2010 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM
MODELING & SIMULATION, TESTING AND VALIDATION (MSTV) MINI-SYMPOSIUM
AUGUST 17-19 DEARBORN, MICHIGAN

HUMAN DIMENSION CHALLENGES TO THE MAINTENANCE OF LOCAL AREA AWARENESS USING A 360° INDIRECT VISION SYSTEM

Jason S. Metcalfe[1], Ph.D., Keryl A. Cosenzo, Ph.D.[2], Tony Johnson[1], Bradley Brumm[3], Christopher Manteuffel[1], A. William Evans III, Ph.D.[2], and Terrance Tierney[3]

[1] DCS Corporation, Alexandria, VA;
[2] U.S. Army Research Laboratory – Human Research and Engineering Directorate, Aberdeen Proving Ground, MD;
[3] U.S. Army Tank Automotive Research, Development, and Engineering Center, Warren, MI.

ABSTRACT

Maintenance of local security is essential for the lethality and survivability in modern urban conflicts. Among solutions the 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.

INTRODUCTION

The ability to maintain local area security is considered critical for the modern Warfighter, who must increasingly conduct complex mobile operations in densely populated urban environments. Underlying the maintenance of local area security is the achievement of situational awareness (SA) sufficient to enable decision making in risky circumstances, often under significant time pressure. As an underpinning of SA, the ability to sustain real-time, full-spectrum local area awareness (LAA) is thought to be facilitated by technologies providing for hemispherical (360°/90°) visualization of the surrounding battlespace. Unfortunately, the ability to provide full-spectrum visual data only offers a partial solution to the problem, largely because the human factors issues that are related to creating 360° LAA from semi-redundant, multi-modal informational sources are only marginally understood. Therefore, the objective of our research program was to develop displays and techniques to facilitate 360° LAA for the Warfighters. Specifically, the present study focused on developing a better understanding of how to facilitate LAA through display solutions for 360° visual data.

Urban Conflict and Situational Awareness

For the Warfighter, the modern strategic context is defined by persistent conflict with a diverse combination of actors that use unpredictable, unconventional, and typically violent means to achieve what are often described as ideological ends[1,2]. While much of the current domain of conflict rests in developing nations, trends towards population growth and globalization have led to a situation in which the primary stage on which conflicts will be enacted is urban in nature[3,4]. More to the point is that, until recently, conventional military capabilities have not been developed and optimized for fighting in urban contexts[5].

The trend towards urbanization of conflict comes with a set of constraints that impose changed and, in some cases, altogether different challenges to those who are developing systems to support modern military operations. Two interacting factors that are intrinsic to any urban area of operbations and must be considered are: (1) the presence of many people, particularly noncombatants, and (2) the multidimensional complexity imposed by a built environment consisting of a dense, diverse, and irregular array of structures[6]. The joint significance of these two factors lies in the need to provide systems supporting a high degree of dynamic SA for the Warfighter. That is, because U.S. Forces follow stringent rules of engagement with respect to intervening amidst civilians and because the urban setting adds an exponential amount of complexity to decision making, the relative cost of failures in maintaining SA is increased dramatically in the urban setting[5,6].

In response to these challenges, the U.S. Army has been modernizing and developing systems aimed at supporting enhanced SA through leveraging advanced technologies for providing real-time, full-spectrum LAA while simultaneously enhancing survivability by means of having Soldiers conduct missions from within "buttoned-up" (closed-hatch) armored vehicles [1,7]. It is thought that enhanced Intelligence, Surveillance, and Reconnaissance capabilities with advanced sensors and systems will serve a fundamental role toward the facilitation of achieving and maintaining SA during the execution of future urban operations[6,8]. Despite the progress of advanced sensor systems, as well as the current enthusiasm for capabilities afforded by modern technologies, significant questions remain regarding how to best structure information so as to enhance overall Warfighter-system performance.

Human Dimension Challenges and 360° LAA

A critical question that has received little attention as engineers proceed with the development of systems intended to support the maintenance of 360°/90° visualization is whether, given sufficient quality and amount of data, a human being is even capable of achieving a sustained, full-spectrum awareness of his or her surrounding environment, especially as that environment becomes increasingly complex and dynamic. That is, whether a human is capable of attending to and holding in working memory many independent elements distributed throughout a 360° environment remains open for investigation.

Humans are limited in the amount of information that can be processed cognitively at any given moment in time. Based on research in fundamental cognitive neuroscience, it is clear that humans are limited in processing visual information in terms of the distribution of targets throughout attentional space[9], the number of items that can be held in visual short term memory/working memory[10], and the temporal dynamics required to process newly received visual information[11]. Together, these basic elements of visual perception limit the ability of an individual human to actively maintain SA in a dynamic 360° environment.

Beyond the level of basic perception, others have identified cognitive biases and contextual factors that diminish the ability to maintain SA. For example, Endsley, Bolté and Jones[12] outlined 8 "demons" to the achievement of SA, at least half of which were associated with what appear to be cognitive biases such as "attentional tunneling" and the so-called "out-of-the-loop syndrome". Research assessing SA using different display configurations in a battlefield context validated that, if presented in certain ways, humans will tend to exhibit such biases. Specifically, when using a visual display that presented environmental data from an egocentric perspective that required manual panning to obtain full 360° information, it was shown that Soldiers "tunneled into" particular aspects of the display (the forward field of view, or FFOV) to the detriment of attention to areas located at the periphery[13]. This cognitive tunneling was associated with poor threat detection and was manifest in reduced accuracy and slow response time. Follow-on research indicated that switching the panning requirement from one necessitating manual interaction to one that was automatic did not facilitate performance, but instead resulted in emergence of the "out-of-the-loop syndrome" wherein the operators became more passive about and overconfident in their acquisition of SA and, ultimately, performed worse at their LAA task[14].

Our experiment examined the factors that can influence performance of a threat detection and identification task using a simulated system of sensors providing 360° LAA. Four different display configurations, representing variations of two 360° Indirect-Vision Display (IVD) concepts were assessed as U.S. Soldiers and civilians performed threat detection using a display representing the view from within a simulated moving vehicle during execution of a presence patrol. Beyond examination of the effectiveness of the display configurations, the dynamics and composition of the simulated environment were structured to allow for statistical investigation of human factors issues related to perception during threat detection. Issues addressed included, but were not limited to, the vehicle-relative range and location of threat onset, amount of time the threat was visible (and potentially viewable by the operator), the nature of the threat (whether armed human, unarmed human or an IED), and whether task dynamics such as operational orders and vehicle mobility status would impact the use of the IVD system.

