THE EFFECTS OF SYMBOL SIZE AND WORKLOAD LEVEL ON STATUS AWARENESS OF UNMANNED GROUND VEHICLES

John F. Lockett, III

Thesis submitted to the faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Industrial and Systems Engineering

Dr. Tonya Smith-Jackson, Chairperson

Dr. Robert Beaton

Dr. Robert Sturges

25 January 2006

Blacksburg, Virginia

Keywords: Human Factors, Displays, Symbology, Workload, Robotics

THE EFFECTS OF SYMBOL SIZE AND WORKLOAD LEVEL ON STATUS AWARENESS OF UNMANNED GROUND VEHICLES

John F. Lockett, III

ABSTRACT

The objective of this study was to determine which size symbols should be used by the U.S. Army for an operator control unit to indicate the status of unmanned ground vehicles (UGVs). Three sizes of symbols were studied. The symbols subtended 20, 40, and 69 minutes of arc corresponding to 0.116, 0.233, and 0.400 inches high when viewed at a distance of 20 inches from a touch screen. Twelve participants were asked to watch the symbols on a map display and touch one of four UGV symbols when it stopped moving. Different numbers (0, 8 and 12) of distracter symbols with the same height as the UGV symbols appeared during the experimental trials. The time to notice that a UGV symbol had stopped (recognition time) and to touch the screen (response time) were measured. Participants were asked for Subjective Workload Assessment Technique (SWAT) ratings for each combination of symbol size and number of distracter symbols. Errors committed while attempting to touch the correct symbol were counted. Participants made very few errors attempting to touch the wrong symbol.

Results for the time and error measures were as expected for changes in symbol size. As symbol size increased, recognition time, response time, and extra touches decreased. Significant differences were seen in these measures between the subtending 20 and 40 minutes of arc and between symbols subtending 20 and 69 minutes of arc. Also, as expected, subjective mental workload increased as symbol size decreased with differences seen between all symbol size levels. No significant differences were observed for workload manipulation (number of distracter symbols) as measured by time and error. However, SWAT scores did show a significant difference as a result of number of distracters. The differences between 0 and 8 distracters and between 0 and 12 distracters were significant. There was no significant interaction between symbol size and number of distracters for any of the measures. Overall results suggest that symbols smaller than those recommended for keypads may be sufficient for interactive map displays. For static platforms with barehanded operators, symbols that subtend 40 minutes of arc may be sufficiently large to ensure adequate touch screen performance under low to moderate workload conditions.

Acknowledgements

This study was sponsored by the U.S. Army Research Laboratory (ARL) Robotics Collaborative Technology Alliance (CTA). As such, I would like to thank ARL management for their support in the form of the time and resources required to complete the study. I would especially like to thank CTA partners Dr. Marc Gacy and Mr. Tom Engh of Micro Analysis and Design for their help with establishing the relevancy of the study, translating the experimental design into simulation, and rounding up resources.

I am grateful to my thesis committee in all its incarnations. I appreciate the patience and flexibility of Drs. Smith-Jackson, Beaton, and Sturges as I balanced other responsibilities. I would like to acknowledge Dr. Sturges' consistent support and Dr. Beaton for his help in scoping the project. Dr. Smith-Jackson was supportive throughout and graciously stepped in as committee chair after personnel changes in the department necessitated a change.

I would like to thank Mr. Richard Kozycki for his assistance in using the Jack human figure modeling software to determine reach distance. The experimental design and data analysis were improved with the help of Dr. Dallas Johnson in his role as statistical consultant to ARL's human use committee. I very much appreciated the support and encouragement of my colleagues throughout the completion of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

The U.S. Army is developing robotic ground systems to reduce dependence on man-in-the-loop operations (Andrews, Schmidt, and Killion, 2001; Dudenhoeffer, Bruemmer, and Davis, 2001). Until full autonomy is achieved, Soldiers will be required to ascertain the status of the unmanned system and, if necessary, intervene by providing corrective guidance. Status may include information such as whether or not the unmanned system is moving, processing data, damaged, overturned, or disabled. For example, the system may encounter an obstacle that it cannot circumvent and may need direction from the operator before it can proceed with the mission.

