than display condition involved variables related to threat target type, range and location of threat onset, and time available to view and process visual information. Additional factors, such as participant type (civilian, Soldier), vehicle mobility, target mobility, and inter-threat interval also appeared to affect specific aspects of threat detection performance (i.e., detection rates vs. reaction time, etc.).

As shown previously in table 5, there were three main effects for threat detection rates that were not involved in any interactions. The participant type main effect appears to have been driven by Soldiers having a higher threat detection rate than civilian participants. Overall, the Soldiers detected 574 out of 1056 potential threatening events, which was a 54.4% detection rate. By comparison, the civilian participants detected 669 out of 1504 potential threatening events, a detection rate of 44.5%. Similarly, a main effect was also observed for vehicle mobility. The detection rate for targets presented when the vehicle was on the move was 50.91% (1140 out of 2239 threatening events) as compared with 32.09% (103 out of 321 threatening events) for targets presented when the vehicle was stationary. As with the effect for participant type, absence of vehicle mobility interaction effects suggests that this difference between detection rates when the vehicle was stationary or on the move was consistent across all levels of other independent variables. The final threat detection rate main effect was due to a tendency for detection performance to increase (t_{2538} = 4.434, p < 0.001) as inter-threat interval increased across a range from ~1.5 s to ~90 s. That is, with more time in between threat detection events, threat detection rate increased slightly. In addition to the performance differences due to front versus rear-presented targets as in the condition × threat location interaction previously discussed, a significant threat target type × threat location interaction revealed a clear additional effect of target type on threat detection.

Shown in figure 14, it seems that IEDs were always detected more frequently than they were missed, although presentations of IEDs to the rear of the vehicle were rather infrequent (constituting 1.33% of all targets presented). For both front and rear presentations of IEDs, detection rates were high (IEDs in front = 78.43%, IEDs in rear = 88.24%) compared with those for human threats. However, detection differences existed between armed and unarmed humans as well. That is, detection rates always appeared below 50% for armed humans (armed humans in front = 39.14%, armed humans in rear = 29.14%) whereas the detection rate for unarmed humans was higher when presented in the front (55.57%) and declined more sharply when presented behind the vehicle (15.58%).

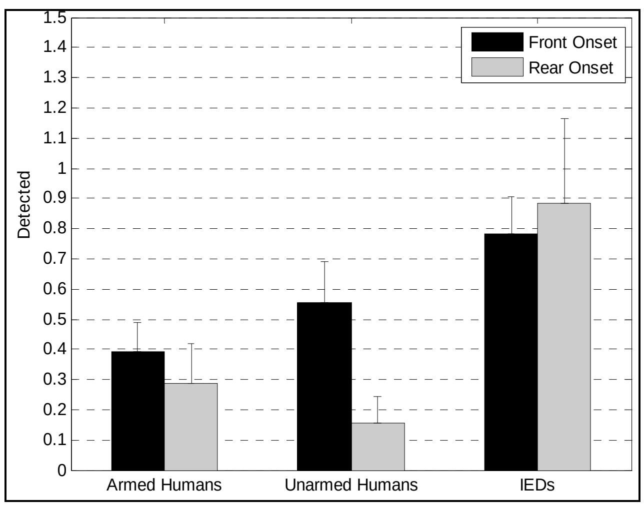

Figure 14. Threat target type × threat location interaction.

Based on mathematical analyses of sensor characteristics, it was expected that detection performance would also be reduced as a function of the range at which the threats were presented. This was supported by the data. Detection performance appeared to decline as threat range increased, particularly for human targets. The threat target type × range interaction was most likely to have been driven by a lack of longer range observations for IED threats. In figure 15, all IED targets appeared to onset and be maintained within a fairly close range of the vehicle, at or below 25 m, whereas the human targets were distributed across a range from 5 to 180 m.

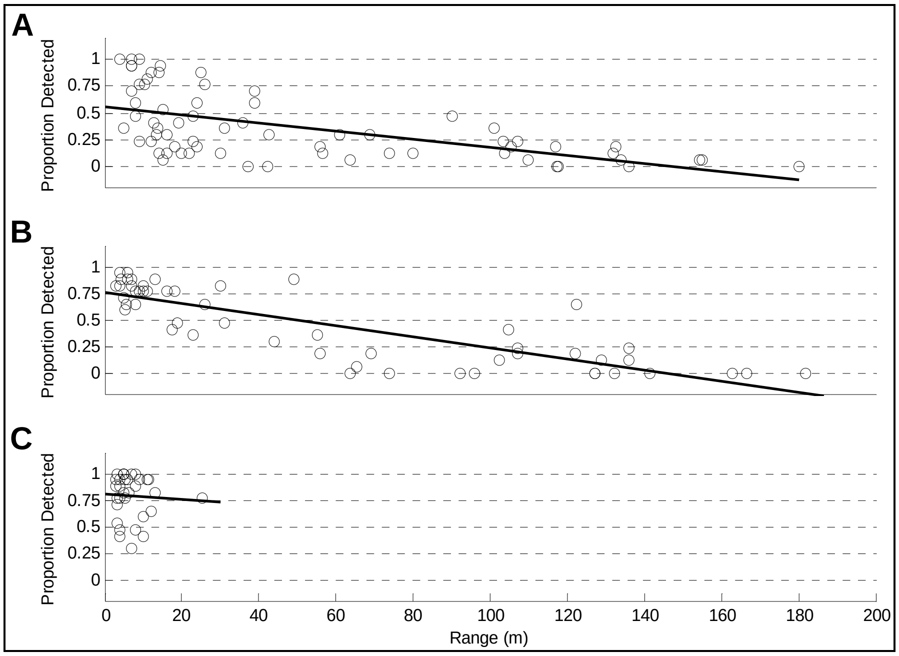

Figure 15. Average threat detection performance as a function of threat target type and range. Target types include (A) armed humans, (B) unarmed humans, and (C) IEDs.

Some evidence was also present for an interaction between armed and unarmed human threats. There appeared to have been fewer observations of low detection performance at close range for unarmed as compared with armed humans (compare figures 15A and 15B at ranges below 40 m, especially examine the left-most portion of the linear fits). This observation was likely a function of a design difference in what made armed and unarmed humans considered “threats”; unarmed humans were considered threatening by their behavior (either staring at or moving towards the vehicle), whereas armed humans were considered threats by virtue of carrying a weapon. However, the fact that unarmed humans were defined as “threatening” by moving towards the vehicle while an armed human could be standing still or even moving away from the vehicle and still be considered a threat points to a possible differential in target salience.

Part of the difficulty in resolving whether such effects were due to target salience is that range exerted a strong influence over performance and thus represented one potential major confound. For example, additional data that would seem to speak to a target salience effect would be a case where moving threats were detected at a higher rate than stationary ones. Salience in such a case would be assumed to go with the target motion; moving entities were likely more attention grabbing than non-moving entities. Yet, the data obtained for the current experiment revealed just the opposite of what would be expected. As shown in figure 16, moving targets were actually detected at a lower rate, on average, than were stationary ones.

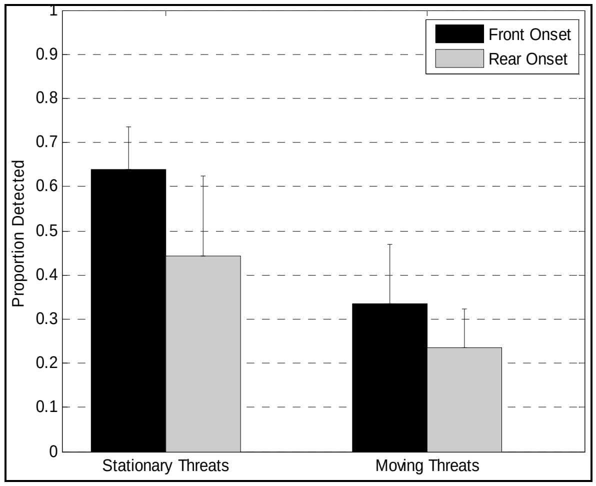

Figure 16. Threat detection rate as a function of threat mobility and threat location.

The reason this effect may represent a potential confound, rather than a counter-result (counter to what was expected), is that observations at different range values were unevenly distributed across the threat mobility factor. In short, there were more threats presented at shorter ranges for stationary targets than for moving targets. This is shown in figure 17, which presents histograms indicating the number of target observations across range within threat mobility and threat onset location factors. Explicitly, because there was a strong tendency for threats presented at shorter ranges to be detected at a higher rate, and because more of the stationary threats were presented at shorter ranges (such as in the urban core), it becomes less surprising that stationary threats were detected at a higher rate than were moving threats. A similar confound was observed for the previously-mentioned main effect for vehicle mobility a greater proportion of threats presented while the vehicle was moving were at shorter ranges than threats presented while the vehicle was stationary, thus explaining part of the improvement in detection performance when the vehicle was moving as compared to when it was stationary (yet another unexpected result).

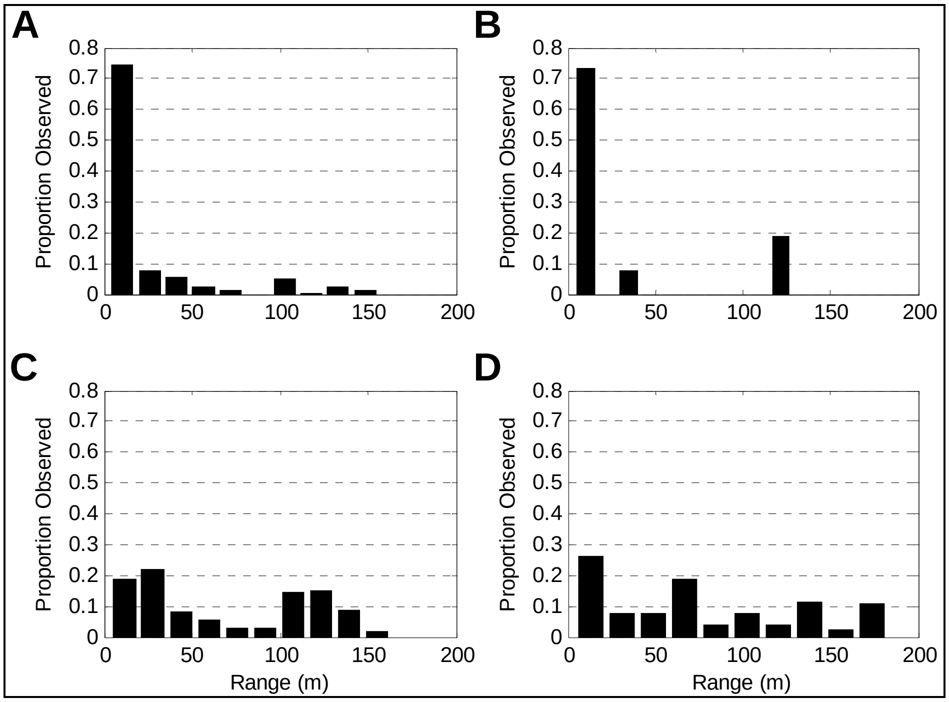

Figure 17. Distributions of threat presentations across range for stationary threats (A, B) and moving threats (C, D). Left panels show front-presented threats and right panels show rear-presented threats.

As previously shown (table 6), for reaction time, the final linear model included threat environment, threat target type, threat location, range, and vehicle mobility as significant predictors of performance. Of the more complicated effects was a three-way interaction of threat environment × threat target type × range. To resolve the interaction, linear models were conducted within each target range. All permutations to resolve this interaction were significant, *p's* < 0.001. Visual inspection of the mean as displayed in figure 18 showed two general patterns of results. For those targets that recognition was relatively less challenging (armed humans carrying large guns in the urban core, IEDs with long, large fuses in both environments), there was a clear scaling of reaction time across range from near to far. That is, for relatively easily recognized threats, reaction time simply increased as range increased. However, for targets where detection was a little more challenging, such as unarmed humans in the urban core or all humans in the outskirts, there seemed to be an optimal range for threat identification wherein reaction time was the fastest to those threats presented at a middle range and increased for threats presented very close to and very far from the vehicle. While entirely speculative, it is likely that this interaction reflects different limitations on the perceptual and cognitive processes associated with simple versus complex threat detection events.

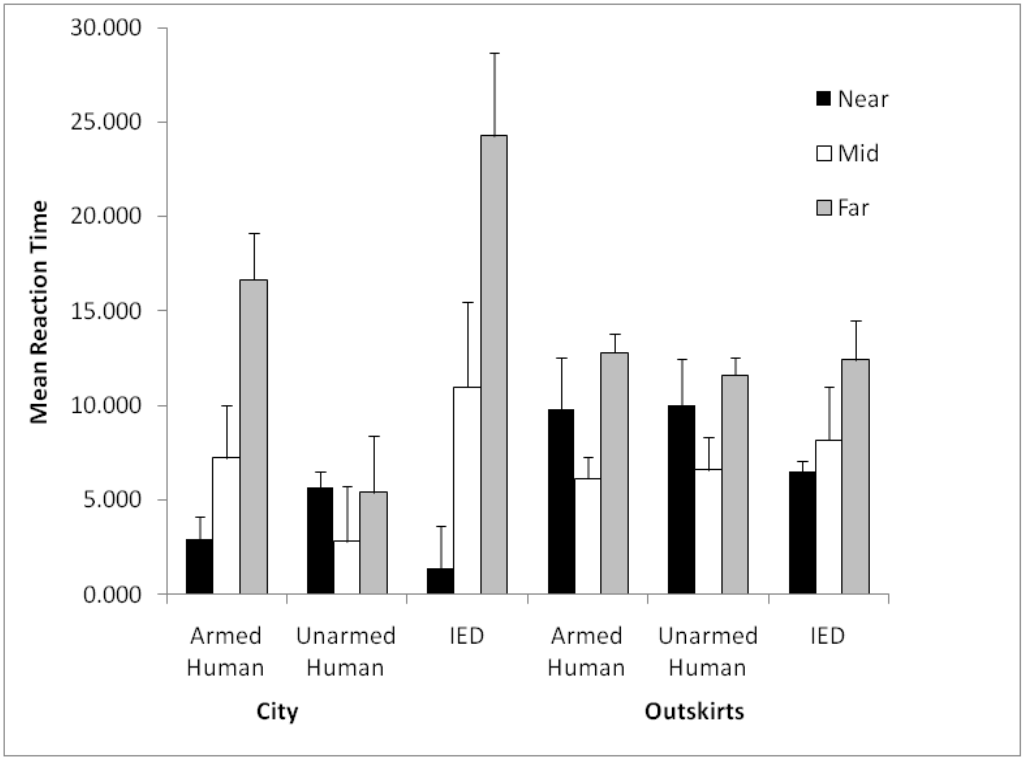

Figure 18. Mean reaction time for target type, range, and threat environment.

As shown earlier in table 6, additional interactions were observed as a function of threat range. To resolve the range × threat location interaction, linear models were conducted, selecting for target range. For mid- and long-range targets, there was a difference between front and rear target locations. More specifically, the increase in reaction time observed across range was considerably greater for threats presented to the front than for threats presented to the rear (table 12). This result likely relates to the sensor usage. Views to the rear of the vehicle were less frequent than views to the front of the vehicle. Moreover, detection rates were lower for threats to the rear. However, it is reasonable to infer that if a threat was successfully detected in the rear of the vehicle, it was probably detected relatively quickly because the consequence of waiting longer to make a decision (which would result in a longer reaction time) for a target to the rear is that, due to vehicle motion, the threats get further away. Conversely, participants may have waited a little longer to make a decision to the front of the vehicle because the consequence of waiting longer was that, on average, the targets would get closer.

Table 12. Mean (±SE) reaction time for targets by threat location and range.

Target Location	Near	Mid	Far
Front	5.99 (0.15)	14.67 (0.58)	20.43 (0.67)
Rear	5.79 (0.38)	7.17 (2.01)	9.43 (1.85)

To resolve the interaction of vehicle mobility × range, linear models were conducted, selecting for vehicle mobility and range. All comparisons were significant, p's < 0.001. As target range increased, RTs increased. This increase was greater when the vehicle was moving relative to when it was stationary (see table 13). As with the result shown in figure 18, this may have reflected a strategy where, while on the move it was, on average, beneficial to wait longer to report on further away targets to the front of the vehicle because, as one waited longer the threats would, on average, get closer to the vehicle and become easier to identify.

Table 13. Mean (±SE) reaction time for targets by vehicle mobility and range.

Vehicle Mobility	Near	Mid	Far
Stationary	4.05 (0.46)	7.65 (0.89)	8.16 (0.50)
Moving	6.10 (0.26)	14.41 (0.42)	21.41 (0.48)

The linear mixed model analyses also revealed significant interactions with vehicle mobility and threat location as predictors. The following sections describe the results of the simple effects analyses for those interactions.

To resolve the interaction of vehicle mobility × threat location, linear models were conducted within target sensor location. If targets were presented in the rear sensors, there was no difference in reaction time to threats regardless of whether the vehicle was moving or not. Targets in the rear of the vehicle were more difficult to detect and identify in general and, as discussed above, responding slowly usually meant that they moved further away from the vehicle and thus there was a premium on responding quickly to threats presented in the rear of the vehicle. In contrast, for targets presented in the front sensors, detections occurred twice as fast when the vehicle was stationary ($X = 5.25$, $SE = 1.17$) relative to when it was moving (× = 11.1, $SE = 0.28$), $p < 0.001$. The results suggest that when performing threat detection with the vehicle on the move, operators might have waited to approach threats in the front before deciding on how to fill out the appropriate SA report.