METHOD

Participants

Seventeen male individuals (n = 7 active duty Army Soldiers and n = 10 civilians) participated in the experiment, which was conducted at the U.S. Army Tank Automotive Research, Development, and Engineering Center (TARDEC) Systems Integration Laboratory facilities in Warren, MI.

Participants were recruited from the local TARDEC population.

Procedure

Prior to the start of the experiment, the participants were briefed on the purpose and procedures and were read a Volunteer Agreement Affidavit as well as a required brief regarding confidentiality as indicated on Department of the Army (DA) Form 5303-R. In order to ensure that personal information could not be revealed, each form was reviewed upon receipt by one of the investigators. Participants who agreed to take part in the study signed the Volunteer Agreement Affidavit and then completed questionnaires for acquisition of information regarding demographics and computer experience.

Next, the participants were given an overview of the baseline scanning system and the three advanced 360° scanning systems. The experimenter reviewed the functionality of the interface configurations, and how to use the scanning and reporting features for local security and target identification. Once the participant had received the introductory brief as well as reviewed a series of training slides, he was then required to complete two training missions: one with Configuration A and one with Configuration D (see "Display Concepts and Configurations" below). Participants were required to repeat the training missions until they could identify at least 50% of the targets as determined by a trained experimenter who was present during the training missions; most participants satisfied this criterion on the first training run in each condition. Participants were offered the opportunity to repeat their training missions if they desired, and they were allowed to do so as many times as they wished in order to ensure complete comfort with the interface technology.

After training was complete, participants then completed the four experimental missions, followed by the completion of a workload assessment. Once the experimental missions had been completed, participants were given a usability questionnaire and an exit interview to assess their overall impressions and preferences with regards to the system.

Mission Execution and Experimental Design

Each mission was 12 minutes long with 8 minutes spent moving through the urban core and 4 minutes in the outskirts of the virtual city. The mission objective was to identify targets through the use of the sensor systems on board the simulated vehicle. Once a target was identified the participant had to send a threat report. For each threat, the participants had to press buttons indicating the object type (armed human, unarmed human, or IED) and location (vehicle-relative clock position in integer increments from 1-12). These threat reports were time tagged (per button press) and served as the basis for calculation of threat identification rates as well as report accuracy and response time.

Within the urban core of the city, participants were required to report armed humans, IEDs, identified high value targets (HVTs), and unarmed individuals performing suspicious behaviors (such as approaching the vehicle). When in the city outskirts, participants' received audio communications that they were in a designated "free fire zone" and during training they were informed that this meant that, in addition to IEDs and HVTs, they were to report on all humans, armed or unarmed, regardless of behavior.

A diverse set of events and targets were developed based on discussion with two Subject Matter Experts on current U.S. operations in the Middle East. In addition, significant independent research for development of events was conducted involving review of materials from several sources including Soldier "blogs", current periodicals and news sources regarding present-day military activities, U.S. Army photo archives, and formal documents such as Army field manuals and other such operational/doctrinal materials.

Figure 1 provides graphic examples of a subset of the entities used within each of the prescribed events. Depending on the needs of each event, these entities were programmed to be either standing still or moving in a scripted manner. Participants were shown examples such as these during their initial experiment brief and training. Table 1 summarizes all events incorporated into the mission scenarios and represents all those thought to be most military-relevant. As described, Table 1 lists a set of over 100 basic events, with a total of 38 per scenario that should have been reported as threats.

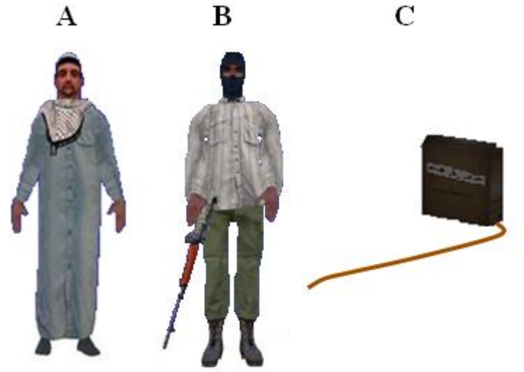

Figure 1: Examples of experimental stimuli including (A) unarmed humans in native dress, (B) armed humans in native dress and (C) IEDs, indicated by visible fuses

Label	Description	Constituent Entities/Events	# per Scenario
Crowd	Group of 10+ non-threatening humans	Market, protest, children playing, 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 were 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 were vehicle motion pauses	2 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 of 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

Table 1: Summary of events used in the simulation environment

Display Concepts and Configurations

The different screen configurations used in this experiment were chosen as four combinations of two different visualization themes including a sensor portal and a banner (Figure 2). The sensor portal was a 64° × 48° (width × height) window that represented the view from of a single directional camera (6 selectable cameras were simulated as comprising the 360° IVD sensor system); the participants could snap (“pan”) from one camera view to another using a sensor control graphic embedded in the top right corner of the sensor portal.

The banner display was representative of a static, “stitched” view composed of a combination of three individual camera views. The banner thus provided a wider horizontal field of view (hFOV = 180°) for the operator with poorer resolution. The banners, when used, showed either a fixed forward or a fixed rearward 180° hFOV image. When available, the banners were always presented in combination with the sensor portal; the design concept was: through a banner an operator could gain an awareness of the overall area and then use the better resolution of the sensor portal to further interrogate a specific area of interest. For both front and rear banners, the view was established as equivalent to that which one would have with “eyes out the window” in the relevant direction. This is particularly important to keep in mind as the lower (rear) banner, which was only used in one experimental condition, was a left-right reversed image as compared with the front banner.