There are several classes of unmanned ground systems (UGVs). The operational requirements document (ORD) for the future combat systems (FCS) and Griffin (2004) mention small unmanned ground vehicles (SUGVs), multifunctional utility/logistics and equipment (MULE) vehicles, and armed robotic vehicles (ARVs). There are several types of ARVs. This study will focus on the ARV reconnaissance, surveillance, and target acquisition (RSTA) variant. The ARV RSTA will consist of a chassis platform with payloads that provide various capabilities such as video and advanced sensors. The variant is intended to support tasks such as providing reconnaissance capability in urban terrain; remotely deploying sensors, firing into buildings and other structures, by-passing obstacles and threats, and remotely assessing and reporting battle damage. ARVs must be capable of switching from semi-autonomous to tele-operation control and back again. Much of the development work for ARVs centers on the Experimental Unmanned

Vehicle (XUV) developed for the Office of the Secretary of Defense (OSD) Demo III Robotics program. The XUV is a four wheeled platform weighing approximately 3000 pounds. It is approximately 10 feet long, 5 feet wide, and 4 feet high. The Demo III XUV is shown in Figure 1.

Figure 1. Demo III experimental unmanned vehicle (XUV).

Human monitoring and control of UGVs is accomplished through user interaction with a remote control station called an operator control unit (OCU). OCUs may include computer generated maps with graphical overlays, video feedback, and vehicle steering devices such as joysticks or steering wheels. They are typically custom built for a particular developmental UGV and as such are not often optimized for human computer interaction. However, as part of the U.S. Army Research Laboratory (ARL) Robotics Collaborative Technology Alliance (CTA) (on-line 2003), OCUs are being developed with an emphasis on usability (Dahn and Gacy, 2002). This study is being conducted as part of the ARL Robotics CTA.

Initial plans call for the ARVs to be operated from a stationary enclosed platform such as a manned, stopped vehicle however operation from a moving vehicle is desired. Preliminary studies have shown that the mental workload associated with operating a single unmanned vehicle from a stationary platform is substantial (Schipani, 2002). However, concepts call for the operators of these systems to be responsible for other tasks such as monitoring and sending communications within their own platform or to other units, therefore intuitive and efficient display of status information is critical. Current concepts for an OCU being developed as part of the Robotics CTA employ a graphical user interface (GUI) that presents the direction and location of the ARV as a symbol overlaid on a dynamic map. Change in status of the ARV is indicated by lines of text shown at the bottom center of the OCU GUI as shown in Figure 2.