To resolve the interaction of threat environment × threat location (table 14), linear models were conducted, selecting for threat location and environment. Continuing the previous pattern, all comparisons were significant, ps < 0.001. In general, if targets were detected in the rear sensors, reaction times were faster. Finally, reaction times for both front and rear sensors were higher in the outskirts than the urban core, another result that was likely confounded by the previously discussed range effect.

Table 14. Mean (±SE) reaction time for targets by threat location and threat environment.

Target Location	City	Outskirts
Front	10.31 (0.33)	11.56 (0.42)
Rear	5.74 (1.039)	7.59 (1.00)

Final linear model analyses revealed that threat target type, range and vehicle mobility significantly affected accuracy. More specifically, the final model for overall accuracy (results in figure 18) included threat environment, threat target type, threat location, range, and vehicle mobility as predictors. It revealed a significant interaction of threat target type × range for detection accuracy, $f\,(4, 1230) = 14.5$, $p < 0.001$ and threat target type × vehicle mobility, $f\,(1, 1230) = 31.0$, $p < 0.001$, the main effects of threat target type, range, and mobility were also significant, $F\,(2, 1230) = 53.9$, $p < 0.001$, $F\,(2, 1230) = 5.3$, $p < 0.001$, and $F\,(1, 1230) = 11.1$, $p < 0.001$, respectively.

To resolve the interaction of threat target type × range, linear models were conducted for each target range. There was no difference in target detection accuracy at short range, however there was a difference at mid and far ranges, $F\,(2, 272) = 11.0$, $p < 0.001$ and $F\,(2,237) = 44.8$, $p < 0.001$. In the mid range, detection accuracy regarding armed humans was significantly poorer than for either unarmed humans or IEDS, *p*s < 0.001. For long-range detection, all pairs were significantly different, *p*s < 0.001. Results are shown in table 15.

Table 15. Mean (±SE) detection accuracy for threat target type × range.

Target Range	Armed Humans	Unarmed Humans	IEDS
Near	88.0 (1.61)	88.7 (2.63)	85.4 (1.51)
Mid	72.9 (2.69)	85.1 (2.90)	87.4 (2.86)
Far	59.2 (3.26)	85.5 (5.58)	94.4 (4.31)

To resolve the interaction of threat target type × vehicle mobility (table 16), linear models were conducted for each target range. There was no difference in target detection for IEDs since these were always present when the vehicle was moving. However, there was a difference for armed and unarmed humans: $F\,(2, 272) = 11.0$, $p < 0.001$ and $F\,(2,237) = 44.8$, $p < 0.001$.

Table 16. Mean (±SE) detection accuracy for threat target type × vehicle mobility.

Vehicle Mobility	Armed Humans	Unarmed Humans	IEDs
Stationary	40.3 (5.13)	91.5 (2.64)	—
Moving	82.9 (1.32)	87.4 (1.30)	87.6 (1.04)

There were many complex, interacting factors affecting the behavioral variability observed in the present experiment. Among the variables that appeared to exert the strongest influences were characteristics of the task context. Variables included the location of a threat onset (whether front or rear of the vehicle), the range from the vehicle to the threat, and whether the vehicle was stationary or on the move. Here was a generalized reduction in performance for detecting threats that onset to the rear of the vehicle and there appeared to be a decrease in detection performance that varied as a function of range. Detection rates and accuracy decreased, on average, as the range between the vehicle and threat increased and reaction time tended to increase. However, the reaction time data suggest that a general strategy was used in terms of responding to threats that onset to the front versus those that onset to the rear of the vehicle. Specifically, when threats were detected at the rear of the vehicle, decisions were made relatively quickly because, especially when the vehicle was moving, they were likely to get further away as the mission progressed. When threats onset to the front it seems that participants may have been waiting a little bit longer to complete their SA reports; the likely interpretation is that they were waiting for the vehicle to approach the target so they could resolve its type for the SA report.

Viewing time also significantly affected performance. If participants were making their reporting decisions, in part based on how long they could view the targets, then one would expect to find a strong relationship between the threat detection performance variables and viewing time (defined as the amount of time each threat was visible on screen before it was reported; see appendix E). As such, the last aspect of the threat detection results closely assesses how viewing time factored into the results of the current study.

4.3 Effects of Viewing Time

A final interaction term that was included in the logistic regression model involved a threat location x viewing time interaction, which is plotted in figure 19. Figure 19 shows that targets presented to the rear of the vehicle were typically viewable in either a sensor or a banner for less time than targets presented in the front. The slope of the increase in detection performance with increased viewing time was greater for the targets presented to the rear of the vehicle (slope = 0.812, t_{2538} = 5.163, $p < 0.001$) than for the targets presented to the front (slope = 0.660, t_{2538} = 10.139, $p < 0.001$), indicating that participants benefited more from additional viewing time when detecting targets appearing to the rear than when detecting forward-presented targets.

There was also a positive relationship between viewing time and detection performance for both sets of target locations, confirming an expected result that given more time with the target in a sensor or banner display, participants would have a higher likelihood of detecting that target. While not profound, this result is important to consider relative to the additional dependent variables, particularly analysis of reaction time.

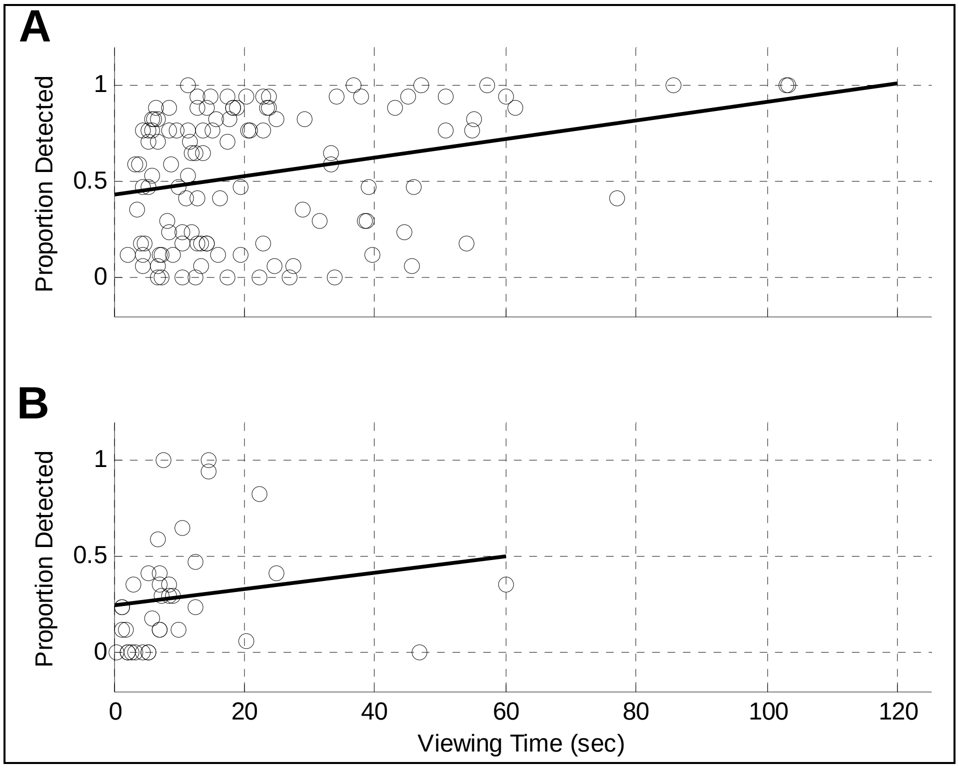

Figure 19. Threat detection performance shown as a function of viewing time and threat onset location for (A) front- and (B) rear-presented targets.

Because of the observation of a relatively strong effect for viewing time on basic threat detection performance, the linear mixed model analyses were conducted while explicitly including viewing time as a covariate. The inclusion of viewing time as a covariate was intended to be a direct acknowledgement of its influence over threat detection performance and, in particular, as a primary determinant of reaction time. This relationship is illustrated in figure 20, showing a strong correlation between viewing time and reaction time.

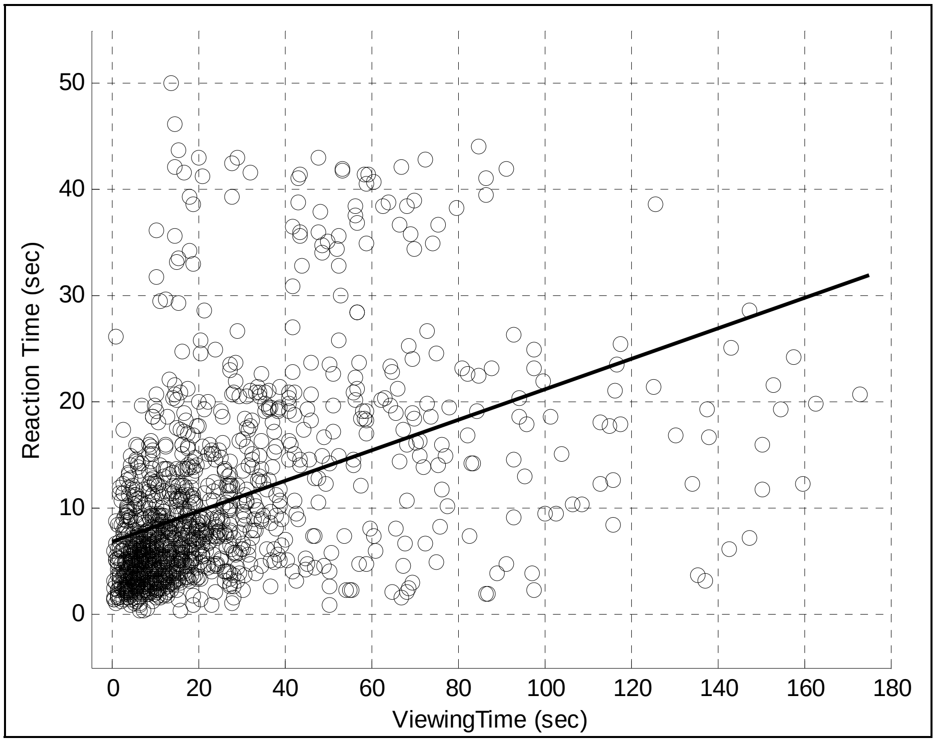

Figure 20. Reaction times for all threat detections plotted as a function of viewing time.

4.4 Concluding Remarks on Target Detection Analysis

Overall, there appeared to be a number of factors influencing the ability of the participants in the current study to detect potential threats of varying characteristics. Perhaps the most influential of these variables were the characteristics of the targets themselves. For example, factors such as location of threat onset, whether the threat was moving or stationary, and the minimum range between the threat and the vehicle following threat onset all exerted strong influences on detection performance. Moreover, there appeared to be some moderate to strong mitigating influences of other factors, such as whether the participant was a Soldier or civilian and whether or not the vehicle was stationary or moving.

Among all of the variables influencing detection performance, perhaps the most powerful predictor of whether or not a target was detected was its location at threat onset. In particular, there were many interactions observed that involved the location variable (i.e., whether the threat first appeared in the front or the back of the vehicle) suggesting an important consideration for system design. Specifically, an effective 360° indirect vision system will be one that not only easily allows (as all display conditions in this study did), but also encourages the operator to scan areas behind the vehicle as frequently as they scan the forward path. In addition to the location variable, range and viewing time also seemed to have significant influence over detection performance. That is, those threats that appeared at longer distances from the vehicle and those threats that appeared in a sensor for a shorter period of time were the ones that were most likely to be missed by the participants. Therefore, if an indirect-vision system is to be effective for threat detection, it will be necessary to improve or enhance range performance as well as maximize or optimize the amount of time the operator potentially can view the target.

Finally, some results indicated that target salience may have been an important factor affecting detection performance, although it was difficult in the current study to definitively identify its influence because of the effects with threat range and other factors. Certainly as intelligent visualization development efforts proceed, aided target/threat recognition tools may become necessary to improve performance beyond passive systems that rely solely on optimizing the images for the human visual system. Rather than making images better and hoping that the operators detection performance will improve with training and visual scanning experience, systems could be designed to actively enhance those features of the environment that make targets more salient, such as contrast enhancement and automated motion detection, especially as applied to detection and identification of humans who may or may not be potential threats to local security.

Although the analysis of performance variables such as detection rates, reaction times, and accuracy was highly informative, some of the conclusions drawn relied heavily on informed interpretation rather than direct observation. Therefore, the final sections of this report address how the display interfaces were used by the participants. In particular, information gleaned from eye movements assessed on a subset of the experimental participants is examined and then conclude the report by discussing subjective questionnaires regarding workload and user preferences.

4.5 Eye Tracker Performance

For this experiment, two infrared (IR) cameras and emitters were used. They were positioned just to the left and right of the experimental display screen, as shown in figure 21. Eye-scanning behavior was measured throughout the experiment.

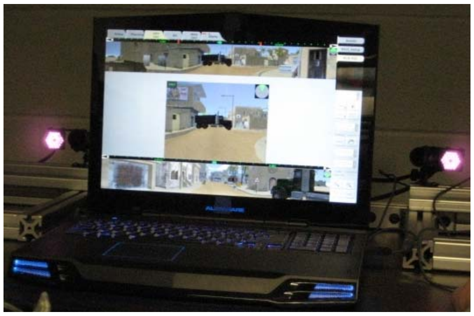

Figure 21. Smart Eye camera placement with illuminated IR emitters.

4.6 Quality Analysis and Data Processing

In order to use the acquired eye tracker data for analysis, it was necessary to assess the quality of each data set obtained. Despite the fact that eye detection and gaze estimation has received a considerable amount of attention over the past few decades and significant progress has been made, eye tracking in general remains a very difficult problem and each system/approach has associated limitations (Hansen and Ji, 2010). Some of the most prevalent limitations encountered with this work included:

1. The use of IR emitters caused unwanted distortions, reflections, and glare in participants who were wearing glasses. This made it difficult for the software algorithms to identify the pupils for participants that had glasses and greatly reduced the reliability of their associated data sets.

2. The Smart Eye system was designed to work with up to six IR cameras; however, for eye tracking, both of the participant's eyes must be visible in at least two cameras at the same time. For this research, only two cameras were available and, therefore, eye tracking measurement was difficult for participants who were either restless (i.e., tended to constantly move about in their seat) or who tended to turn their head outside of the calibrated viewing range.

3. World models were used to obtain intersections between eye gaze and display screen coordinates; however, subtle differences in calibration for each participant often resulted in slight errors, which often appeared as translational shifts in the data.

Given these limitations, a number of pre-processing steps were used to determine the quality and usability of each data set. The first step involved filtering out all data points that occurred outside of the experimental time window (i.e., prior to the start or after the completion of each condition), as well as all points in which no reading was possible due to the fact that one, or both of the participant's eyes were not visible in both cameras. The second step involved creating a heat map for each data set and then subjectively assessing each set to determine its usability. The results of this step were used to provide initial reports on participant behavior, to guide further analysis, and to provide a point of comparison for later quantitative analysis techniques.

The manual inspection of the eye-tracking data sets involved overlaying each heat map with a wire-frame model of the current experimental condition (i.e., screen layout) and then labeling that data set as high, medium, or low quality. Figure 22 shows examples of high, medium, and low quality data. For reference, eye-tracker data were considered to be of high quality if, upon visual inspection, it appeared to be virtually noise free (i.e., very few points outside of viewing regions) and there was an apparent strong spatial correlation between the individual patterns of activity and screen layout. Eye-tracker data were considered to be of medium quality if there was good spatial correlation between the observed patterns of activity and the screen layout, and if the apparent noise did not mask these patterns of activity. Eye-tracker data were considered to be of low quality if no discernable relationship could be identified between the patterns of activity and the screen layout, or if there appeared to be more noise than data. As an example, consider figure 22A, which was labeled high quality, the strongest patterns of activity can be intuitively matched to the center of the sensor view and the sensor control (i.e., the small square in the upper-right corner of the sensor view). In addition, another pattern of activity is observed over the report panel at the right side of the screen, and there were very few points that occurred outside of these three regions of interest. In figure 22B, which was labeled medium quality, there appeared to be patterns of activity corresponding to the top banner, sensor view, and report panel, but there was also a noticeable amount of noise that made these patterns appear slightly more diffuse than one would expect, given the screen layout.

Of course, it was difficult to filter or simply remove these points from the data set because it was possible that in this particular experiment the participant had trouble staying focused on the task and as a result their eyes tended to "wander" more. In figure 22C, however, there was no apparent correlation between the patterns of activity and the screen layout, and this data set was subsequently labeled as low quality.