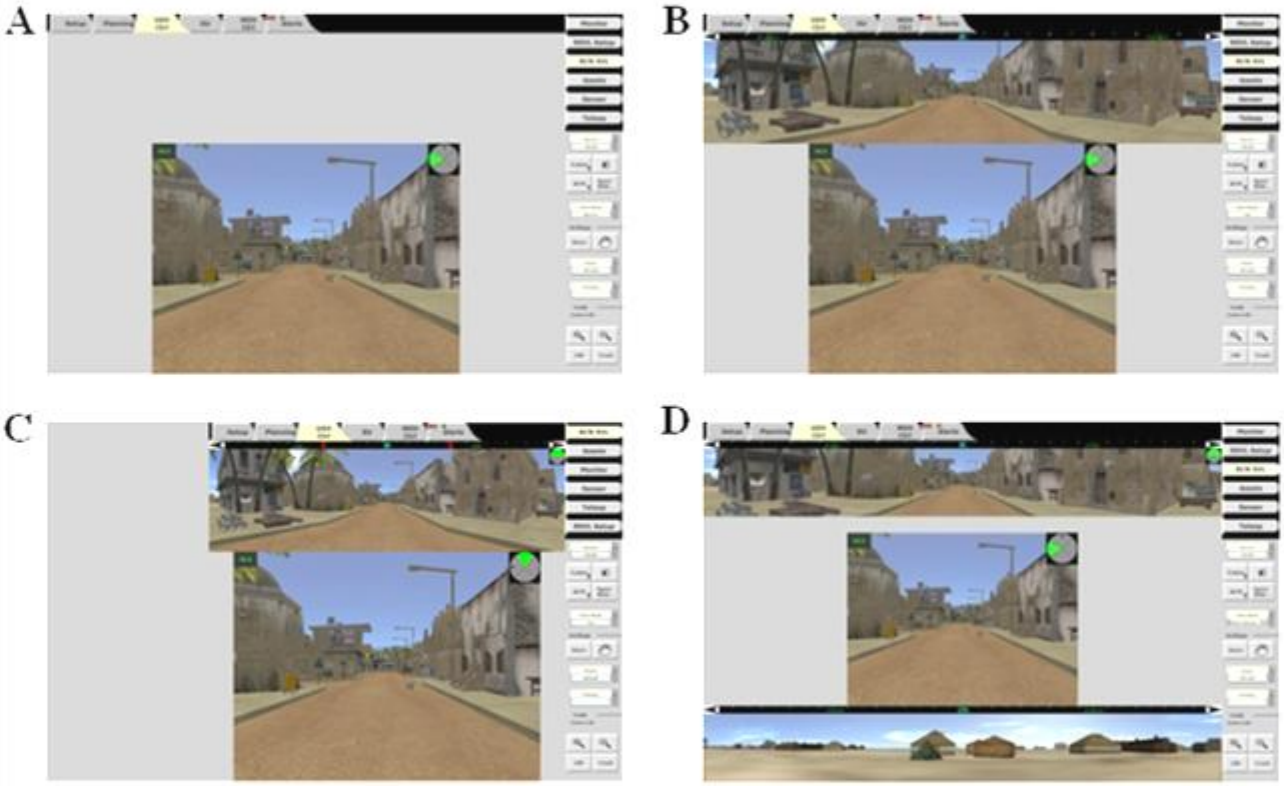

Figure 2: Display configurations (each corresponding with A-D as labeled) used in the current experiment. See text for details.

Four display configurations were tested (Figure 2). Configuration A (Figure 2A) was the sensor portal displayed in full screen mode. This mode utilized the single 64° × 48° FOV sensor at 1024 × 768 resolution. The participants could snap ("pan") the view of the sensor portal through 6 discrete steps to obtain 360^{o} LAA using the circular control graphic in upper right hand corner of sensor portal window. Configuration B (Figure 2B) was the sensor portal combined with a single banner. This included the 64° × 48° FOV sensor as well as a 180° hFOV forward facing banner located above the sensor display. As with Configuration A, the participants could snap the view of the sensor portal through 6 discrete steps to obtain 360^{o} vision, however the banner remained fixed on the forward 180° throughout the mission. Configuration C (Figure 2C) involved the same banner and sensor portal combination as Configuration B. However, in this condition, screen space was reserved for small portals that would provide additional mission-relevant functionality. The reduced screen space resulted in a consequent reduction in overall resolution for the banner as compared with Configuration B; the same 180° hFOV was displayed in each, but the banner in Configuration C was compressed to accommodate the small portal space. The sensor portal, on the other hand, remained at full resolution. Configuration D (Figure 2D) involved the use of the sensor portal and two banners, one for each of the front (top) and rear (bottom) views. As with the other configurations, the participants could snap the sensor portal through 6 discrete steps to obtain 360^{o} vision. Neither of the two banners had selectable views; they remained fixed on either the front or rear 180^{o} throughout the mission. As with Configurations A and B, this was a full screen mode with no screen space reserved for small portal functions. Despite using full screen mode, however, all images had to be reduced to accommodate use of overall screen space.

Simulation and Data Acquisition Environment

The chosen experimental crewstation surrogate was a laptop computer that provided the same screen size (17") and resolution (1920x1200) as the next-generation crewstation that has been under development by TARDEC and its partner, the U.S. Army Communications-Electronics Research, Development and Engineering Center (CERDEC). Participants used a mouse to interact with the system when using all interface configurations.

The laptop served as the Warfighter Machine Interface (WMI) and provided sensor displays and sensor controls. Sensor displays (sensor portal and banner) were handled using the system's internal graphics processor, while sensor control was enabled through an interactive graphic positioned in the sensor portal (see circular graphic overlaid on the top right corner of the sensor portal window in each image contained in Figure 2).

An Embedded Simulation System (ESS) was connected to the WMI and provided two major services: communication throughout the distributed simulation system and the vehicle dynamics model. The communication provided by the ESS was crucial for synchronized information interchange, while the vehicle dynamics model was used to maintain the position of the vehicle within the simulated environment. The Intelligent System Behavior Simulator (ISBS) was another key process that ran on the laptop and provided vehicle control. The ISBS was responsible for driving the vehicle by using route plan information received from the ESS. The ISBS then provided the ESS with appropriate actuator data to keep the vehicle inside pre-specified route conformance parameters, thus emulating a simplified autonomous mobility system.

A separate Event Server controlled execution of all events using vehicle location information from the ESS in combination with pre-defined virtual trip lines, which served as triggers for event onsets. The trip lines were used to indicate when the participant reached prescribed locations in the current scenario and then scripts would run to place the appropriate entities in the correct locations, behaving in specified ways. 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, respectively. In cases where an audio command was needed, a pre-recorded audio file appropriate to the event was triggered. In cases where image generation was needed, the Scenario Populator received events from the Event Server and output Distributed Interactive Simulation (DIS; defined under IEEE Standard 1278) packets representing a set of entities moving around the database. These DIS packets were interpreted by the ESS, which passed them back to the internal graphics processor running the WMI for rendering of appropriate entities and behaviors within the virtual environment.

All events were recorded and time-stamped in event log files for later use during data reduction and analysis. 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 a commercial eye tracking system (Smart Eye; data not reported herein). 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 Line of Sight Log, which was used as a basis for calculating threat detection rates, response times and accuracy. The subjective measures were recorded separately and data files for each corresponded to the information provided by the NASA-TLX, the usability assessment, and post-experiment exit interview.

Data Reduction and Analysis

The experiment was run as a within-subjects design. There was one primary independent variable of interest, display configuration, with four levels. 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, representing qualitatively different combinations of the mission events described above. Assignment of each display condition to a mission scenario was counterbalanced and the order of condition presentations was randomized across participants. Other contextual and task factors, such as target characteristics (range, location of onset, etc) and mission context (city vs. outskirts, stationary vehicle vs. on the move) were factored into the statistical analysis for assessment of human dimension influences on threat detection performance and, more importantly, served to assist the interpretation of how the different display configurations were used.

All dependent measures were calculated from a reduced, collated, and time-synchronized set of variables extracted from the raw data set composed of the six log files described above. 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 threat report. That is, for each threat report, it had to be determined 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 threat 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 threat report – entity association algorithm is left to its detailed description elsewhere[15].

Following the application of the threat report – entity association algorithm, a number of variables were calculated from the reduced data. The dependent measures were delineated as task performance measures, physiological measures (eye tracking based) and subjective questionnaire responses. *Task Performance Measures* included measures of threat detection, response time, report accuracy, and sensor portal usage. This last construct, sensor portal usage, was assessed in multiple ways including the number and rate of times the sensor view was changed and the proportion of mission duration spent looking at each of the six possible sensor views. *Physiological Measures* were recorded as additional indicators of performance, but due to the extensive processing required and scope limitations on the current presentation, will not be discussed further here. *Subjective Questionnaire Responses* included the NASA-Task Load Index (NASA-TLX [16]) as well as questions regarding the usability of the interface configurations and the responses to an open-ended exit interview.

The data were analyzed using a combination of techniques, each appropriate to the statistical properties of the variables being assessed. For threat detection, which was a binary variable representing detection (reported or not) on each of the 2585 possible valid threat events, multiple logistic regression was used. For the continuous variables, which included response times, report accuracy, sensor usage statistics and the NASA-TLX scores, analyses were conducted using linear mixed-model regression. In all cases, regression model building steps were included that preceded formal analyses. That is for each variable, a variety of statistical models were entertained that included a diverse selection of independent variables and covariates. Model building proceeded from most complex (including the greatest number of variables, covariates and interaction terms), through a process similar to backward selection, to the most simple models that explained the significant sources of variation in the performance data. In all cases, the final statistical model that was used for analysis only included terms that were significant. Further details regarding the analytic technique can be found in the full technical report[15].

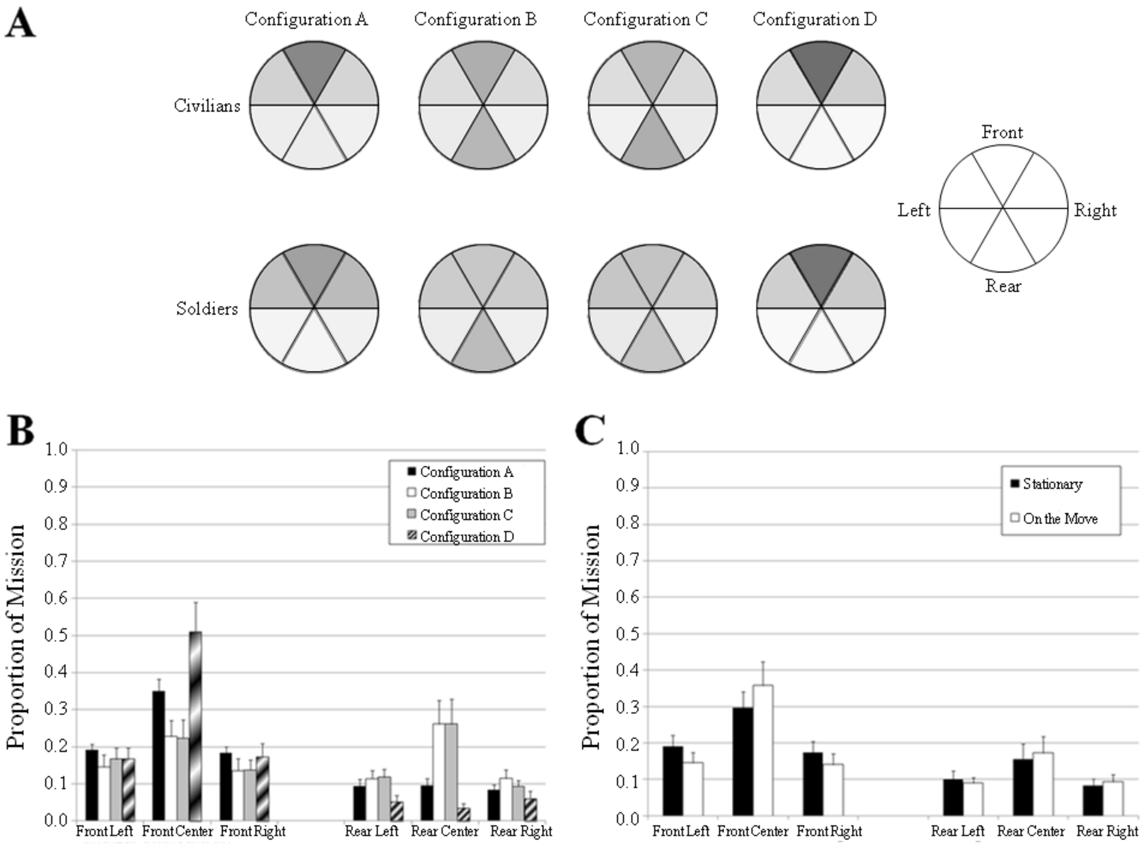

Figure 3: Results for usage of the central sensor portal. (A) Proportional usage as a function of sensor view direction, display configuration and participant type (Soldier vs. Civilian); darker shading indicates greater proportion of use. (B) Interaction effect for the Sensor Portal Direction × Display Configuration. (C) Interaction effect for Sensor Portal Direction × Vehicle Mobility.

RESULTS

Active Scanning using the Sensor Portal

Prior to analysis of threat detection performance, it was important to assess the use of the interface. Recall that in all but Configuration D, which had two banners providing total 360° coverage, participants had to use the sensor portal if they were to successfully acquire sufficient LAA. The original design concept was that the participants would make use of the sensor portal to further interrogate areas in which they had identified potential threats through the use of the banners (when present). This first analysis assessed how the sensor portal was actually used throughout each mission.