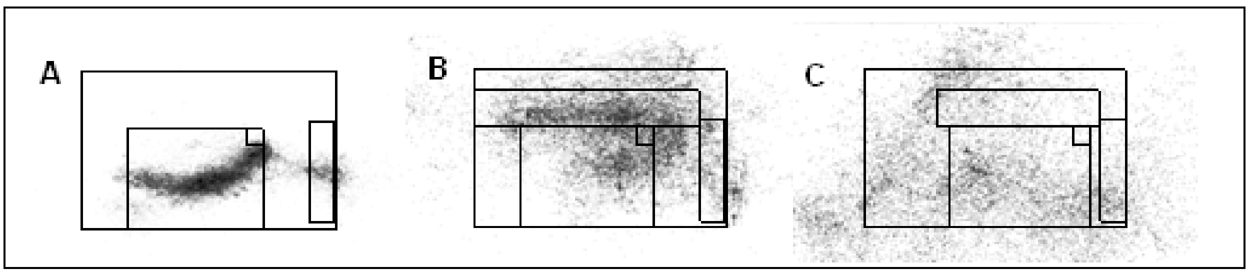

Figure 22. Heat maps of (A) high-, (B) medium-, and (C) low-quality eye-tracker data.

Once the individual missions were labeled for data quality, an overall participant label was assigned by aggregating the results over all four missions for a single participant (in a winner-take-all fashion). In other words, if a participant had three missions that were labeled medium quality and one labeled low quality, then the participant's overall rank was medium. Ties were broken manually by re-examining the data. This approach was taken because the variables most affecting data quality (i.e., participant, eye-tracker calibration, and presence of glasses) were consistent across an individual participant's four missions. Table 17 presents the results of this process.

Table 17. Results of manually ranking participant eye-tracker quality.

Participant	Rank – Manual
4	High
7	High
1	Medium
2	Medium
3	Medium
5	Medium
6	Medium
11	Medium
14	Medium
15	Medium
17	Medium
8	Low
9	Low
10	Low
12	Low
13	Low
16	Low

After each participant's data had been manually labeled, the third pre-processing step was designed to perform a less subjective analysis of data quality. However, this was a difficult process due to the unsupervised nature of the problem: there was no labeled training, or calibration, data set with which to develop a system model. Therefore, to obtain labeled training data this processing step focused on eye activity immediately prior to button presses. Because participants interacted with the system using a mouse, it was assumed that when a participant pressed a button that there was a high probability the participant had to look at the button to do so. For this analysis, the 'Spot/BDA' button was chosen due to its location on the report panel (i.e., it was spatially isolated from the rest of the screen regions) and because unlike the buttons used to indicate target type and location, the 'Spot/BDA' button had to be pressed for every report.

The algorithmic analysis process began by isolating each 'Spot/BDA' button press for each mission and then extracting a small segment of the eye-tracker data, $\boldsymbol{\Delta}_i^m$, that occurred immediately prior to that button press. Here, $m \in M$ denotes mission number and $i \in I$ indexes each button press. The time window was set to 500 m/s, ending at the moment the button was pressed. For each $\boldsymbol{\Delta}_i^m$ set of data, the mean viewing location, $\boldsymbol{v}_i^m$, was calculated as a 2D vector of {x, y} screen coordinates. Figure 23 shows two examples in which all $\boldsymbol{v}_i$ have been plotted for two separate missions. In the example on the left, the entire set $\boldsymbol{v}_i$ was tightly packed near the 'Spot/BDA' button indicating that the data from the eye tracker agreed with the participant's behavior. In the example on the right, the set $\boldsymbol{v}_i$ appeared randomly distributed across the entire screen leading to the conclusion that either (1) the participant could successfully and repeatedly use the mouse cursor to press the 'Spot/BDA' button while looking in another direction or (2) that the data from the eye tracker were not reliable.

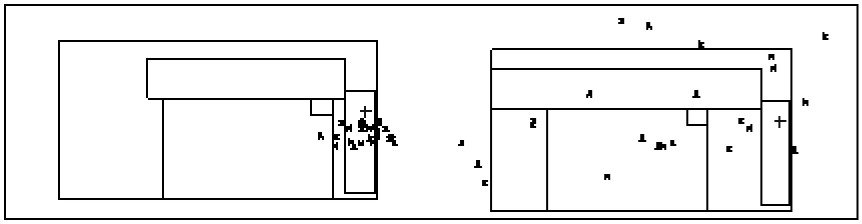

Figure 23. Example data plotting v_i^m, for all $i \in I$, (*) along with screen layout and the coordinates of the 'Spot/BDA' button (+).

Once the $\boldsymbol{v}_i^m$ data sets had been obtained, three descriptive parameters were extracted from each set. The first two parameters extracted were the variances along each of the two principal component dimensions, while the third parameter extracted was the difference between the overall mean of $\boldsymbol{v}_i$ for each $m = \{1, 2, \ldots, M\}$ and the known screen location for the 'Spot/BDA' button. These parameters were chosen to provide a measure of both the amount of noise and translational error present in the data. By normalizing these parameters on the scale [0, 1] and using them to represent the eye-tracker data, each mission was ranked from best to worst, or from 1 to 68, and an overall participant rank was determined by taking the average of each participant's four missions. The overall participant rank is shown in table 18, along with whether or not the participant had glasses as well as the previously determined subjective rank. As can be seen in table 19, participants with glasses tended to produce the lowest quality eye-tracker data and there was strong agreement between the results from the algorithm and the previously described manual approach.

Table 18. Algorithmically derived participant rank.

Participant	**Rank – Algorithm**	**Rank – Manual**	**Glasses?**
7	1	High	No
4	2	High	No
14	3	Medium	No
17	4	Medium	Yes
2	5	Medium	No
15	6	Medium	No
1	7	Medium	No
6	8	Medium	Yes
3	9	Medium	Yes
8	10	Low	Yes
11	11	Medium	No
13	12	Low	Yes
10	13	Low	Yes
5	14	Medium	No
9	15	Low	Yes
12	16	Low	Yes
16	17	Low	No

Using the algorithmically determined ranks, the seven participants with the best eye tracker data were chosen for more detailed analyses. This selection provided a combined total of 28 missions to analyze four Soldiers and three civilians and limited the number of participants with glasses to one. Once these missions had been selected, the fourth and final pre-processing step involved manually translating each data set to improve the overall alignment with the screen layout. This was done on a case-by-case basis and involved mapping the observed patterns of activity (identified via heat maps) to individual screen regions. An example is shown in figure 24 in which a large vertical translational error appeared in an arguably otherwise usable data set.

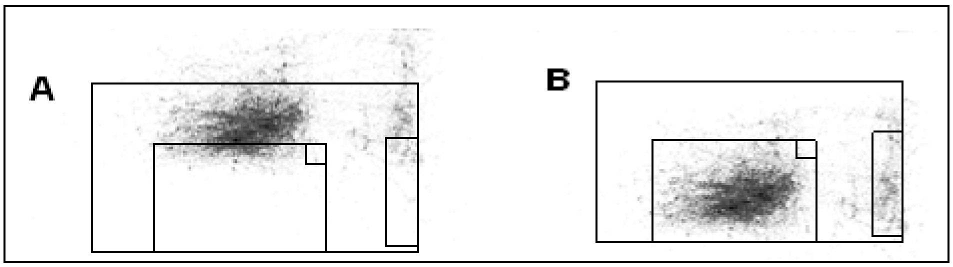

Figure 24. Example showing the need for data translation of eye-tracker data. Original data shown in (A) and translated data in (B).

As a final note on the pre-processing steps, it is understood that there remains considerable uncertainty when assessing the exact {x, y} screen locations for a participant's eye gaze at a specific instance of time. However, it is believed that the selected data should be still useful for: (1) identifying general search patterns, (2) assessing the degree to which participants utilized different screen regions, and (3) analyzing relative patterns of eye movement as they occurred independently of screen region.

4.7 Screen Usage and Performance Analysis

Tables 19–22 summarize the viewing proportion for each screen region, as well as threat detection per condition. For each condition, the viewing proportion was normalized to exclude all points that did not lie in one of the four screen regions: top banner, sensor view, bottom banner, and report panel. It should also be noted that the value for threat detection is only representative of the seven participants selected, and not of the whole participant set. Therefore, any deviations from earlier results should be viewed as an effect of down sampling the available participant set to obtain a subset of only those participants with usable eye-tracker data.

Table 19. Viewing proportion and threat detection for condition A.

	All Participants	**Soldier**	**Civilian**
Top banner (%)	—	—	—
Sensor view (%)	93.31	93.93	92.47
Bottom banner (%)	—	—	—
Report panel (%)	6.69	6.07	7.53
Threat detection (%)	54.35	58.30	49.07

Table 20. Viewing proportion and threat detection for condition B.

	All Participants	**Soldier**	**Civilian**
Top banner (%)	26.90	25.41	28.89
Sensor view (%)	66.45	67.62	64.89
Bottom banner (%)	—	—	—
Report panel (%)	6.66	6.98	6.23
Threat detection (%)	56.09	60.26	50.54

Table 21. Viewing proportion and threat detection for condition C.

	All Participants	**Soldier**	**Civilian**
Top banner (%)	19.64	15.94	24.57
Sensor view (%)	72.91	74.60	70.67
Bottom banner (%)	—	—	—
Report panel (%)	7.45	9.47	4.76
Threat detection (%)	54.30	62.91	42.82

Table 22. Viewing proportion and threat detection for condition D.

	All Participants	**Soldier**	**Civilian**
Top banner (%)	24.23	27.28	20.17
Sensor view (%)	66.22	64.70	68.26
Bottom banner (%)	3.30	2.28	4.66
Report panel (%)	6.25	5.75	6.91
Threat detection (%)	52.65	54.96	49.57

Two initial observations available from tables 19–22 are: (1) in all conditions, the sensor view received the most attention from both soldiers and civilians, and (2) in condition D, the bottom banner received very little attention and could, for the most part, be considered neglected. Further analysis yielded no direct, discernable relationship between screen viewing strategy and threat detection per condition. As an example, figure 25 plots threat detection as a function of increasing sensor view (%) for each condition. Though they are not shown here, similar results were obtained when including participant type, accuracy, detection range, or reaction time.

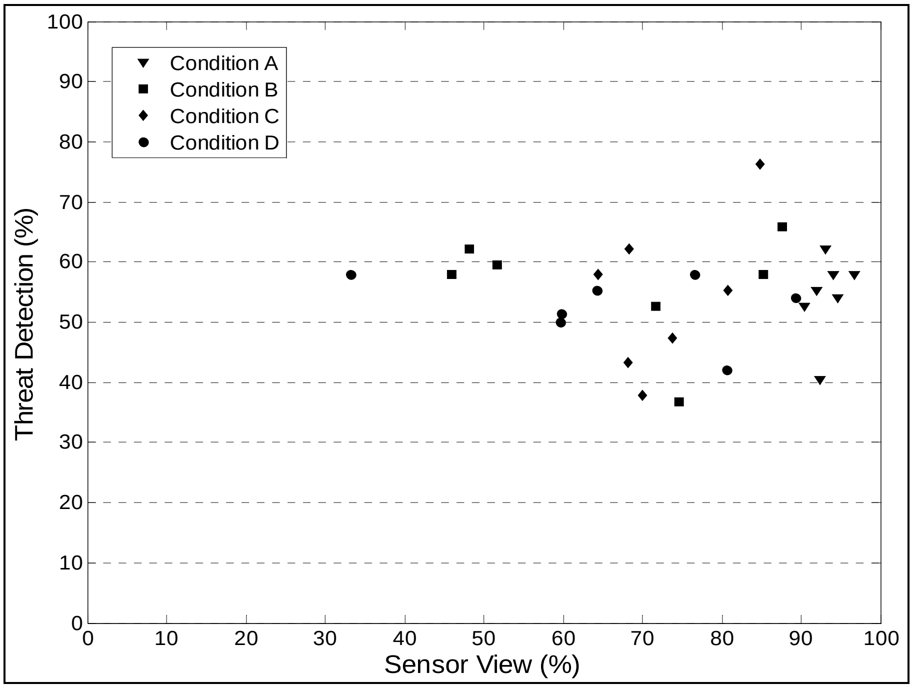

Figure 25. Threat-detection rates as a function of sensor portal viewing percentage.

The conclusion drawn from this initial investigation was that different individuals employed different search strategies, but that search strategy itself did not predict performance. However, it was noticed that as participants spent more time using the banners, the number of sensor view changes decreased (figure 26). The conclusion drawn is that while performance did not vary given search strategy, participants who heavily relied on the sensor view had to maintain a higher level of interaction with the system to reach the same performance levels as participants who tended to use the banners more.

Following this analysis, the investigation proceeded to considering where participants were looking immediately prior to filling out a report. The intention was to determine to what degree participants used the banners as a detection aid rather than as their sole means of detection. For each report submitted, analyses were performed on the previous 2 s of eye-tracker data and then a voting scheme was used to determine the screen region to which attention had last been given.

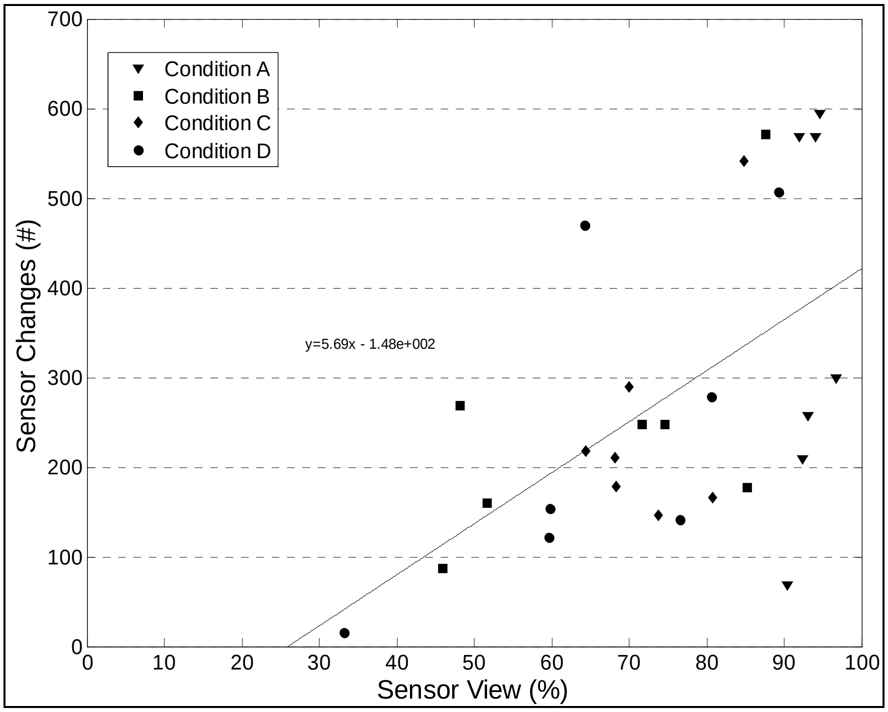

Figure 26. Sensor changes as a function of sensor portal viewing percentage, with linear trend line.

As shown in figure 27, as the usage of a particular screen region increased, the likelihood that a participant was using that screen prior to initiating a report also increased (note: figure 27 represents averages across conditions, thus the leveling off observed for the sensor view-based reports). From this trend, and the point noted earlier that mean reaction time did not differ with screen region usage, it was concluded that participants tended to not use the banners as a detection aid, but rather when participants used the banner they relied on it for extracting all information related to the report. In other words, the behavior—detect target in banner, use the sensor to get a closer look, and then initiate the report—was not observed. Instead, operators tended to submit reports immediately following detection of a threat, regardless of the screen region being used at the time.

Additional investigations, such as analyzing the probability of detecting moving targets, human targets, or IEDs when the vehicle was stationary or moving, or when target onset was in the front 180^{o} or rear 180^{o} of the vehicle also yielded no direct relationship between banner usage and performance that could not otherwise be explained as a difference between soldiers and civilians or between conditions.