Figure 3A provides an overview of how the sensor portal was used across the four display configurations as a function of Soldier and Civilian participants. Qualitatively, the pattern of sensor portal usage was that participants tended to disproportionately orient on the forward view, with generalized a bias towards orienting on the central position within the forward view. This pattern even held in Configuration A despite the fact that it provided no other alternative for viewing the rest of the environment. Figure 3B shows a significant Display Configuration × Sensor Portal Direction interaction ($F_{15,\ 815} = 10.09$, $p < 0.001$) indicating that this “front-center bias” was more strongly observed when participants used Configurations A and D as opposed to Configurations B and C. This significant interaction also indicated that in Configurations B and C, participants used the sensor portal to more frequently scan to the rear of the vehicle, supporting the qualitative pattern seen in Figure 3A. Overall, scanning to any of the side views was typically observed less frequently than scans to the central sensor positions. A second interaction, shown in Figure 3C, was detected as a significant Sensor Portal Direction × Vehicle Mobility interaction ($F_{5,\ 815} = 10.33$, $p < 0.001$), which indicated that this “front-center bias” became slightly stronger when the vehicle was in motion as compared to when it was stationary; a finding suggesting that the presence of forward motion further encouraged participants to fixate on the front-center view.

An important question that was not answered by examining proportional time spent viewing each sensor portal direction was that of how much work was required to obtain 360° LAA. That is, the previous analysis focused on what portion of the total mission was spent looking at each sensor view but did not look at how many times the sensor view was toggled throughout mission execution. Such a metric was seen as important in that it was an indicator of how much user-interface interaction was required to obtain LAA. Assessing the absolute number of sensor changes provided an expected pattern of results wherein the participants toggled their sensor view more than twice as many times when using Configuration A as compared with all other display configurations. This indicated that, as a strategy, not only did participants spend a greater proportion of time on the front-center sensor view in Configuration A, but they would toggle back to it quite frequently after looking to either of the two front side sensors. Configuration D revealed the fewest sensor portal view changes. Thus, while the proportional time spent on the front-center sensor view made Configurations A and D appear similar, the low frequency of sensor changes in Configuration D indicated a very different strategy in that it seemed participants relied only on the two banners and, for the most part, did not utilize the sensor portal when the full 360° visual array was present; this was despite the fact that the banners provided poorer resolution for identifying potential threats than did the sensor portal. The sensor change frequency effect just described, shown in Table 2, was statistically significant across Display Configuration ($F_{3,\ 64} = 10.95$, $p < 0.001$).

Configuration	View Changes (changes per mission)	View Change Rate (changes per minute)
A	381.5	28.2
B	171.4	12.6
C	189.2	13.8
D	131.0	9.6

Table 2: Sensor view change statistics as a function of Display Configuration.

Threat Detection Performance

The data from the current experiment provided scant evidence of an independent influence of display configuration on threat detection performance. Performance was influenced by participant, environment, and target characteristics more than by a particular display configuration. One effect, however, provided some insight suggesting that the influence of display condition differed based on the location at which a given threat was initially presented. Specifically, a significant Condition × Location interaction ($F_{3,\ 2538} = 12.99$, $p < 0.001$) appears to have been influenced by variations in threat detection performance to targets presented in the rear of the vehicle. The data suggest that the variation in the amount of performance change due to threat location was driven by how much detection rates were reduced when threats were first presented in the rear cameras. This trend can be seen upon close inspection of the differences between the black and gray bars in Figure 4. The largest decrease in detection performance between targets presented in the front and rear was seen in condition D (31% lower for rear targets), followed by Condition A (28% lower). Given that the smallest change in performance was observed in Condition C (24% lower), 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 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.

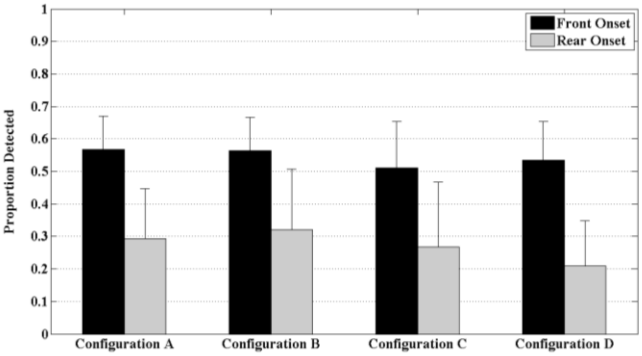

Figure 4: Significant Display Configuration × Location interaction effect

There was also a significant Target Type × Location interaction ($F_{2,\ 2538}$ = 16.46, $p < 0.001$) that revealed an additional effect of target type on threat detection. Shown in Figure 5, the IEDs were always detected more frequently than they were missed, although IED presentations 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. Detection differences existed between armed and unarmed humans as well. Detection rates were 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%).

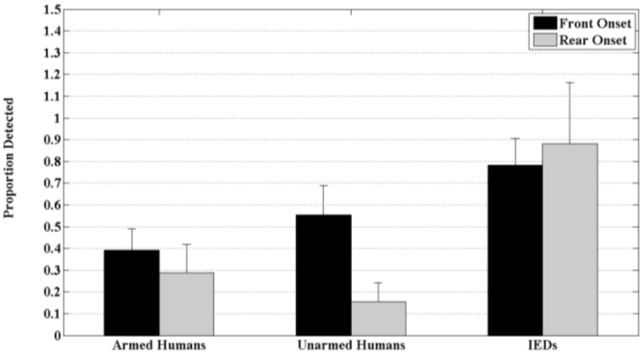

Figure 5: Target Type × Location interaction

Based on prior mathematical analyses of the optical performance characteristics of the sensors that were simulated in this study, it was also expected that detection performance would be reduced as a function of the range at which the threats were presented. This expectation was supported by the data. Detection performance declined as threat range increased, particularly for human targets. The Target Type × Range interaction ($F_{2,\ 2538}$ = 16.48, $p < 0.001$) was most likely to have been due to the lack of longer range observations for IED threats. Demonstrated in Figure 6, all IED targets onset and were maintained within a fairly close range of the vehicle, at or below 25 meters, whereas the human targets were distributed across a range from 5 to 180 meters.