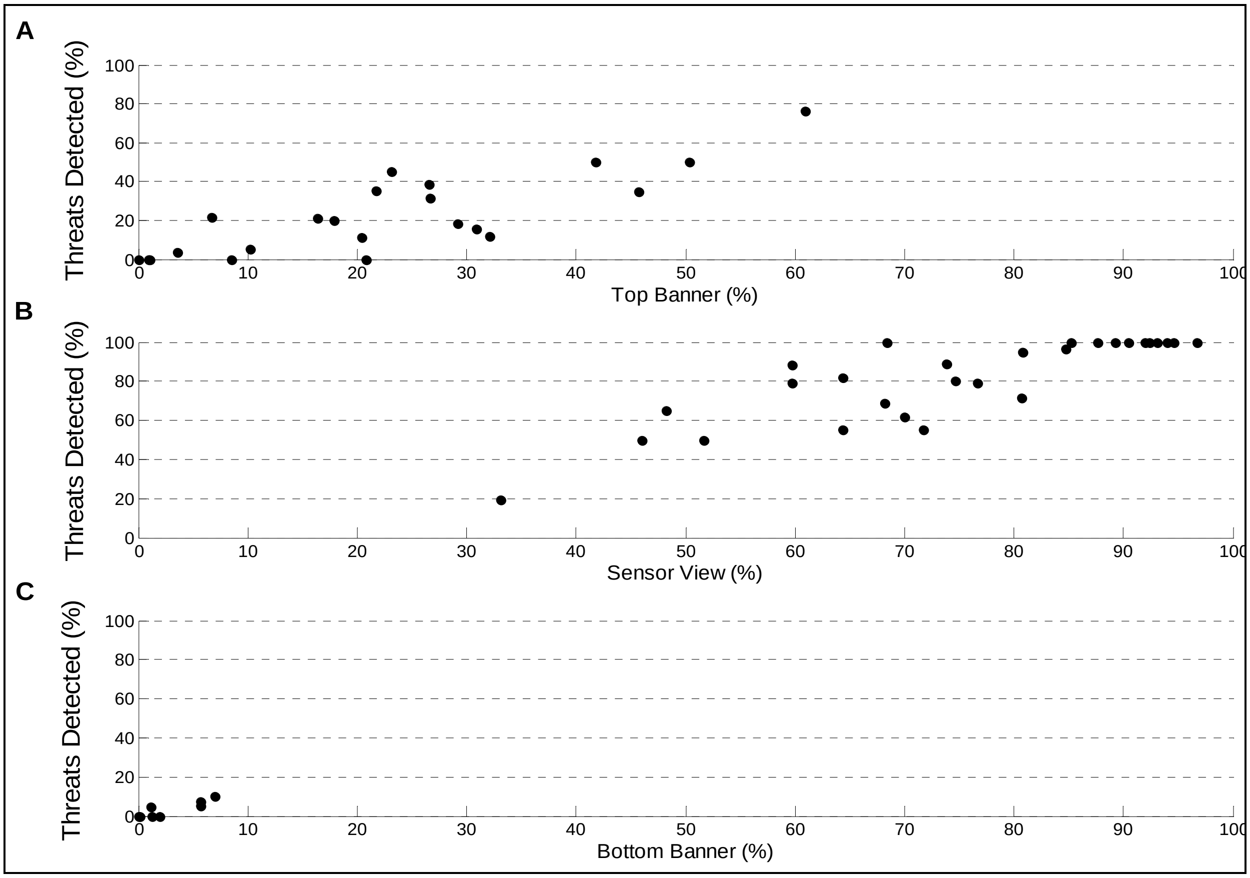

Figure 27. Percentage of threats detected in which participant's last recorded gaze position was on the (A) top banner, (B) sensor portal, or (C) bottom banner.

4.8 Eye Tracker Workload Analysis

Following the assessment of the specific relationship between eye position and screen usage, the investigation shifted focus on to other potential uses of eye tracker data for inference of the operators experience in using the 360° sensor system. In particular, following work presented in Fidopiastis et al. (2009), eye-movement patterns were investigated by using the Nearest Neighbor Index (NNI). The NNI is a metric for assessing distributions of point-based patterns originally developed by Clark and Evans (1954). The value of each NNI was obtained by first extracting eye fixations, or periods in which visual gaze was maintained at roughly the same location, and then analyzing eye movements between fixation points (Di Nocerra et al., 2006; Salvucci and Goldburg, 2000). The final NNI value was then determined to be the ratio of the sum of nearest neighbor distances (for all points) and the distance one would expect if the fixations had been generated randomly. NNI values near 1.0 suggest a random distribution, or potentially low workload, while values less than 1.0 suggest focused searching, and thus potentially high workload (Fidopiastis et al., 2009). The intent with investigating NNI was to determine if there was a noticeable difference in the behavior patterns (and possibly workload) for participants with better overall detection rates, accuracies, and reaction times. For a more complete discussion and review of how eye-movement patterns and other eye-related metrics (e.g., blink duration and pupil dilation) may be related to operator workload, the interested reader is directed to Andreassi (2000), Van Orden et al. (2001), Marshall et al. (2003), and De Greef et al. (2009).

NNI scores were extracted for each of the 28 missions in the reduced data set. Fixations were identified using a velocity filter set to 10^{o}/s, with minimum allowable fixation duration of 150 m/s. The size of the search area was chosen to be the size of the convex hull surrounding all fixations. The results shown in figure 28 indicate that as participant's search patterns became more random (i.e., NNI values closer to 1.0), threat detection tended to decrease. This trend, however, was less apparent with regards to mean accuracy (figure 29) and mean reaction time (figure 30).

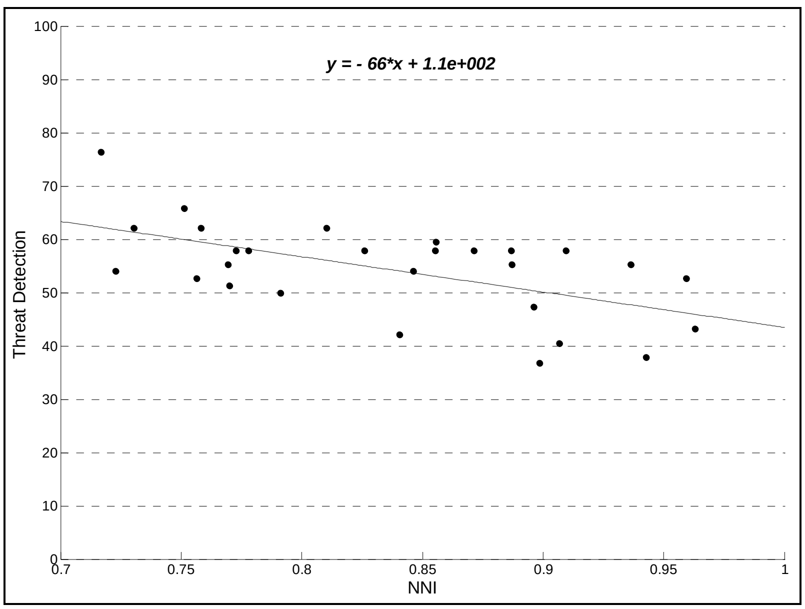

Figure 28. Threat detection across condition as a function of NNI, with linear trend line.

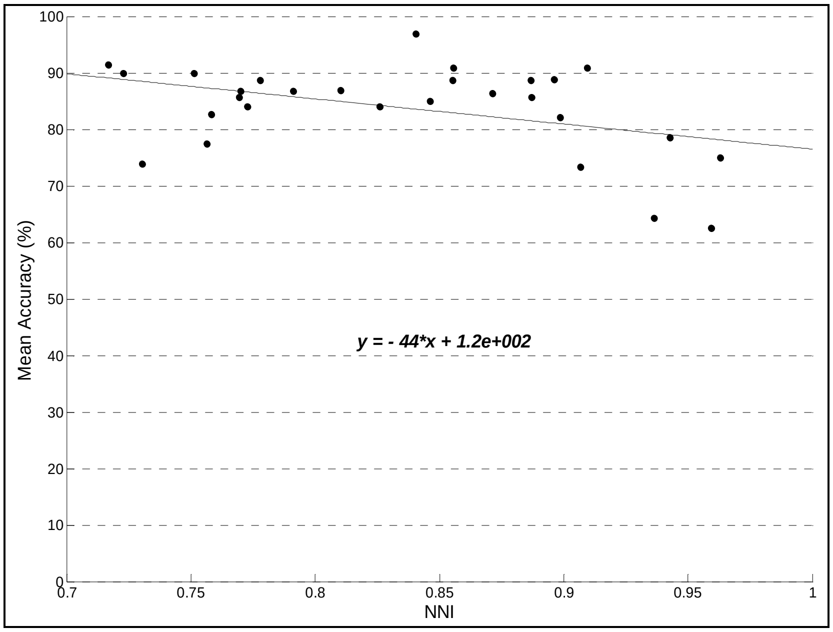

Figure 29. Mean accuracy across condition as a function of NNI, with linear trend line.

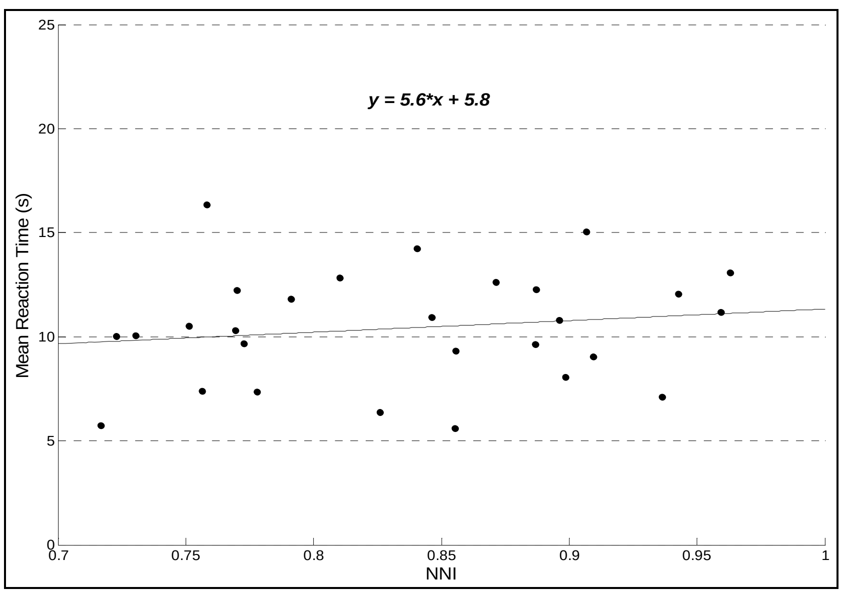

Figure 30. Mean reaction time across condition as a function of NNI, with linear trend line.

Figure 31 separates threat detection into a function of both NNI and condition. Similar trends were seen for conditions A, B, and C that were observed for the entire data set, but interestingly in condition D, the same trend is not observed. That is, the participants whose eye movements were measured to be more random did not appreciably perform worse with respect to threat detection in condition D. The most likely explanation for this result is that search strategy was less of an influencing factor on performance in this condition, where participants had both top and bottom banners providing a 360^{o} FOV.

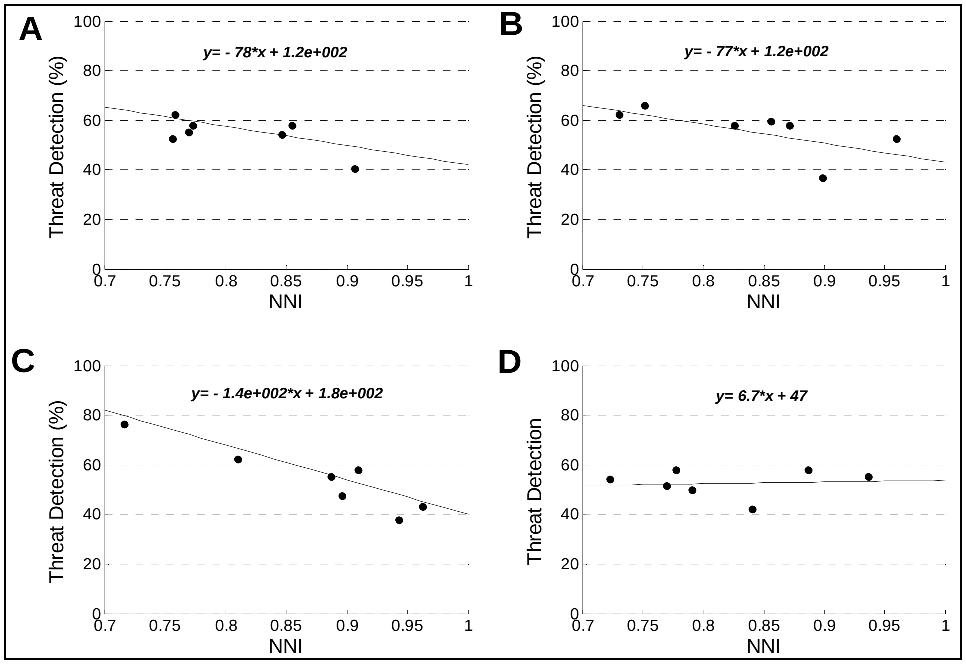

Figure 31. Threat-detection rate as a function of NNI and condition, with linear trend lines. Panels correspond with (A) condition A, (B) condition B, (C) condition C, and (D) condition D.

Finally, figure 32 plots the observed NNI values versus the computed workload measures from the NASA-TLX for each of the four conditions. While a word of caution is warranted when interpreting these results (due to the overall reduced data size and various uncertainties attached to the quality of the eye tracker data), it is still intriguing to note that the recorded eye tracker data shows potential as surrogate measure of workload and can also potentially be related to performance.

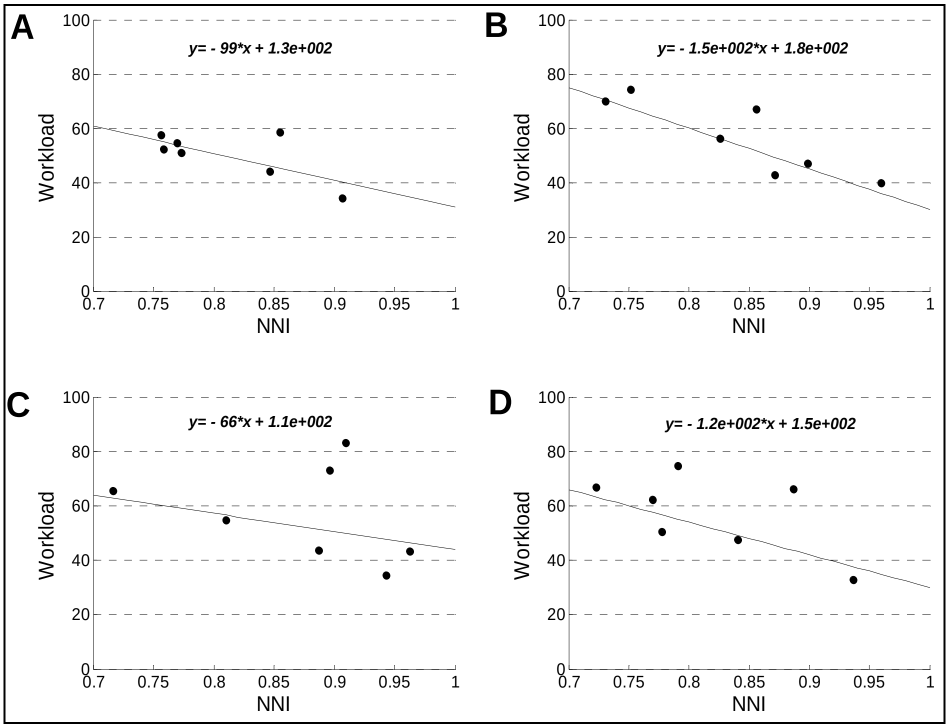

Figure 32. Reported workload as a function of NNI and condition, with linear trend lines. Panels correspond to (A) condition A, (B) condition B, (C) condition C, and (D) condition D.

4.9 Concluding Remarks on Eye-Tracker Analysis

Through the current analysis of the available, and usable, eye tracker data, some initial conclusions can be drawn: (1) the use of banners was not a good predictor of overall performance, (2) in condition D, the bottom banner was used rarely, if at all, and (3) the data obtained from the eye tracker has the potential to be used for workload analysis and, under certain conditions, participant's search patterns may be indicative of performance. However, as previously stated, these results are only tentative due to a number of reasons, specifically the reduced size of the available data set and the difficulties removing noise and simultaneously proving the reliability of the measured eye gaze information.

For improved reliability and quality of eye tracker data in future experiments, it would be worthwhile to obtain specific calibration information from each participant. This would require additional experimental preparation time, but would provide labeled training points which would allow an algorithmic process to be designed that assesses, and corrects, data quality with respect to known values. Furthermore, reducing the number of participants with glasses, or requesting that participants who can wear either glasses or contacts to wear their contacts for the experiment would reduce a lot of the currently observed noise effects. It is, of course, understood that obtaining participants who do not wear glasses may not always be possible, and thus some noise may need to be tolerated. This further enforces the need to obtain participant-specific calibration data.

4.10 Subjective Workload

A repeated measures MANOVA was performed to examine the participants' perceived workload (weighted NASA-TLX scores), with interface condition as the within-subject factor. There were no significant differences in perceived workload between the interface conditions, *F*s < 1.0. (See figure 33 and table 23.)

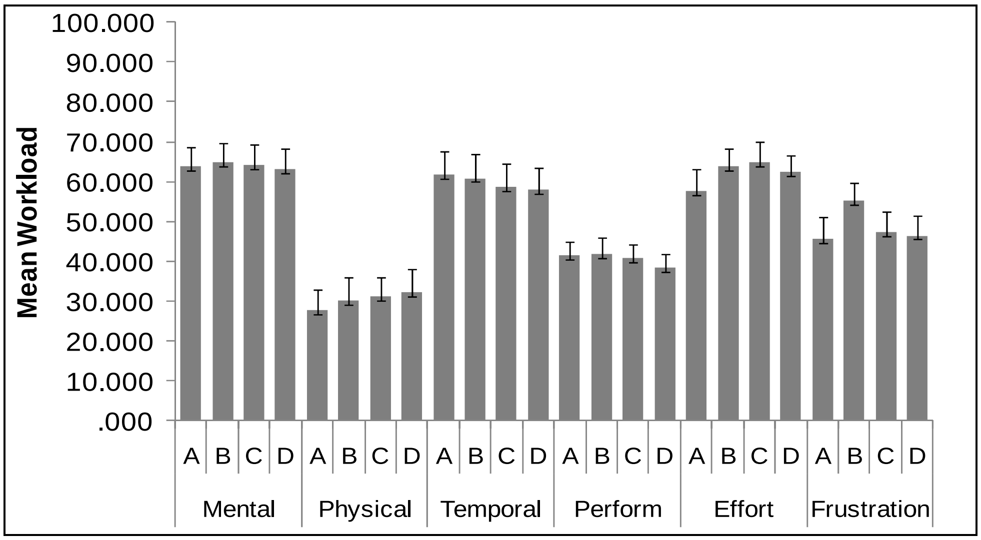

Figure 33. Mean workload rating for the NASA-TLX subscales.