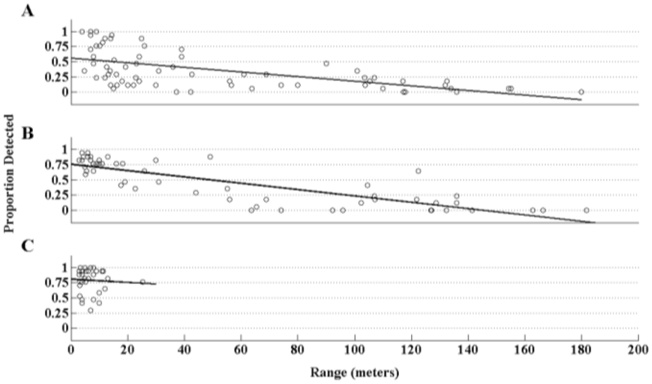

Figure 6: Average threat detection performance as function of Range and Target Type. Target types include (A) armed humans, (B) unarmed humans, and (C) IEDs.

Some evidence was also present for an interaction between armed and unarmed human threats. Overall detection performance at close range was poorer for armed as compared with unarmed humans (compare Figure 6A with Figure 6B at ranges below 40 meters). 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 (i.e. staring directly at or moving towards the vehicle) whereas armed humans were considered threats by virtue of carrying a weapon. That unarmed humans were defined as threatening by moving towards the vehicle while an armed human could have been standing still or even moving away from the vehicle and still be considered a threat points to a possible difference in target salience. That is, motion towards the vehicle may have served as an added cue helping participants identify the unarmed human threats.

Although threat detection rates provided a base level of information regarding performance, without examining additional variables the understanding of the acquisition of 360° LAA would be incomplete. That is, while providing useful high-level information, threat detection rates on their own could not be informative regarding the cognitive and behavioral processes by which 360° LAA was formed. As a first step towards understanding such processes, additional information may be obtained by examining both the temporal properties of response generation (i.e. through response times) as well as the pattern of errors associated with different task conditions (i.e. through report accuracy).

Threat Response Time

Examination of the response times revealed a pattern wherein the performance data were largely reflective of the amount of time available to view and subsequently decide on the nature of the threat. For example, as shown in Table 3, a significant Range × Location interaction ($F_{2,1213} = 11.878$, $p < 0.001$) indicated that the influence of range varied depending on whether the threat had onset in the front or the rear of the vehicle. The increase in response time that was observed across range was considerably greater for threats presented to the front than for threats presented to the rear. A reasonable inference that follows from this is 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 was that, due to forward vehicle motion, the threats would get further away. Conversely, it makes sense that participants waited a little longer to make a decision about threats in front of the vehicle because the consequence of waiting longer was that, on average, the targets would get closer and more readily identifiable.

Location	< 50m	50 – 100m	> 100 m
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)

Table 3: The interaction between Range and Location (mean ± standard error in parenthesis)

In similar fashion, response times were seen to increase as target range increased. Specifically, as indicated by a significant Range × Vehicle Mobility interaction ($F_{2,1213} = 5.21$, $p < 0.01$), this range-based increase in response time was greater when the vehicle was moving as compared to when it was stationary (see Table 4). While the vehicle was moving it was beneficial to wait longer to report on further away targets because as one waited longer the threats would, on average, get closer to the vehicle and become easier to identify.

Vehicle Mobility	< 50m	50 – 100m	> 100 m
Stationary	4.05(.46)	7.65(.89)	8.16(.50)
Moving	6.10(.26)	14.41(.42)	21.41(.48)

Table 4: The interaction between Range and Vehicle Mobility (mean ± standard error in parenthesis)

The two interactions just discussed appear to have fallen out from a pattern that was reflected in a significant generalized relationship between response time and a variable that was denoted as Viewing Time. Simply, Viewing Time was the sum total time that a given entity appeared in any one of the displays (Viewing Time = time visible in a banner + time visible in the sensor portal). Because of the observation of a relatively strong interaction effect for Viewing Time × Location on basic threat detection performance ($F_{1,\ 2538} = 33.92$, $p < 0.001$), the linear mixed model analyses for response times 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 pervasive influence over threat detection performance and, in particular, as a primary determinant of response time. This relationship is illustrated in Figure 7.

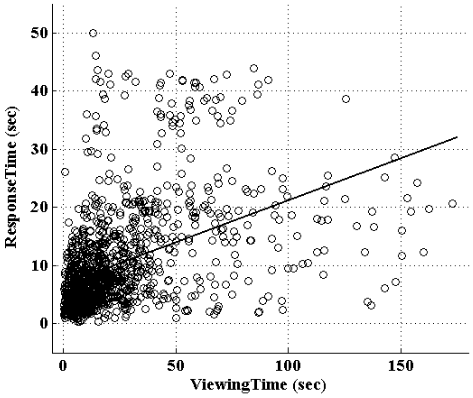

Figure 7: Response time as a function of Viewing time.

Report Accuracy

Report accuracy was scored as a combined value based on whether the participants responded correctly in terms of the type of threat (armed human, unarmed human or IED) as well as the vehicle relative location (bearing) as reported in clock positions in integer increments between 1 and 12. For threat type accuracy, participants received either a score of 1 or 0 indicating that their reported threat type either did or did not match the actual entity type. For location accuracy, participants were scored as correct if their reported location was within ± 1 integer increment of the actual clock position and, likewise, were scored as incorrect for all other reported positions. An overall accuracy score was the average of these two component scores with 0 indicating an incorrect response, 0.5 indicating that one component was correct, and a 1 indicating that both components were correct.

Linear mixed-model regression indicated two significant interactions for accuracy. A Target Type × Range effect ($F_{4,\ 1230} = 14.55$, $p < 0.001$), appeared to have been due to a range influence that was specific to the armed human threats but was not present for the other two entity types. As displayed in Table 5 the accuracy scores were generally high and because they averaged values greater than 0.5, were indicative that, more often than not participants were accurate on both location and threat type components. However, for threats that were armed humans, there appeared to be a range-based attenuation of accuracy where, for ranges greater than 100m, accuracy approached 50%.