Table 23. Mean overall workload by condition.

Condition	Overall Workload
A	56.6 (3.83)
B	60.0 (3.07)
C	59.5 (3.54)
D	57.1 (3.22)

4.11 Usability Assessment: Operator Preferences

While, in general, usability data are highly subjective, in this case, operators' opinions of the technologies used in this study line up very well with their corresponding performance measures. Of particular interest here is the relationship that the usability data have with sensors view changes. Participants were asked to rank order the sensor system used in this study according to preference (from 1 to 4). These results were then averaged across all participants and the outcome of these calculations shows that condition A (the sensor only configuration) was overwhelmingly chosen as the least favored option (mean score = 3.88). Condition C (the sensor + top banner, small-portal configuration) was the second-least favored option with a mean score of 2.88. Condition B (the sensor + top banner, full-screen configuration) and condition D (the sensor + 2 banner, full-screen configuration) rated the best with mean scores of 1.65 and 1.59, respectively, with condition D scoring the greatest number of first-place votes (11 out of 17, or 67%).

When matching this data up with the scores for sensor view changes, you can see that the conditions rank order in the same way (table 24). Given that overall, target identification performance was not dramatically different from one sensor condition to the next, the quasi-workload measure of sensor view changes shows that operators were more efficient in the task when working with those systems that included at least one banner, in a full screen mode (i.e., conditions B and D). Further, combining this outcome with the usability data supports the idea that operators were cognizant of this increase in efficiency.

Table 24. Comparison of operator's rank of experimental displays against sensor view changes.

Interface Configuration	**Total Sensor View Changes**	**Sensor Views Per Minute**	**Operators Average Rank Order of Interface**
A	381.5	28.2	3.88
B	171.4	12.6	1.65
C	189.2	13.8	2.88
D	131.0	9.6	1.59

In addition to these findings, exit interviews were conducted to establish any additional preferences or concerns expressed by the operators in this study (table 25). While completely subjective in nature, data such as these can indicate areas of success and potential failure within system design.

In the case of the current study, every one of the 17 operators expressed a preference toward using the display interface configurations that included a banner as opposed to those without a banner display. Further data from the exit interviews show that only five of the 17 participants expressed a sense of being overwhelmed by information during the study. However, only one of those five was a non-civilian (i.e., having a military background). Both of these findings support the use of banners in display configurations as a useful and wanted display feature.

Table 25. List of preferences and concerns organized by participant (1–17).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**% All**	**% Sld**
Banner preferred	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	100	100
Wants zoom	—	×	—	×	×	—	—	—	×	—	×	×	—	—	×	—	×	47.1	71.4
Sa report problem	×	—	—	×	—	×	×	×	—	—	—	—	—	—	×	×	×	47.1	42.9
Sensor control (clicking) problem	×	—	×	—	—	—	—	×	×	—	—	—	—	×	—	—	—	29.4	28.6
Wants continuous panning	—	×	—	—	—	—	—	—	—	×	—	×	—	—	—	×	×	29.4	14.3
Sensor aimed at rear	—	—	×	—	—	×	—	×	—	—	×	—	—	—	×	—	—	29.4	14.3
Wants interactive targeting	—	—	—	—	—	×	×	×	—	—	—	—	—	×	—	—	—	23.5	14.3
Overwhelmed by information	—	—	—	—	—	×	—	—	—	—	×	—	×	—	—	×	×	29.4	14.3

Notes: Shaded columns indicate Soldiers, % All is calculated as a percentage of 17 participants, and % Sld is a percentage of the seven Soldiers.

Still some potential issues may have been exposed by the exit interviews. While none of the areas mentioned by the operators were expressed by the majority, the issues are still important to point out as potential pitfalls for the implementation of a display system based on the designs found in this study. Table 25 shows a complete listing of the preferences and areas of concern generated from the exit interviews of operators in this study.

5. Conclusions

The main objective of the experiment was to compare four different interface configurations for representing 360° SA. In general, the data suggest that the presence of banners has potential for improving efficiency of operators maintaining local area awareness. However, even with a 360° FOV available to the operator, target detection remains a difficult task.

We anticipated that the presence of banners would enhance target detection performance. Overall, there was no resounding evidence for dramatic effects of display condition on target detection in the main overall analyses. Several smaller effects were observed, but were clearly not the main sources of influence over the variability in observed performance. However, inspection of operator interactions with the SMI and other variables in the experiment showed an advantage for the use of banners. Though overall performance did not differ radically, more interactions with SMI were required to obtain the same performance in condition A as compared to conditions B through D. Fifty-five percent more sensor changes occurred in the condition with only a sensor view (A) than one where the operator had a forward looking banner (B). Soldier preferences mirrored these results, noting that the no banner condition was more difficult to use, even though the banner condition took more time to learn.

When banners were present, operators looked more frequently to the rear of the vehicle than when no banners were present, which is an important component of local area awareness. This fact was confirmed by the sensor usage data and eye tracking data which showed that rear views were infrequent without banners. The banners allow operators to keep their eyes forward while on the move and still scan the rear of the vehicle. Banners provide partially redundant display information, increasing the possible viewing time and allowing for threat detection as well as identification. Importantly, Soldiers showed a preference for the two conditions that included banners and single sensor portal view. In fact, in exit interviews, Soldiers described a useful technique in which the front 180° banner was used to observe forward for targets and vehicle mobility, while a sensor view at the 6 o'clock position was used to simultaneously scan for targets to the rear of the vehicle.

Poor performance for threats presented to the rear despite available technology to view the rear of the vehicle indicates that rear target detection is an important issue and warrants further investigation. In addition to display factors, task factors were also important determinants of threat detection and identification performance. Threat type, threat range, location of threat onset, threat density (inter-threat interval), and mobility were all factors affecting performance. There was a generalized decrease in probability of detection with increased range, but a generalized increase in detection probability with increased time to view a threat. Threat onset location, as previously mentioned, also affected performance such that performance was poor for threats that onset to the rear of the vehicle. Viewing time was an important organizer of performance; increased viewing time increased the likelihood there would be detection because the operators had more time make discrimination. The presence of redundant displays (banner + sensor portal) increased viewing time and in turn could enhance detection. Mobility also affected performance. Target detection was improved when the vehicle was on the move relative to stationary. One explanation for this may be that, on the move, operational tempo (OPTEMPO) is really high which can heighten operator awareness. In addition, when on the move the viewing time for a target is increased and as the vehicle get closer to a target the probability of correct detection increases since the target is more clearly discernable.

These data provide a justification for operators choosing a display that maximizes the opportunity to view threats and thus, redundant displays would be predicted to lead to benefit based on current data. Indeed, usability data on Soldiers' preference indicates that they would agree with this statement: choosing the display configurations with the largest primary viewing area and most judicious use of banners (conditions B and D).

In addition to the traditional human factors measures of objective performance and subjective workload, a non-invasive physiological measure (eye tracking) was included in the paradigm. Eye movement patterns were successfully tracked in seven of the 17 participants. Screen region usage was assessed with the performance data. The screen usage data showed that the bottom banner was mostly neglected. There were no identified relationships between banner usage and any of the selected performance variables. There was, however, a relationship between banner usage and interaction with sensor (number of sensor view changes). Participants that neglected the banners had to rely on constantly changing sensor position to achieve similar results. This was validated by the significant effect of display configuration on sensor view changes. The eye tracking and sensor use data also indicated variation in workload throughout the experiment, with some variation attributable to display condition. Eye-tracking data indicated, based on NNI results, that increasing workload may have a deleterious impact on threat detection performance.

The final recommendation from these results is to provide a banner solution for the 360° system. The use of the banner combined with a sensor portal view allows for both enhanced field of view as well as enhanced range performance. Further study is needed to resolve optimal display and usability characteristics for appropriate banner technology. Technology is needed to aid the operator in target detection especially for mid- and long-range targets across the 360° spectrum and when the vehicle is on the move. Range performance is an important factor and sacrificing range for the sake of FOV is not recommended for maintaining local area awareness. Real-time image processing may ultimately be necessary to account for other factors affecting detection and identification performance and particularly useful would be ways of enhancing target salience in cluttered environments.

6. References

Andreassi, J. L. *Psychophysiology: Human Behavior and Physiological Response*. Lawrence Erlbaum Associates: Hillsdale, NJ, 2000.

Clark, P. J.; Evans, F. C. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology* **1954**, *35*, 445–453.

De Greef, T.; Lafeber, H.; Van Oostendorp, H.; Lindenberg, J. Eye Movement as Indicators of Mental Workload to Trigger Adaptive Automation. In *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience.* Springer: Berlin, Germany, 2009; pp 219–228.

Di Nocerra, F.; Terenzi, M.; Camilli, M. Another look at Scanpath: Distance to Nearest Neighbor as a Measure of Mental Workload. In *Developments In Human Factors In Transportation, Design, And Evaluation;* de Ward, D., K.A. Broohuis, K.A., Toffetti, A., Eds.; Shakier Publishing: Maastricht, Netherlands, 2006; pp 1–9.

Draper, N. R.; Smith, H. *Applied Regression Analysis;* 3rd ed. New York: John Wiley & Sons: New York, NY, 1998.

Fidopiastis, C. M.; Drexler, J.; Barber, D.; Cosenzo, K.; Barnes, M.; Chen, J. Y. C.; Nicholson, D. Impact of Automation and Task Load on Unmanned System Operator's Eye Movement Patterns. In *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience.* Springer: Berlin, Germany, 2009; pp 229–238.

Hansen, D. W.; Ji, Q. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2010**, *32* (3), 478–500.

Hart, S. G.; Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload;* Hancock, P. A., Mechkati, N., Eds.; North Holland Press: Amsterdam, 1998.

Marshall, S. P.; Pleydell-Pearce, C. W.; Dickson, B. T. Integrating Psycho Physiological Measures of Cognitive Workload and Eye Movements to Detect Strategy Shifts. *International Conference on System Sciences*, 2003.

Salvucci, D. D.; Goldburg, J. H. Identifying Fixations and Saccades in Eye-Tracking Protocols. *Proceedings of the Eye Tracking Research and Application Symposium*. ACM Press: New York, NY, 2000; pp 71–78.

Van Orden, K.; Makeig, S.; Jung, T. P. Eye Activity Correlates of Workload during a Visual Spatial Memory Task. *Human Factors* **2001**, *43*, 111–121.

INTENTIONALLY LEFT BLANK.

Appendix A. Informed Consent

This appendix appears in its original form, without editorial change.

Consent to Participate in Research Informed Consent Form
Army Research Laboratory, Human Research & Engineering Directorate
Aberdeen Proving Ground, MD 21005

IRB USE ONLY: **ARL-20098-**
Institutional Review Board
Aberdeen Proving Ground, MD
Approval Date:
Expiration Date:

Title of Project: Impact of 360° Sensor Information on Vehicle Commander Performance

Principal Investigator: Keryl A. Cosenzo
Soldier Performance Division
Cognitive Sciences Branch
Robotics Program Team
410-278-5885, kcosenzo@arl.army.mil

Jillyn Alban
Intelligent Ground Systems
U.S. Army Tank Automotive Research, Development, and Engineering Center
Soldier Machine Interface Team
586-574-3941, jillyn.alban@us.army.mil

Purpose of the Study: The objective of this research is to evaluate the effect of the new 360/90° sensor system on the vehicle commander's ability to maintain local area awareness and disseminate the information to his crew.

Procedures to be followed: Upon arrival at the lab, the experimenter will brief you on the purpose of the experiment and to determine if you would like to participate. You will then be given a description of the local security systems and interface and the primary goals of the experiment will be briefly outlined. You will be asked to play the role of a vehicle commander and conduct security on the move in complex urban terrain using two different indirect vision sensor systems. After receiving the overview you will be set up with the eyetracker and Lifeshirt. You will then complete four training missions to ensure that you understand the task. You can repeat the training if needed. After training, you will complete four missions. Following each mission you will complete the NASA-TLX to assess your subjective workload.

Discomforts and Risks: The risks that will be encountered in this study are minimal. There are minimal risks associated with the head and eye-tracking apparatus used in this experiment. The infrared eye-tracking pods used in this experiment are well within the Maximum Permissible Exposure (MPE) limit, based on the International Electrotechnical Commission (IEC) Standard 60825-1 V1.2. Also, the pod's minimum operating distance is well beyond the distance that represents an ocular hazard (based on the IEC Standard 60825-1 V1.2.).There are minimal risks associated with the physiological recording equipment. There is no risk of shock from the LifeShirtTM. You may develop skin irritation in response to the skin preparation, electrode gel, or electrode adhesive. This possible irritation is similar to what you might develop in response to a bandage or skin lotion that you do not typically use.

Benefits: Although there are no direct benefits for your participation in this experiment, you will contribute to our understanding of how a 360/90° sensor affects Soldier performance. You will receive the personal satisfaction of providing valuable information to the Army's manned and unmanned ground vehicle research robotics research.

Duration: It will take ~4 hours to complete this experiment. A 5 minute break is scheduled between each mission. Breaks can be taken as often as needed during the experiment.

Payment for participation: You will not be paid for your participation in this study in addition to your regular salary.

Confidentiality:

Your participation in this research is confidential. The data will be stored and secured at Building 459 at Aberdeen Proving Ground, MD, in a password protected computer file. Publication of the results of this study in a journal or technical report or presentation at a meeting will not reveal personally identifiable information. This consent form will be sent to Army Research Laboratory's Institutional Review Board, where it will be retained for a minimum of three years.

No personally identifiable information will be shared with anyone outside the research staff. Officials of the U. S. Army Human Research Protections Office and the Army Research Laboratory's Institutional Review Board may inspect the records obtained in this study to insure compliance with laws and regulations covering experiments using human subjects.

Participation terminated by the investigator: N/A

Contact Information for Additional Questions:

You have the right to obtain answers to any questions you might have about this survey research both while you take part in the study and after you leave the research site. Please contact the researcher listed at the top of the first page of this consent form for more information about this study. You may also telephone Dr. Paul Rose of the Army Research Laboratory Human Research and Engineering Directorate Institutional Review Board at (410) 278-5992 with questions, complaints, or concerns about this survey research. He can also answer questions about your rights as a research participant. You may also call this number if you cannot reach the research team or wish to talk to someone else.

Voluntary Participation: Your decision to be in this research involving completing questionnaires and interview is voluntary. You can stop at any time. You do not have to answer any questions you do not want to answer. Refusal to take part in or withdrawal from answering these questions will involve no penalty or loss of benefits you would receive by staying in it.

Military personnel cannot be punished under the Uniform Code of Military Justice for choosing not to take part in or withdrawing from this study, and cannot receive administrative sanctions for choosing not to participate. Civilian employees or contractors cannot receive administrative sanctions for choosing not to participate in or withdrawing from this study.

You must be 18 years of age or older to take part in this research study. If you agree to take part in this research study based on the information outlined above, please sign your name and indicate the date below.

You will be given a copy of this consent form for your records.

____ ____________________

Participant Signature Date

____ ____________________

Person Obtaining Consent Date

INTENTIONALLY LEFT BLANK.

Appendix B. Demographics Questionnaire

This appendix appears in its original form, without editorial change.

Participant ID# __________

Demographic and Computer Experience Questionnaire

1. AGE: _____
2. GENDER: ___Male ___ Female
3. Do you wear glasses? ___ Yes ___ No

4. Do you have any reason to believe that you have a hearing impairment? ___Yes ___ No

5. Do you have an apparent hearing impairment? ___Yes ___No

6. Please indicate your highest level of education:
   ___ High School Diploma
   ___ Undergraduate Degree
   ___ Some graduate courses
   ___ Graduate Degree
   ___ Other

7. Are you in the military? ___Yes ___No If yes, what Branch? _________________

   For how many years? ___Less than 5 years ___5-10 years ___ 11-15 years ___16-20 years ___ 20 years or more

   What is your rank? _____ What is your MOS? ___________________

8. Does your job require you to use a computer on a regular basis? ___Yes ___No

9. How long have you been using a computer?

__Less than 1 year ___ 1-3 years ___4-6 years ___7-10 years ___10 years or more

10. How often do you use a computer?

___Daily ___Weekly ___Monthly ___Once or twice a year

11. . Do you have a computer in your house? ___Yes ___No

12. . Do you use the computer to play games? ___Yes ___No

   If yes, how often? ___Daily ___Weekly ___Monthly ___Once or twice a year

13. Do you have operational experience in complex urban terrain? ___Yes ___No

   If yes, where ____

14. Have you ever conducted security patrols in complex urban terrain? ___Yes ___No

   If yes, where ____

15. Have you ever used an indirect vision system to conduct local security? ___Yes ___No

   If yes, which systems ____

Appendix C. National Aeronautics and Space Administration Task Load Index (NASA-TLX) Questionnaire

This appendix appears in its original form, without editorial change.

NASA TLX Questionnaire

Participant ID:________________

TLX Workload Scale

Please rate your workload by putting a mark on each of the six scales at the point which matches your experience.

Mental Demand

Low **High**

Physical Demand

Low **High**

Temporal Demand

Low **High**

Performance

Good **Poor**

Effort

Low **High**

Frustration

Low **High**

Appendix D. Usability and Exit Interview Questionnaires

This appendix appears in its original form, without editorial change.

Usability

PIN___________ Date______________ Time____________ Comments

1. Text size was large enough to read |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
2. It was hard to select some items when using the mouse |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
3. Buttons/controls are arranged conveniently so that I can quickly find the ones I need at any given time |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
4. The terminology used throughout the interface was appropriate |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
5. The sensor orientation icon was easy to interpret |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
<u>Sensor Only WMI Configuration</u> Please fill out this section for the Sensor Only Configuration	
6. The Sensor Window size was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
7. The Sensor Window field of view was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
8. I always knew where my sensor was pointed relative to my vehicle |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
9. It was difficult to determine where a target was relative to my vehicle |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	

10. The Sensor Only configuration allowed me to effectively <u>search for</u> targets |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
11. The Sensor Only configuration allowed me to effectively <u>identify</u> targets |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
12. The interface allowed me to effectively control sensor orientation |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
13. Overall, I think the Sensor Only WMI configuration was easy to learn |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
<u>Sensor + Top Banner Configuration, Full Screen</u> Please fill out this section for the Sensor +Top Banner, Full Screen Configuration only	
14. The Sensor Window size was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
15. The Sensor Window field of view was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
16. The Banner size was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
17. The Banner horizontal field of view was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
18. The Banner vertical field of view was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	

19. The Banner orientation was appropriate for the tasks I completed

|-----------|-----------|-----------|-----------|-----------|-----------| N/A
1 2 3 4 5 6 7
Strongly Disagree ... Strongly Agree

20. I always knew where my sensor was pointed relative to my vehicle

|-----------|-----------|-----------|-----------|-----------|-----------| N/A
1 2 3 4 5 6 7
Strongly Disagree ... Strongly Agree

21. It was difficult to determine where a target was relative to my vehicle

|-----------|-----------|-----------|-----------|-----------|-----------| N/A
1 2 3 4 5 6 7
Strongly Disagree ... Strongly Agree

22. The Sensor + Top Banner, Full Screen configuration allowed me to effectively <u>search for</u> targets

|-----------|-----------|-----------|-----------|-----------|-----------| N/A
1 2 3 4 5 6 7
Strongly Disagree ... Strongly Agree

23. The Sensor + Top Banner, Full Screen configuration allowed me to effectively <u>identify</u> targets

|-----------|-----------|-----------|-----------|-----------|-----------| N/A
1 2 3 4 5 6 7
Strongly Disagree ... Strongly Agree

24. The Sensor + Top Banner, Full Screen configuration allowed me to effectively use the vehicle sensor system

|-----------|-----------|-----------|-----------|-----------|-----------| N/A
1 2 3 4 5 6 7
Strongly Disagree ... Strongly Agree

25. Overall, I think the Sensor + Top Banner, Full Screen configuration was easy to learn

|-----------|-----------|-----------|-----------|-----------|-----------| N/A
1 2 3 4 5 6 7
Strongly Disagree ... Strongly Agree

<u>Sensor + Top Banner Configuration, Small Portal Mode</u>

Please fill out this section for the Sensor +Top Banner, SP Configuration only

26. The Sensor Window size was appropriate for the tasks I completed

|-----------|-----------|-----------|-----------|-----------|-----------| N/A
1 2 3 4 5 6 7
Strongly Disagree ... Strongly Agree

27. The Sensor Window field of view was appropriate for the tasks I completed

|-----------|-----------|-----------|-----------|-----------|-----------| N/A
1 2 3 4 5 6 7
Strongly Disagree ... Strongly Agree

28. The Banner size was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
29. The Banner horizontal field of view was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
30. The Banner vertical field of view was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
31. The Banner orientation was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
32. I always knew where my sensor was pointed relative to my vehicle |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
33. It was difficult to determine where a target was relative to my vehicle |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
34. The Sensor + Top Banner, SP mode configuration allowed me to effectively <u>search for</u> targets |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
35. The Sensor + Top Banner, SP mode configuration allowed me to effectively <u>identify</u> targets |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	
36. The Sensor + Top Banner, SP mode configuration allowed me to effectively use the vehicle sensor system |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree ... Strongly Agree	

37. Overall, I think the Sensor + Top Banner, SP mode configuration was easy to learn |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
<u>Sensor + 2 Banner Configuration, Full Screen</u> Please fill out this section for the Sensor +Top Banner Configuration only	
38. The Sensor Window size was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
39. The Sensor Window field of view was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
40. The Banner size was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
41. The Banner horizontal field of view was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
42. The Banner vertical field of view was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
43. The Banner orientation was appropriate for the tasks I completed |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
44. I always knew where my sensor was pointed relative to my vehicle |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
45. It was difficult to determine where a target was relative to my vehicle |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	

46. The Sensor + 2 Banner, Full Screen configuration allowed me to effectively <u>search for targets</u> |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
47. The Sensor + 2 Banner, Full Screen allowed me to effectively <u>identify</u> targets |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
48. The Sensor + 2 Banner, Full Screen configuration allowed me to effectively use the vehicle sensor system |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
49. Overall, I think the Sensor + 2 Banner, Full Screen configuration was easy to learn |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
General	
50. Overall, I liked the banner concept for displaying extended field of view |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
51. I sometimes felt lost in the banner imagery |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
52. It would like to tailor the presentation of banners (appearance, field of view, orientation) throughout a mission or from mission to mission depending on goals |-----------|-----------|-----------|-----------|-----------|-----------| N/A 1 2 3 4 5 6 7 Strongly Disagree … Strongly Agree	
53. Please rank the test conditions in terms of your preference for the missions you completed (1= best, 4 = worst) ____ Sensor Only ____ Sensor + Top Banner, Full Screen ____ Sensor + Top Banner, Small Portal Mode ____ Sensor + 2 Banners, Full Screen	

Exit Interview

PIN__________ Date_____________ Time___________

Note: This questionnaire is to be given verbally by the experimenter.

Sensor Window Only Configuration – Show participant screen shot

1. **What would you improve about the Sensor Only WMI configuration that you used?**

2. **What are the strengths of the Sensor Only WMI configuration?**

3. **What was your general strategy or technique for searching for targets with the Sensor Only WMI configuration?**

Sensor + Top Banner Configuration, Full Screen – Show participant screen shot

4. **What would you improve about the Sensor + Top Banner, Full Screen WMI configuration that you used?**

5. **What are the strengths of the Sensor + Top Banner, Full Screen WMI configuration?**

6. **Where did you generally first see a target of interest on the WMI (sensor window or banner)?**

7. **What was your general strategy or technique for searching for targets with the Sensor + Top Banner, Full Screen WMI configuration?**

Sensor + Top Banner Configuration, SP Screen Mode – Show participant screen shot

8. **What would you improve about the Sensor + Top Banner, SP Screen Mode WMI configuration that you used?**

9. **What are the strengths of the Sensor + Top Banner, SP Screen Mode WMI configuration?**

10. **Where did you generally first see a target of interest on the WMI (sensor window or banner)?**

11. **What was your general strategy or technique for searching for targets with Sensor + Top Banner, SP Screen Mode WMI configuration?**

Sensor + 2 Banner Configuration, Full Screen – Show participant screen shot

12. **What would you improve about the Sensor + 2 Banner, Full Screen WMI configuration that you used?**

13. **What are the strengths of the Sensor + 2 Banner, Full Screen WMI configuration?**

14. **Where did you generally first see a target of interest on the WMI (sensor window or banner)?**

15. What was your general strategy or technique for searching for targets with Sensor + 2 Banner, Full Screen WMI configuration?

General

16. Would you use a 360 SA sensor system similar to this on a vehicle today?

_____ Yes

_____ No

17. Do you have any recommendations for additional features or information that would help you use a 360 SA sensor system?

18. The biggest issue I had during the missions was….

19. The feature I liked best about the system I experienced was…

INTENTIONALLY LEFT BLANK.

Appendix E. Situational Awareness Report – Entity Association Algorithm

Advanced development of technologies and systems for use in the military operational environment persistently challenges scientists and engineers to make use of increasingly complex experimental scenarios for testing and technology evaluation. Traditionally, simulation environments used for testing behavioral and psychological performance have involved minimal complexity in terms of the operational events requiring action by experiment participants. For example, events in simulated operational scenarios have been controlled in a manner such that stimuli were presented sequentially with no overlap in time and/or were confined to a limited spatial distribution. Even if the event sequence unfolded at a demanding temporal pace and spatial frequency, experimenters using such paradigms could nearly always identify a unique mapping between stimulus and response by virtue of such highly controlled and simplified emulations of the operational environment. In cases where interpretation was confounded by overlapping events, the standard protocol would be to simply discard both events from further analyses (provided enough other data points existed for analysis).

Despite the importance of experimental control during development, modern technologies are designed to enable successful Soldier-system performance in tasks and environments that are not nearly as clean, or uncluttered, as those that have been applied in previous experimental endeavors. Such new systems are meant to enable Soldiers to manage multiple, temporally and spatially overlapping events unfolding in a distributed and extended task space. As a result, designers of experimental scenarios and engineers that implement such scenarios in high-fidelity three-dimensional simulation environments must respond by developing tools and techniques to allow presentation of stimuli of increasing correspondence with the real operational environment during experimental test and evaluation activities.

Inherently, the incorporation of temporally and spatially overlapping operational events amidst a myriad of possible decoy/distracter events (e.g., ambush activity, suspicious “spotter”-type behavior of individuals in a crowd, presence of a partially disguised IED, etc.) in a distributed spatial layout leads to ambiguity in interpretation and understanding of the operator’s behavior. In the current experiment, such ambiguity was manifest in the assessment of Spot/BDA reports submitted by experimental participants as they were maneuvered through a complex environment with simultaneous presentation of multiple entities that could or could not constitute a military-relevant threat. In order to assess performance measures such as “response time” and “report accuracy” in such an environment, someone or some automation (ultimately, a combination of both) had to judge specifically which environmental entity was the subject of each report.[*]

In addition to the complexity inherent in the scenario was complexity due to variability in the performance of each individual experimental participant. That is, although significant time and effort were expended to carefully script controlled-but-complex scenarios, there were no

* In the current study, there were over 1400 Spot/BDA reports across 17 participants operating in four scripted scenarios. Indeed, assessing each of the 1400+ reports would be a laborious and error-prone process if left to the default solution of human manual intervention.

guarantees at run time that a given participant would respond to the scenario as it was scripted (and as they were trained). In fact, as with many well-designed experiments on human performance, error-free behavior was not expected and would have been less informative about experimental manipulations were it observed. Among ways in which the participants could diverge from the "script" were behaviors such as: (1) reporting on targets out of scripted sequence, (2) correctly or incorrectly identifying characteristics of an entity or entities designed to be a non-threatening decoy, (3) incorrectly identifying aspects of an otherwise correctly identified threat, (4) failing to report on a scripted event, and/or (5) generating duplicate reports for events already spotted and reported. Indeed, even with the advantage of complex tools for continuously keeping track of all entities on which the human could be reporting (e.g., automated line of sight [LOS] entity detection), conclusively associating a single entity with a single Spot/BDA report in light of possible human error/behavioral variation was rather challenging.

To facilitate this step in the data reduction, that is, the association of reports with entities, an algorithm was developed using both quantitative and Boolean logic. In essence, the algorithm assessed all possible entities present in the environment at the time a Spot/BDA report was initiated – defined as spanning 10 s prior until the moment when the report was initiated – and then determining which entity was most likely to be the subject of the report. The main assumption on which the algorithm was developed was that, at any given moment in the simulated scenario, there was a finite subset of entities among which one was distinguished as the entity of interest. Second, it was assumed that the entity being reported on would, on average, share many characteristics in common with the information contained in the report (even in the presence of reporting errors) and would have characteristics relative to the observer/vehicle (such as distance and bearing) that would increase the likelihood that it would be identified as the subject of the report. The challenge, therefore, was to develop a methodology that would accurately characterize the likelihood of a positive relationship between a given report and each candidate entity in the environment at the time the report was generated.

Quantitatively, the algorithm made use of variables describing relatively simple assumed relationships between entity characteristics and the likelihood of being a subject of a given report. For example, entities located within a small range of the clock position reported were judged as more likely to be the subject of the report than those that were in another location. Likewise, entities that were located at a shorter range from the vehicle were considered to be more likely to be observed and reported on than those that were further away. In short, for the quantitative logic portion of the algorithm, each entity was scored on seven such variables, with increases in each variable representing an increase in the likelihood of being the subject of the current report. To be conservative, most variables were binary in nature; either the characteristic was present (score = 1) or absent (score = 0). However, as will be discussed momentarily, some variables appeared to demand more complex functions to account for presumed nonlinearities in perceptual and cognitive aspects of target identification.* Regardless, all scores were confined to a maximum range of ±1 to facilitate standardization of the relative contributions of each variable to the total score indicating likelihood that a given entity was the subject of the report being assessed.

Table E-1 provides a description of the variables used in the quantitative portion of the algorithm. Of the variables used to represent the likelihood of being the subject of a given report, only three were specified as non-binary; these variables included TypeMatch, viewing proportion and InRange. Each of these variables was included specifically to account for some of the complexities (i.e., apparent nonlinearities and cross-variable interactions) of judging whether an entity was a good candidate for a particular report. For each of the binary values, the null answer (false) was represented with a zero. One of the binary variables, is moving, used a value of 0.5 rather than 1.0 to describe an answer of "true" simply because it was judged as having less weight than the other decision factors; that is, the intermediate value of 0.5 was used to give a small advantage to moving entities under the assumption that moving entities were slightly more salient during target detection tasks than are static entities. For the other two binary variables, Current LOS and InClockPosition, buffers were used to define acceptable ranges of error before assigning a null value.

Table E-1. Summary of the variables used to assign quantitative scores to each entity serving as a candidate for a particular Spot/BDA report.

Variable Name	**Description**	**Range of Values**
IsATarget	Was the entity a scripted target?	{0, 1}
IsMoving	Was the entity moving?	{0, 0.5}
Current LOS	Was current line of sight (LOS) established?	{0, 1}
InClockPosition	Was the entity in the reported threat position?	{0, 1}
TypeMatch	Did the entity match the reported threat type?	{–1, 0, 0.5: 1.0}
ViewingProportion	During what proportion of the previous 10 s was the entity visible in one of the operator's sensor views?	{0: 1}
InRange	Was the entity within viewing range of the simulated vehicle?	{–1: 1}

Note: Curl brackets denote an inclusive range of values adopted by a variable, a comma separates an instance of a discrete value or set of values from another discrete value or set of values, and a colon is used to represent a continuous range between two numbers in a set.

For current LOS, an entity was judged as having a current LOS if such was established within the prior 1.5 s. This was included to give some advantage to entities that were most recently within LOS of the vehicle. Similarly, it was desirable to give extra weight to entities that were closer to the reported location at the time of report initiation, but because bearing was specified

* Before developing this algorithm, a team comprised of five engineers and scientists scrutinized all Spot/BDA reports associated with one of the four experimental scenarios (383 valid and invalid reports). In this process, the factors used to judge the reports as well as the entities associated with each report were noted and those factors were later translated into elements of the algorithm. The complex functions used for some variables resulted from complex considerations used in the decision process as discussed among the data processing team.

as integer increments of positions on a clock face (1–12), some ambiguity was expected in the correspondence between reported and actual entity clock position. Thus, the in-clock position variable only adopted a null value when the candidate entity was greater than 2 integer increments from the reported bearing. For example, if the reported entity was identified at 1 o'clock and a candidate entity was actually located at 12 o'clock, it was scored with a 1.0 indicating that it was within an acceptable range of the reported bearing. However, an entity located at 10 o'clock would have been judged as too far from the reported bearing to be a likely candidate based on position. Of course, because this was only one of seven variables, a null score on in-clock position (or any other variable for that matter) would not necessarily remove a given entity from candidacy for being the subject of a report; instead, it would lower the likelihood that the entity would end up being judged as the subject of the report. This allowed for positive association to occur despite potential confusion about (or errors in) judging the correct clock label to use for describing the entities bearing.

The type match variable was designed to allow for cases in which the experimental participants became confused when generating a report. For instance, it was possible that an armed human – due to viewing angle and viewing time restrictions – could have been confused for and reported as an unarmed human (or vice-versa). At the same time, given our experimental scenario and participant training protocols, it was highly unlikely that a human target of either type (armed or unarmed) could have been confused with an IED. Moreover, the likelihood of certain types of confusion was increased by the range factor. That is, as viewing distance increased, the visual distinctiveness of armed versus unarmed humans diminished altogether. Therefore, the potential for confusion among different target types was compounded by a dynamic function across viewing distance.