Target Range	**Armed Humans**	**Unarmed Humans**	**IEDS**
< 50m	88.0 (1.61)	88.7 (2.63)	85.4 (1.51)
50-100 m	72.9 (2.69)	85.1 (2.90)	87.4 (2.86)
> 100 m	59.2 (3.26)	85.5 (5.58)	94.4 (4.31)

Table 5: The Target Type × Range interaction for report accuracy. (Means ± standard errors in parenthesis)

The second main result was an interaction of Target Type and Vehicle Mobility ($F_{1,\ 1230} = 31.05$, $p < 0.001$). This particular interaction appeared to be driven by two aspects of the data. First, no IEDs were detected while the vehicle was stationary thus resulting in a blank cell in the analysis. Second, however, was the observation that vehicle mobility influenced the detection of armed human threats in a manner opposite to the detection of unarmed human threats. In short, unarmed humans appeared to have been slightly more accurately identified when the vehicle was stationary (stationary = 91.5 ± 2.6; moving = 87.4 ± 1.3) whereas armed humans seemed to be detected much less accurately under similar circumstances (stationary = 40.3 ± 5.1; moving = 82.9 ± 1.3).

Subjective Questionnaire Responses

In an attempt to thoroughly characterize the participants' experience using the different display configurations in the present experimental task, two subjective questionnaires were employed along with an open-ended exit interview. The questionnaires included an assessment of the subjective workload via the NASA-TLX and an assessment of the system usability via a custom-written Usability Questionnaire. While the NASA-TLX proved to be uninformative in that no statistically significant differences were detected between the different display configurations, the Usability Questionnaire and Exit Interview both shed some light on the preferences of the participants.

For the Usability Questionnaire, participants ranked the various display configurations in terms of which they preferred to use. The rankings were from 1-4 with 1 being the most preferred and 4 being the least. The resulting rank order was: Configuration D (1.59), Configuration B (1.65), Configuration C (2.88), Configuration A (3.88). These rankings were further validated by the results of the Exit Interviews, which are summarized in Table 6. All of the participants preferred to have a banner solution when completing the experimental task used in the present study.

Perhaps more interesting was how the responses to the Usability Questionnaire and Exit Interview related to the performance results of the study. For example, a majority of the Soldier participants (71.4%) and nearly half of all participants (47.1%) wanted a zoom capability, which one may infer was related to the detrimental effects of target range on both threat detection rate and response times. An interesting disconnect, however, was that few participants indicated a need for rear-facing sensor or a need for some form of continuous panning, which seemed to indicate a lack of self-knowledge regarding the "front-center" sensor use bias that was observed so strongly in the data as well as a lack of self-knowledge regarding poor task performance on threats appearing to the rear of the vehicle. Likewise, while the average threat detection rates were only moderate, less than a third of the participants (and only a single Soldier) indicated a sense of being overwhelmed by information during execution of the task. The observed average subjective workload rating of less than 60 would seem consistent with this qualitative self-rating. Thus, one may infer that reductions in performance were legitimately due to a lack of awareness rather than due to overt workload limitations. Of course, because workload was not explicitly manipulated in this experiment, such a conclusion should be treated as tentative until future confirmation with more precise measures of human neurocognitive function for the assessment of 360° LAA and the maintenance of full-spectrum situational awareness.

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

Table 6: List of preferences and concerns organized by participant (1-17). Shaded columns indicate Soldiers. % All is calculated as a percentage of 17 participants and % Sld is a percentage of 7 Soldiers

DISCUSSION

The trends towards globalization and population growth have created new challenges for modern military operations; an urbanization of conflict has demanded changes in military technology and force structure. In particular, because of additional complexity imposed by urban environments, the ability to support the achievement and maintenance of full-spectrum (360°/90°) situational awareness, or SA, has taken a central position of focus for military modernization efforts. While technologies designed to enhance the maintenance of local area awareness (LAA) sufficient to support dynamic, 360°/90° SA have advanced at a breakneck pace over the last several decades, they have generally evolved independent of the requisite understanding of how humans sense, perceive, integrate and apprehend the complex informational arrays that they provide. Indeed, how humans construct and maintain a real-time sense of LAA in complex environments remains both a poorly-defined and a poorly understood question. Yet, in order to enhance LAA abilities for the modern Warfighter, it would seem essential to understand the perceptual-cognitive constraints that affect its dynamics. The present investigation was designed in an effort to understand some of those constraints.

The main results indicated that issues associated with human perception and cognition exerted a far stronger influence over threat detection and identification performance than did the presentation of information through a variety of visualization techniques. Despite this, it was determined that certain interface options had a potential to offset the challenges imposed by natural human perceptual-cognitive function and thus, should be the focus of additional technology development efforts.

Perhaps the most striking observation in the present experiment was the emergence of a front-center bias, discussed elsewhere as “the keyhole effect” or cognitive tunneling[13,14]. Despite the clear instructions to the participants that they were solely responsible for scanning their full surrounding environment (all 360° horizontally), they tended to focus disproportionately on the central aspect of their forward view. This trend appeared particularly strongly when using the only display configuration that provided no alternatives for simultaneously viewing both the front and the rear of their vehicle. That is, using a display configuration that limited viewing of the surrounding world to a selectable set of views, covering 64° (horizontally) at a time, the participants tended to focus most of their scanning activity on the front three sensors (covering the forward 180°) with a specific bias to orient on the front-central view (see Figure 3A, Configuration A). A similar pattern of scanning was observed when the participants used a more advanced set of displays (see Figure 3A, Configuration D), but in this case the neglect of the selectable sensor portal views was understandable given that other options were available for the participants to view the full 360°.

Although there was a clear difference in the usage of the different display configurations, this was not manifest in a strong independent influence of display configuration on threat detection and identification performance. Instead, it seemed that the primary factors influencing performance were those associated with human perception and cognition. For example, there was a generalized reduction in performance when threats onset to the rear of the vehicle. Of course, given that the participants did not look to the rear of the vehicle as frequently as they should have, it is not surprising that a preponderance of threats that were missed were those that first appeared in the rear camera views. Though a small effect, some evidence was observed that this attenuation of threat detection performance for rear-presented targets was less dramatic when display configurations were used that allowed the operators to simultaneously maintain front and rear views. Participants appeared to have less of a performance reduction for rear targets when using Configurations B and C. Of course, that such an effect was not seen with Configuration D, which always provided simultaneous front and rear views, was somewhat surprising. However, one must keep in mind that there was a cognitive transformation involved in using the two banners simultaneously in Configuration D. Because the rear-view banner was left-right reversed as compared with the front-view banner, it is assumed that there was a tendency for participants to poorly integrate information from the lower banner into their overall sense of LAA. Thus, it may have been the case that unique threats presented in the rear banner were misidentified as non-unique and thus, were not reported. Moreover, because of the screen space required to maintain all of the visual information provided in Configuration D, all displays were reduced in size which may have also lead to an increased likelihood of missing threat presentations.

Similar to the location (front vs. rear) issue, there was also considerable variation in performance due to other factors such as the range of threat onset as well as the specific nature (type) of the threat that was presented. Of these factors, the range effect was the most anticipated because there are known distance-related detriments to resolution of optical systems. More interesting than the range effects, however, was the effect of target type on detection performance.

In general, detection and identification performance was the best for IEDs. Overall IED detection rates were approximately 80%, nearly double the combined detection rates for unarmed and armed humans. Perhaps more interesting, however, was the observed pattern of responding for human threats. First, the lowest overall detection rate was for unarmed humans that onset to the rear of the vehicle; this was not an unexpected finding. Second, was that the highest detection rate for humans was for unarmed humans presented to the front of the vehicle; especially those that were presented to the front and at a close range. That unarmed humans were more readily detected than armed humans at a similar range (Figure 6) was definitely a surprise. Referring back to Figure 1, the presence of large weapons in the hands of the armed humans was expected to be an important factor influencing threat detection performance, especially at close ranges. However, what seemed to be more important was whether the threats were moving at a close range to the vehicle. That is, the likely explanation for why unarmed humans were more readily detected was that their behavior is what made them candidates for threat reports. The threatening unarmed humans were typically close to and were engaging the vehicle by either staring directly at it or walking/running towards it; such differences point to a possible effect of target salience, which was not explicitly manipulated in this experiment but should be in future studies of 360° LAA.

Finally, in terms of response time and accuracy, the results were fairly clear in that they pointed towards a cognitive strategy of the participants. The data from the response time and accuracy metrics showed that participants were prioritizing report accuracy over report speed. As noted in the results section, report accuracy was generally high, on the order of 80%. The only reductions in report accuracy appeared to be for armed humans and indications were that these were mostly the armed humans appearing at long ranges where they were likely confused for unarmed humans (the weapons in their hands were less visible as range increased). Although there was little systematic variation in report accuracy, response times appeared to vary much more considerably. Response times scaled with how much time was available to view the threats. Those threats appearing to the front, further away, and while the vehicle was moving were responded to more slowly whereas the converse was true for threats appearing to the rear, closer, and while the vehicle was stationary.

The final recommendation based on this research was that a banner solution mitigates a natural tendency of humans to “tunnel into” the forward field of view, however the mitigation provided by the banner solution was far from complete. Performance still varied dramatically based on whether the threats were presented to the rear of the vehicle, how far away the threats appeared, and what type of threats they were. Some of the results were taken as an indication that target salience was a factor and thus, future study and technology assessments involving explicit manipulations of salience in a similar military-relevant context seem warranted. If salience is shown to be a critical factor in threat detection and identification performance, then additional mitigations possibly involving real-time image processing may be required for enhancement of LAA in operational contexts. Further, the results point towards a need to assist the Soldiers in ways that offset their natural cognitive and perceptual tendencies. Technical solutions, such as better optics or implementation of zoom features may suffice to account for the range-based detriment in detection and identification performance. However, accounting for the “tunneling” bias may require additional types of technologies, such as the development of intelligent systems that detect, in real time, where the Soldiers are looking as well as storing where they have been looking in recent history and then cueing examination of neglected areas of the operational environment. Ultimately, enhancement of SA most commonly takes the form of displays and systems that provide for synthesized information for the user, rather than simply the provision of additional raw data for the Soldiers to parse and integrate themselves.

REFERENCES

[1] McHugh, J. M. and Casey, G. W., "A statement on the posture of the United States Army 2010", (2010).

[2] Speakes, S. M. and Martin, G. M., "Army modernization in an era of persistent conflict", Army Magazine 58(1) , 31-36 (2008).

[3] Metz, S. and Millen, R. A., [Insurgency and Counterinsurgency in the 21st Century: Reconceptualizing Threat and Response], Strategic Studies Institute, United States Army War College, Carlisle, PA, (2004).

[4] Taw, J. M. and Hoffman, B., [The urbanization of insurgency: The potential challenge to U.S. Army operations], The RAND Corporation, Santa Monica, CA, (1994).

[5] [Block by Block: The Challenges of Urban Operations], U.S. Army Command and general Staff College Press, Fort Leavenworth, KS, (2003).

[6] Glenn, R. W., Steeb, R., and Matsumura, J., [Corralling the Trojan Horse: A Proposal for Improving U.S. Urban Operations Preparedness in the Period 2000-2025], The RAND Corporation, Santa Monica, CA, (2001).

[7] Department of the Army, [U.S. Army Weapons Systems 2010], OASA (ALT) Strategic Communications & Business Transformation SAAL-ZG, Arlington, VA, (2010).

[8] Ceranowicz, A. and Torpey, M., "Adapting to urban warfare", Journal of Defense Modeling and Simulation 2(1), 3-15 (2005).

[9] Intriligator, J. and Cavanagh, P., "The Spatial Resolution of Visual Attention", Cognitive Psychology 43, 171-216 (2001).

[10] Todd, J. J. and Marois, R., "Capacity limit of visual short-term memory in human posterior parietal cortex", Nature 428, 751-754 (2004).

[11] Marois, R. and Ivanoff, J., "Capacity limits of information processing in the brain", Trends in Cognitive Sciences 9(6), 296-305 (2005).

[12] Endsley, M. R., Bolté, B., and Jones, D. G., "SA demons: The enemies of situation awareness", in [Designing for Situation Awareness: An Approach to User-Centered Design], Taylor & Francis, London, 31-42 (2003).

[13] Thomas, L. C., Wickens, C. D., and Merlo, J., "Immersion and Battlefield Visualization: Frame of Reference Effects on Navigation Tasks and Cognitive Tunneling", US Army Research Laboratory Technical Report #ARL-99-3/FED-LAB-99-2(1999).

[14] Thomas, L. C. and Wickens, C. D., "Effects of display frames of reference on spatial judgements and change detection", US Army Research Laboratory Technical Report #ARL-00-14/FED-LAB-00-4(2000).

[15] Cosenzo, K., Metcalfe, J., Gordon, S., Johnson, T., Feng, T., Brumm, B., Evans, A. W., and Capstick, E., "Impact of 360° Sensor Information on Vehicle Commander Performance", U.S. Army Research Laboratory Technical Report (2010, submitted).

[16] Hart, S. G. and Staveland, L. E., "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research", in P. A. Hancock and N. Meshkati (Eds.) [Human Mental Workload], North Holland Press, Amsterdamn, 239-250 (1988).