Table E-2 shows the "confusion matrix" implemented for determination of the type match variable in the current study. Arranged along columns are values associated with the known candidate entity types and arranged along the rows are the possible values that could have been included in the Spot/BDA report. The numbers in the different cells give the score allotted to each combination. Where single values are shown, that is what was assigned for that type of match, regardless of vehicle distance from the entity. As can be seen, exact matches were represented with the highest value (= 1.0) and complete mismatches, such as human vs. IEDs, were represented using the minimum possible value (= –1.0). For cases of possible confusion, scores ranged between 0.5 and 1.0. This feature was included to account for the diminishing distinctiveness between armed and unarmed humans as vehicle distance from the entity increased. These "partial credit" scores were determined as an exponential function of the range (x), in meters, from the vehicle to the entity (equation E-1). No score was given in cases where the participant omitted the type variable from his or her Spot/BDA report.

$$1 - \frac{0.5}{e^{0.025x}} . \qquad \text{(E-1)}$$

Table E-2. Depiction of the "confusion matrix" used to account for potential confusion in Spot/BDA reports regarding different entity types.

		Candidate Entity Type			
		Decoy	**Armed Human**	**Unarmed Human**	**IED**
Reported entity type	No answer	0	0	0	0
	Armed human	{0.5: 1.0}	1.0	{0.5: 1.0}	–1.0
	Unarmed human	1.0	{0.5: 1.0}	1.0	–1.0
	IED	–1.0	–1.0	–1.0	1.0

Note: All decoys were unarmed humans but were distinguished from threatening unarmed humans by their behavior as described in the main body of this report.

Figure E-1 shows the form of the function defined by equation E-1. An exponential function was used to capture the fact that as distances grew progressively longer, the distinction between armed and unarmed humans disappeared. That is to say, whether armed or unarmed, if an operator was identifying a human at 200 m, there was no way to distinguish whether that human was indeed carrying a weapon or not. At the longest distances, no score was detracted for confusing armed for unarmed humans or vice-versa. At the shortest distances then, we allowed for partial credit to be given if a human was identified at all, even if incorrectly. Thus, even at a viewing distance of 0 m, incorrectly calling an unarmed human and armed human (or vice versa) would earn a score of 0.5; effectively representing partial credit for a partially correct answer.

Use of the exponential function also allowed avoidance of selecting a breakpoint beyond which all values were to equal 1.0. That is, it was recognized that a linear increase in partial credit with increasing viewing distance would have been a slightly more conservative approach, however, to achieve a similar "plateau" characteristic as the exponential would have required the selection of a point where the score was to level off at 1.0. Absent any data to drive the selection of such a breakpoint, use of a more continuous characterization was considered more appropriate. Using the exponential provided for a gradual transition between a score of 0.5 (partial credit) and a score of 1.0 (full credit) as viewing distance increased.

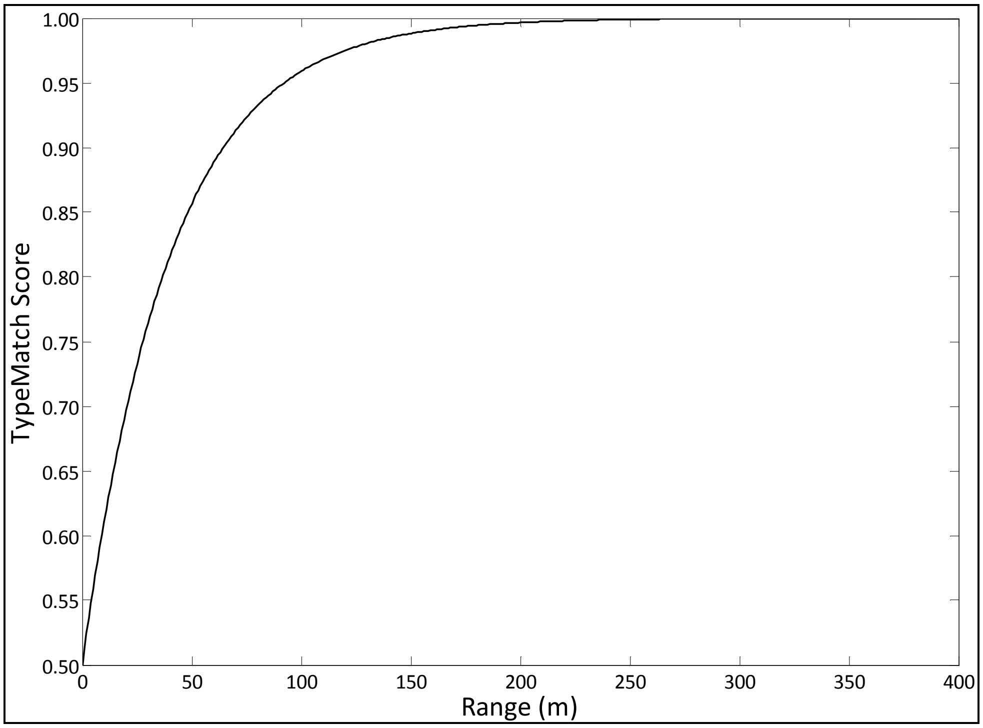

Figure E-1. The exponential function modulating partial type match credit as a function of increasing range.

For reasons similar to those underlying selection of the form of the dynamic allotment of partial credit for the type match variable, an exponential function was used to define the score allotted based on viewing proportion, that is the proportion of the previous 10 s that the entity was in view in either the banner or sensor portal (represented by *x* in equation E-2).

$$1-\frac{1}{e^{10x}} \quad . \tag{E-2}$$

Because experimental conditions varied in terms of what type of visualization was available to the operators, an average viewing proportion was determined from viewing proportions calculated on each individual sensor type. What this variable represented was, for each entity in the local area at the time of generation of a report, how much of the previous 10 s was spent visible either in the sensor or a banner view. The form of equation E-2 is shown in figure E-2. Succinctly, use of the exponential in this case allowed two desired characteristics including: (1) a plateau at a maximum score of 1.0 as viewing proportion increased and (2) a faster accumulation of score for smaller viewing proportions – implying that small increases in viewing time at the lower end of the viewing proportion scale had a greater effect than similar-sized increases at the high end. Otherwise, as with the TypeMatch exponential (equation E-1), there was no intended explicit physical or psychological interpretation to be inferred from the form of these functions or from the specific values of their rate-constants. Both functions were implemented as approximations of the perceptual judgment processes based on how well they worked in the algorithm and should not be interpreted in terms of psychophysical properties of the neurocognitive system.

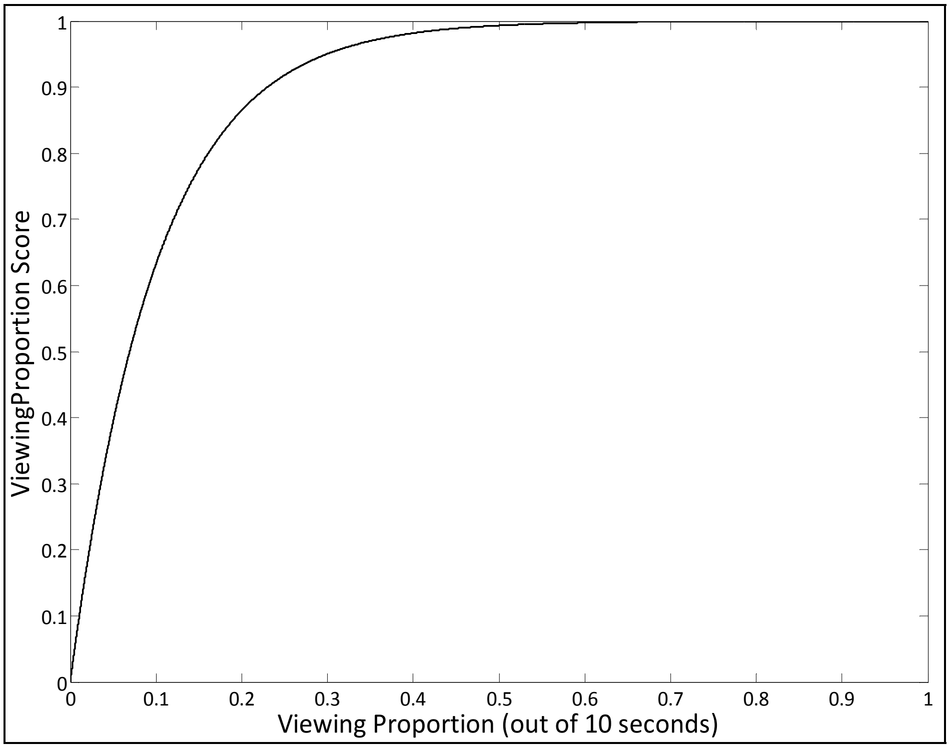

Figure E-2. The exponential function modulating viewing proportion score as a function of increasing proportion of the previous 10 s spent in either the sensor or banner view.

Finally, the variable used to capture how increases in likelihood of being recognized as a target varied as a function of distance from the vehicle to the entity, or Range, a piecewise linear function was used as shown in figure E-3. Unlike with the previous two variables, we utilized two assumptions to drive selection of breakpoints in the linear function for assigning score to specific range values.* That is, the low end of the range (below 30 m; dash-dot line in figure E-3) encompasses all viewing distances encountered during portions of the mission spent in urban zones. As such, the range from 0–30 m was considered the range where all entities should be relatively equally detectable and thus were given an InRange score of 1.0. Conversely, all entities located beyond 200 m were considered equally less identifiable and were thus given

* Whereas the exponential "partial credit" function used in assigning the TypeMatch variable appears to conflict with the piecewise linear function used to assign the InRange score, they were included to separately account for the difference between stimulus detection and object identification. In the current study, the InRange variable was meant to characterize likelihood of detection of any target as a function of distance whereas the TypeMatch variable was meant to characterize variation in object identification (armed vs. unarmed human vs. IED) as a function of distance.

InRange scores equal to –1.0. For all ranges falling in between 30 and 200 m, InRange decreased linearly with a slope of –0.0118 points/meter; this linear portion of the InRange variable transitioned from increasing likelihood (positive values) to decreasing likelihood (negative values) at a range equal to 114.65 m (dashed lines in figure E-3).

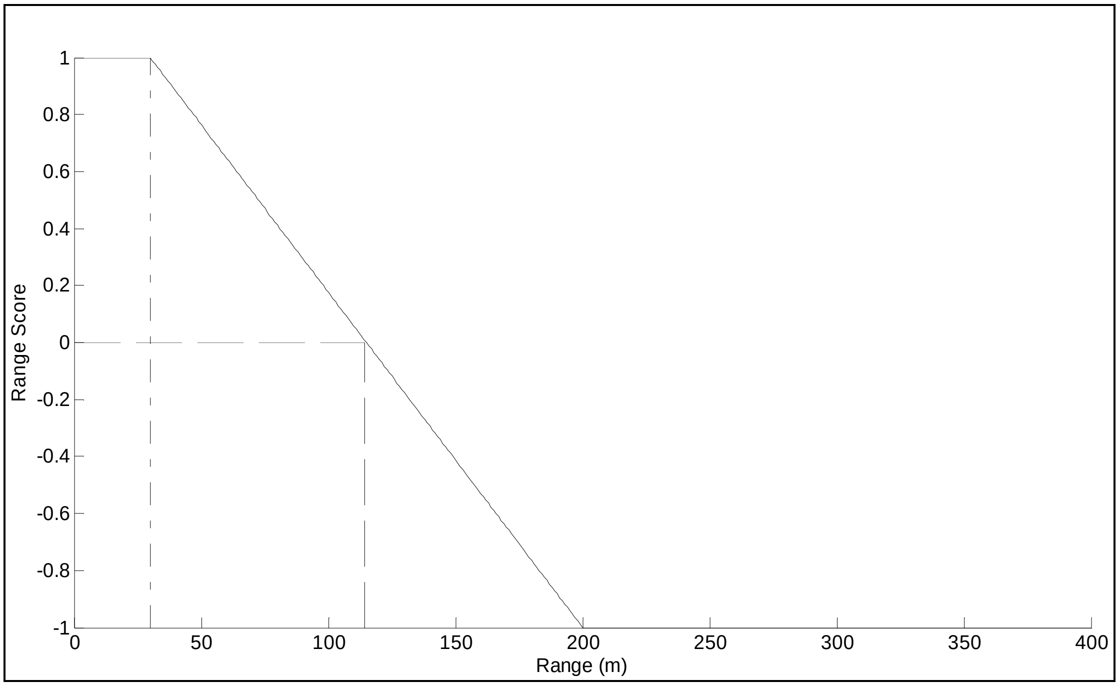

Figure E-3. Piecewise linear function representing change in range score as a function of entity range from the vehicle in meters.

Using all of these variables, a final score for each candidate entity was determined simply by adding all variable values together and then ranking the entities according to their overall score. This step concluded the quantitative logic portion of the algorithm by returning a set of up to five entities for each Spot/BDA report; it was always possible that there were fewer than five, but at least one, candidate entities at the time of a given report. Each of the top candidate entities was accompanied by its script ID number (a unique number indicating the scripted target event of which it was a part) and a confidence value. The confidence value was simply the proportion of total possible points achieved based on the values from each of the variables described above. In short, there were a total of 6.5 points possible (see table E-1) and thus the confidence value was defined as entity total score/6.5.

A Boolean structure was applied to examine relationships between the top candidates provided by the quantitative logic described above in order to determine a final set of suggested entities corresponding to the set of submitted SPOT/BDA reports. This Boolean logic had two exclusionary decision points as well as three basic conditions that could lead to a return of an ambiguous result requiring manual intervention from an informed human supervisor. Shown in figure E-4, the first exclusionary criterion was whether or not the operator pressed the cancel button at any point during completion of the report. If the cancel button was pressed, the report was flagged to be omitted from the data analysis phase and, therefore, the report-entity association algorithm was not applied. The second exclusionary criterion was applied within the quantitative logic section of the algorithm. As each entity was processed for assignment of its quantitative score, its associated raw viewing proportion variable was assessed to determine if it was actually zero, indicating that an LOS could be drawn between the entity and the vehicle but the entity was never in the LOS of any sensor that the operator was looking at. If a given entity showed a raw viewing proportion of zero, which meant the operator had no opportunity to report on it because they had no chance to view it and thus, that entity was excluded from consideration as a candidate for the current report being assessed.

Other than the two exclusionary criteria, there were three conditions under which the algorithm would return an "ambiguous result" for a given report. The presence of an ambiguous result was intended to be a flag to the experimenter that manual intervention was needed to resolve the subject of the report in question; this was meant to provide a means of catching potentially false positives and prevent them from distorting the analysis of performance variables. As can be seen in figure E-4, the three conditions leading to the return of an ambiguous result included: (1) the case where more than one entity received a perfect score from the quantitative logic, but each entity was associated with a different target event, [*] (2) the case where none of the top candidate entities had confidence values exceeding a preselected confidence threshold (α), or (3) the case where the top two entities were associated with different target events, received confidence scores above the preselected confidence threshold, but they did not differ from one another by the preselected difference threshold (δ). In each of these three cases, the algorithm essentially "decided" that no one entity stood out as an obvious choice for the target event associated with the submitted report. It should also be noted that, because the Boolean phase of the decision algorithm proceeded iteratively on a report-by-report basis, it was able to adjust the candidates for successive reports based on those selected in the current iteration. That is, there were many instances when certain entities appeared as candidates for multiple reports. In order to allow for likely, association for successive reports.

* Note that multiple entities could be part of the same target event, such as during an ambush or during the suspicious movement of a crowd.

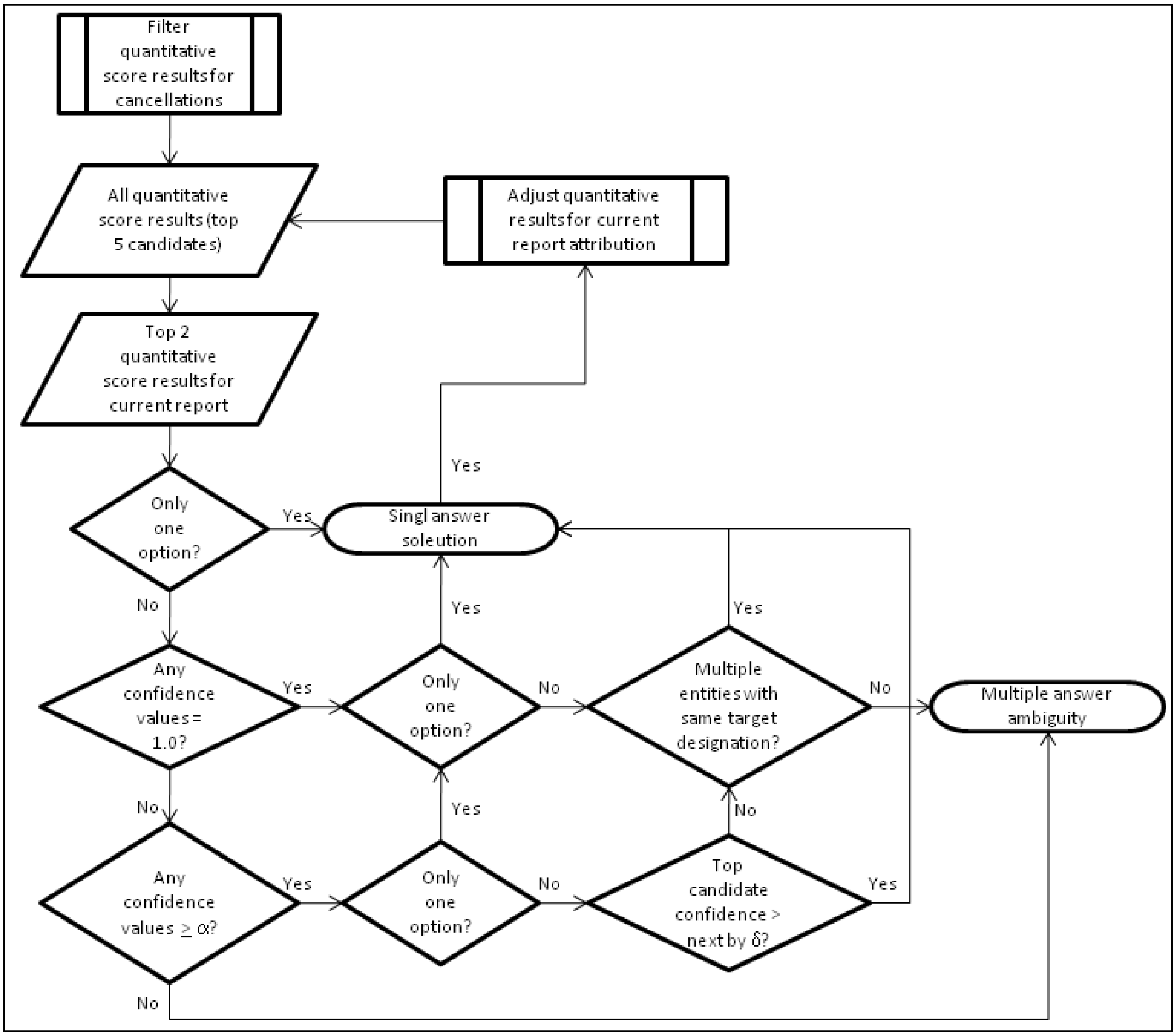

Figure E-4. A summary of the Boolean logic used to finalize recommendations as to which entities were associated with which SA report.

The final processing step in this entity-report association algorithm was a manual one; that is, trained human experimenters examined all reports to verify correct selections. Of course, owing to the use of this algorithm, this manual processing was more efficient and less error prone than having a human (or team of humans) complete the entire process unassisted by computational intelligence. The primary roles of the human experimenters in this manual step involved: (1) manual selection of the correct entity for reports deemed by the algorithm as a multiple entity ambiguity and (2) verification of reports deemed by the algorithm as false alarms (i.e., reports submitted about decoy entities). In the process of resolving the ambiguities, the human experimenters were also able to perform quick checks to verify correct associations between all entities and targets made by the algorithm.

Worthy to note, after the current algorithm was developed, its results were compared against the selections made by a team of five engineers and scientists on 339 valid reports for one of the experimental scenarios (some of the original 383 were omitted from this comparison for technical problems associated with a single participant who did not perform the reporting task consistent with instructions). This comparison showed 83% of the reports were accurately associated with entities by the algorithm, 12% were returned as ambiguous and thus required manual confirmation/intervention, and 5% were judged as erroneous associations. Of the 5% erroneous associations, 2.6% were false negatives that would be caught by manual verification (as they were flagged as “false alarms” and would trigger manual inspection) and another 2.4% were false positives, of which ~1/2 would have been rectified as a consequence of resolving false alarms and ambiguous results. So the ultimate rate of “undetectable error” by the algorithm, according to this assessment, was ~1%–2% of all reports.

Appendix F. Statistical Analysis Details

Because the current experiment aimed to achieve a relatively high degree of ecological validity (i.e., correspondence with real military operational environments), there were many factors and variables that acted to influence the human performance results. Variable factors that were intrinsic to the experiment included the following:

1. Condition – This was the main independent variable of interest in the current study. There were four unique display conditions within which experiment participants performed their threat detection task.

2. Scenario – Four scenarios were created by the teamwork of two different engineers and it was considered that there could have been subtle (or not so subtle) differences between scenarios.

3. Threat Target Type – There were three different types of targets utilized as both threats and decoys. These included armed humans, unarmed humans, and IEDs. Each of these target types, because they had their own unique behaviors, were considered to potentially exert unique influence on threat detection performance.

4. Threat Location – In addition to different threat target types, each threat onset in a different location (bearing) relative to the vehicle. While it was possible to quantify threat location based either on clock position (1–12) or based on which sensor view (1–6) it first appeared in, initial analyses suggested that the effect on performance was structured based on whether the threat onset to the front or to the rear of the vehicle. Thus, when entered as a factor into the statistical models, threat location was treated as a binary variable with 0 = front and 1 = rear onset.

5. Threat Range – As with location, range (distance from the target to vehicle) was also an important variable. There were several ways in which range could have been accounted for which included: range at threat onset, range at threat report, minimum range after onset, and range category (near = range < 50 m, mid = 50 m $\leq$ range < 100 m, far = range $\geq$ 100 m). For some variables such as threat detection, initial analyses indicated that minimum range at threat report was the most highly influential variable while for others, such as reaction time and accuracy, analyses ended up focusing on range category.

6. Vehicle Mobility – There was reasonably strong qualitative and quantitative evidence that the state of vehicle motion (stationary, on the move) had a significant overall impact on performance and, in some cases, interacted with other influential variables.

7. Target Mobility – While less influential in an overall sense, whether or not targets were moving was another variable that interacted with other factors in determining detection performance.

8. Inter-threat Interval – Because of the manner in which the scenario unfolded dynamically, there was a variable amount of time in between each successive threat presentation. As

such, it was considered important to assess whether or not the time in between threat presentations was a factor that needed to be accounted for in the statistical models.

9. Viewing Time – Owing to different display conditions having a different number of "viewing portals" (sensors, banners), an important variable that was affected was the amount of time entities were present to and viewable by the operator. In condition A, for example, there was only one sensor view and the operator had to have the sensor pointed directly at a target for some minimum amount of time before he detected it. However, in the other conditions, the presence of a banner increased this viewing time because it could be capturing imagery that was not present in (or was redundant with) that shown in the sensor view. Finally, because the scenario was dynamic and both the vehicle and targets were potentially moving, there was also a variable amount of time that each threat appeared to the operator. As it turns out, this was a major factor affecting threat detection performance for most variables.

10. Threat Environment – because the reporting criteria varied depending on whether the vehicle was located within the city or within the outskirts, it was considered possible that threat environment (city, outskirts) affected some of the performance results.

11. Interactions – beyond the individual, independent effects of the previously mentioned 10 factors on performance was the fact that each of the factors could have impacted or exerted an influence on any of the other variables. For example, viewing time could have had a different effect for threats presented to the front than for threats presented to the rear (which would indicate a viewing time x threat location interaction). As such, at a minimum, initial models considered all possible two-way (two variable) interactions as well as main effects. As models were simplified, some higher order (three-way) effects were assessed as possibly influential.

Given the number of potentially influential variables outlined above, it was desirable to assess all factors for potential significant influences over variations in human performance, at least at an omnibus level. Yet, with as many variables as were potentially influencing the data, it was also impossible to run a full factorial statistical model that would simultaneously account for all possible sources of variation. In short, a methodology was needed that would allow for a selection of a subset of variables/factors that were most significantly impacting the performance results. Thus, from the outset, statistical analysis of the data from the current experiment was both concept-driven as well as exploratory and required a stepwise, decision-based approach that would lead to the application of tractable statistical models which could be applied for reliable data analysis.

As described in the main body of the current report, the basic methodology was relatively straightforward. As a first pass, a large omnibus model that included a majority of the factors thought to be most related to threat detection performance was assessed. Afterwards, successive models were examined that were simplified by the removal of parameters (variables) that did not appear to exert a significant influence on the results. These successive models were iteratively assessed and parameters removed until the point where a smaller, more tractable model was achieved. The results of the initial models are presented in tables F-1 through F-3.

Table F-1. Overall linear mixed model results for reaction time. Shaded rows indicate parameters considered significant and were passed on to subsequent model reduction steps.

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	1162	846.916	0.000
Condition	3	1162	1.149	0.328
Threat environment	1	1162	30.881	0.000
Threat target type	2	1162	45.915	0.000
Range (mid, near, far)	2	1162	114.797	0.000
Vehicle mobility	1	1162	19.211	0.000
Condition × threat environment	3	1162	0.576	0.631
Condition × threat target type	6	1162	0.185	0.981
Condition × range	6	1162	0.698	0.652
Condition × vehicle mobility	3	1162	0.374	0.772
Threat environment × threat target type	2	1162	8.176	0.000
Threat environment × range	2	1162	22.138	0.000
Threat environment × vehicle mobility	1	1162	0.402	0.526
Threat target type × range	4	1162	10.046	0.000
Threat target type × vehicle mobility	1	1162	0.822	0.365
Range × vehicle mobility	2	1162	11.020	0.000
Condition × threat environment × threat target type	6	1162	1.046	0.394
Condition × threat environment × range	6	1162	1.566	0.154
Condition × threat environment × vehicle mobility	3	1162	0.291	0.832
Condition × threat target type × range	12	1162	0.971	0.474
Condition × threat target type × vehicle mobility	3	1162	0.636	0.592
Condition × range × vehicle mobility	6	1162	0.152	0.989
Threat Environment × threat target type × range	4	1162	5.468	0.000
Threat target type × range × vehicle mobility	1	1162	0.003	0.954

Table F-2. Overall linear mixed model results for accuracy. Shaded rows indicate parameters considered significant and were passed on to subsequent model reduction steps.

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	1162	2438.068	0.000
Condition	3	1162	0.655	0.580
Threat environment	1	1162	0.451	0.502
Threat target type	2	1162	15.039	0.000
Range	2	1162	0.638	0.529
Vehicle mobility	1	1162	5.646	0.018
Condition × threat environment	3	1162	0.731	0.534
Condition × threat target type	6	1162	1.515	0.170
Condition × range	6	1162	0.665	0.678
Condition × vehicle mobility	3	1162	0.691	0.558
Threat environment × threat target type	2	1162	0.756	0.470
Threat environment × range	2	1162	0.986	0.374
Threat environment × vehicle mobility	1	1162	0.227	0.634
Threat target type × range	4	1162	8.475	0.000
Threat target type × vehicle mobility	1	1162	4.787	0.029
Range × vehicle mobility	2	1162	0.420	0.657
Condition × threat environment × threat target type	6	1162	0.349	0.911
Condition × threat environment × range	6	1162	0.581	0.745
Condition × threat environment × vehicle mobility	3	1162	0.512	0.674
Condition × threat target type × range	12	1162	1.323	0.199
Condition × threat target type × vehicle mobility	3	1162	1.264	0.285
Condition × range × vehicle mobility	6	1162	1.199	0.304
Threat environment × threat target type × range	4	1162	1.273	0.279
Threat target type × range × vehicle mobility	1	1162	1.576	0.210

Table F-3. Summary of the logistic regression model and partial sums-of-squares f tests for threat detection. Shaded rows indicate parameters considered significant and were passed on to subsequent model reduction steps.

Source	SS Difference	df_{reg}	$F(df_{reg}, 2499)$	Sig.
Scenario	6	3	2.0495	0.1049
Participant type	63	12	5.8368	0.0000
Condition	164	24	7.6205	0.0000
Threat target type	95	18	5.8685	0.0000
Threat location	103	11	10.3761	0.0000
Vehicle mobility	27	10	3.0226	0.0008
Target mobility	11	8	1.5188	0.1453
Minimum range after onset	368	9	45.4816	0.0000
Inter-threat interval	23	7	3.7049	0.0005
Viewing time	287	8	39.8973	0.0000
Participant type × condition	6	3	2.2665	0.0788
Participant type × threat target type	5	2	2.9305	0.0536
Participant type × threat location	1	1	1.5531	0.2128
Participant type × vehicle mobility	0	1	0.0003	0.9855
Participant type × target mobility	0	1	0.0004	0.9843
Participant type × min. range	0	1	0.2169	0.6414
Participant type × inter threat interval	0	1	0.0692	0.7925
Participant type × viewing time	0	1	0.0770	0.7814
Condition × threat location	24	3	8.8535	0.0000
Condition × vehicle mobility	1	3	0.3976	0.7548
Condition × target mobility	6	3	2.0875	0.0998
Condition × min. range	5	3	1.9371	0.1214
Condition × inter-threat interval	1	3	0.4231	0.7364
Condition × viewing time	17	3	6.3462	0.0003
Threat target type × threat location	26	2	14.6222	0.0000
Threat target type × vehicle mobility	6	2	3.1169	0.0445
Threat target type × target mobility	1	2	0.4936	0.6105
Threat target type × min. range	23	2	12.9712	0.0000
Threat target type × inter-threat interval	2	2	0.8896	0.4109
Threat target type × viewing time	4	2	2.3579	0.0948
Threat location × vehicle mobility	7	1	7.4223	0.0065
Threat location × target mobility	29	1	32.6323	0.0000
Threat location × min. range	2	1	1.7350	0.1879
Threat location × viewing time	27	1	30.2822	0.0000
Vehicle mobility × target mobility	0	1	0.0000	1.0000
Vehicle mobility × min. range	4	1	4.7783	0.0289

NO. OF COPIES	ORGANIZATION
1 (PDF only)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC HRR 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL CIM L 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL CIM P 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL D 2800 POWDER MILL RD ADELPHI MD 20783-1197

ABERDEEN PROVING GROUND

1	DIR USARL RDRL CIM G (BLDG 4600)

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY – HRED RDRL HRM A J MARTIN MYER CENTER BLDG 2700 RM 2D311 FORT MONMOUTH NJ 07703-5601
1	ARMY RSCH LABORATORY – HRED RDRL HRM C A DAVISON 320 MANSCEN LOOP STE 115 FORT LEONARD WOOD MO 65473
2	ARMY RSCH LABORATORY – HRED RDRL HRM DI T DAVIS J HANSBERGER BLDG 5400 RM C242 REDSTONE ARSENAL AL 35898-7290
1	ARMY RSCH LABORATORY – HRED RDRL HRS EA DR V J RICE BLDG 4011 RM 217 1750 GREELEY RD FORT SAM HOUSTON TX 78234-5002
1	ARMY RSCH LABORATORY – HRED RDRL HRM DG K GUNN BLDG 333 PICATINNY ARSENAL NJ 07806-5000
1	ARMY RSCH LABORATORY – HRED ARMC FIELD ELEMENT RDRL HRM CH C BURNS THIRD AVE BLDG 1467B RM 336 FORT KNOX KY 40121
1	ARMY RSCH LABORATORY – HRED AWC FIELD ELEMENT RDRL HRM DJ D DURBIN BLDG 4506 (DCD) RM 107 FORT RUCKER AL 36362-5000
1	ARMY RSCH LABORATORY – HRED RDRL HRM CK J REINHART 10125 KINGMAN RD BLDG 317 FORT BELVOIR VA 22060-5828
1	ARMY RSCH LABORATORY – HRED RDRL HRM AY M BARNES 2520 HEALY AVE STE 1172 BLDG 51005 FORT HUACHUCA AZ 85613-7069

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY – HRED RDRL HR MP D UNGVARSKY POPE HALL BLDG 470 BCBL 806 HARRISON DR FORT LEAVENWORTH KS 66027-2302
1	ARMY RSCH LABORATORY – HRED RDRL HRM DQ M R FLETCHER NATICK SOLDIER CTR AMSRD NSC WS E BLDG 3 RM 343 NATICK MA 01760-5020
1	ARMY RSCH LABORATORY – HRED RDRL HRM AT J CHEN 12350 RESEARCH PKWY ORLANDO FL 32826-3276
1	ARMY RSCH LABORATORY – HRED RDRL HRM AT C KORTENHAUS 12350 RESEARCH PKWY ORLANDO FL 32826
1	ARMY RSCH LABORATORY – HRED RDRL HRM AS C MANASCO SIGNAL TOWERS BLDG 29808A RM 303A FORT GORDON GA 30905-5233
1	ARMY RSCH LABORATORY – HRED RDRL HRM CU 6501 E 11 MILE RD MS 284 BLDG 200A 2ND FL RM 2104 WARREN MI 48397-5000
1	ARMY RSCH LABORATORY – HRED FIRES CTR OF EXCELLENCE FIELD ELEMENT RDRL HRM AF C HERNANDEZ 3040 NW AUSTIN RD RM 221 FORT SILL OK 73503-9043
1	ARMY RSCH LABORATORY – HRED RDRL HRM AV S MIDDLEBROOKS 91012 STATION AVE RM 348 FORT HOOD TX 76544-5073
1	ARMY RSCH LABORATORY – HRED RDRL HRM CN R SPENCER DCSFDI HF HQ USASOC BLDG E2929 FORT BRAGG NC 28310-5000

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY – HRED RDRL HRM DW E REDDEN BLDG 4 CL 60 FORT BENNING GA 31905-5400
1 (CD only)	ARMY G1 DAPE MR B KNAPP 300 ARMY PENTAGON RM 2C489 WASHINGTON DC 20310-0300
	ABERDEEN PROVING GROUND
5	DIR USARL RDRL CIM G S FOPPIANO RDRL HR L ALLENDER T LETOWSKI RDRL HRM B J LOCKETT RDRL HRS D B AMREIN

NO. OF COPIES	ORGANIZATION
	ABERDEEN PROVING GROUND
3	DIR USARL RDRL HRS C T FENG RDRL HRS E K COSENZO A EVANS