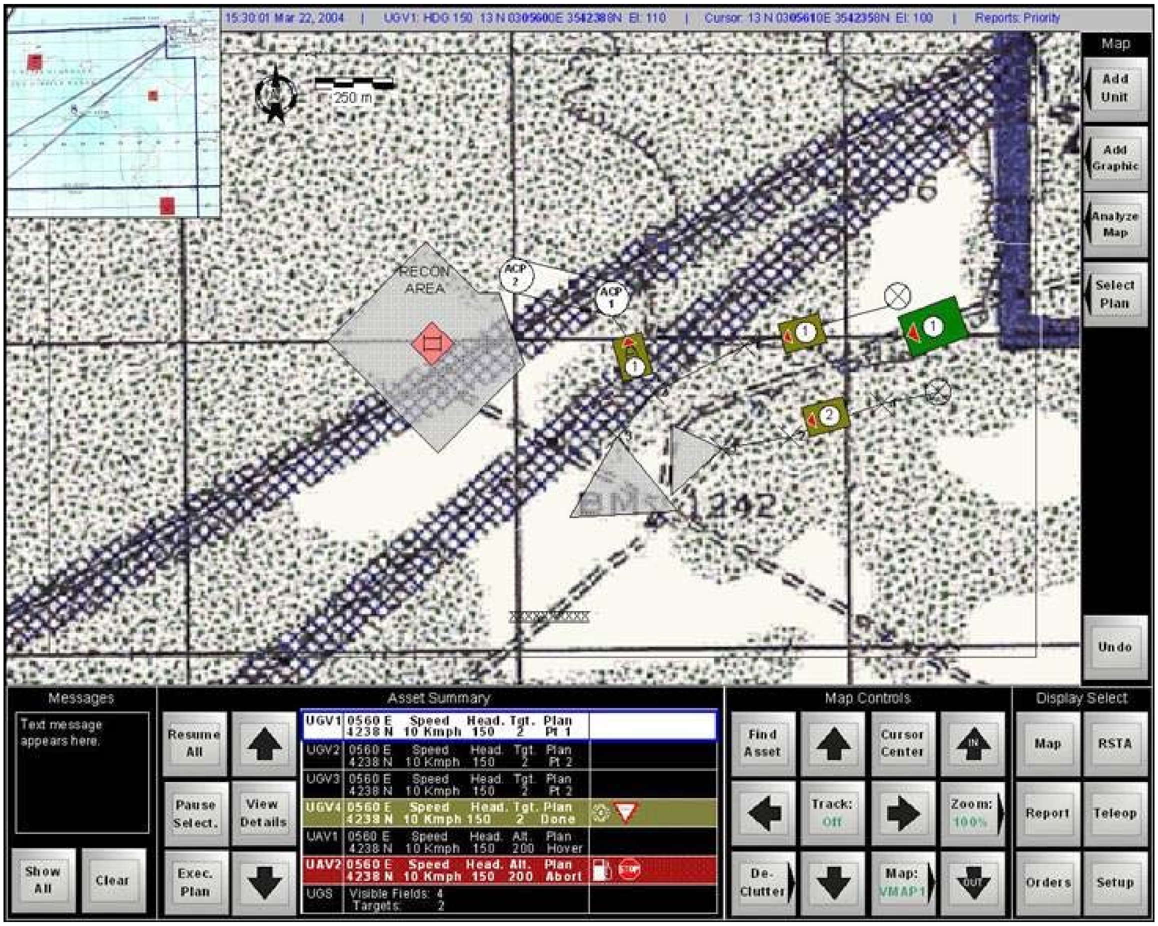

Figure 2. Concept graphical user interface for OCU.

An important issue in the GUI identified by the CTA is determination of the optimum size for the ARV symbols. The answer must consider that the method of interacting with the OCU map display is via a touch screen. The background mental workload of the OCU operator must also be considered. The symbols must be large enough to recognize changes in ARV status and to touch but if smaller symbols are used, less map information will be obscured. Lack of motion is one of the most important indicators that the vehicle status has changed and that intervention may be necessary. This is particularly true of map-based OCUs. The focus of this study is to explore the relationship between symbol size and mental workload in a stationary control platform when using touch screen technology while attempting to follow the constraints of military standards for symbols.

Visual displays operated via touch screen are increasingly common. This research into symbol size and mental workload may have implications for the user interfaces of other remotely controlled vehicles such as those used in undersea operations, mining, space exploration, and explosive demolition. While time pressure may differ in each of these settings, use of cluttered displays and complex symbology is common. If small symbol size proves to be a source of mental workload then this may add to the workload imposed on drivers when using devices such as in vehicle manually operated navigation systems. This may add to the literature on driving and distraction. Complex displays involving maps are common in air traffic control. While direct control of the aircraft is not analogous to UGV control, tracking and checking the status of an aircraft by querying a symbol under moderate to high mental workload is analogous.

Background

Determination of Symbol Size

Military Standard 2525B (MIL-STD 2525B) (US Department of Defense, 1999) covers the design of symbols for military systems. The size of a symbol or point graphic is directly related to the viewing distance of the operator from the display surface on which the object is presented. The following formula can be used to determine object size for a given implementation:

$$L = \frac{(VA)(D)}{(57.3)(60)}$$

where: VA is the visual angle in arc minutes,

D is the viewing distance in inches, and

L is the object size in inches.

Because the crew stations in which the OCU will be used have not been finalized, the viewing distance (D) was determined analytically. An anthropometric analysis model such as Jack version 4.0 (Badler, Phillips, and Webber, 1993 and UGS Corp, 2005) can be used for this purpose. The panel on which the symbols are viewed is an 18.1 inch liquid crystal display (LCD) made by Landmark Technology with a 3M MicroTouch touch screen overlay. The display has a resolution of 1280 x 1024. Because the panel is operated by touch, it should not be located further away than the reach of the segment of the user population with the shortest arm reach. A female figure with 5^{th} percentile functional arm reach and 5^{th} percentile seated eye height was generated and placed in a standard seated position. The anthropometric data was derived from the Army anthropometric survey in 1988 (ANSUR 88) database (Gordon, et al 1989). An 18.1 inch panel was centered on the medial plane of the figure and angled 15 degrees off vertical. The display was positioned at a distance such that a reach across the user's body to the far corner of the display could just be accomplished. It was assumed that the user will be restrained with a seat belt and only shoulder motion could be used to extend the reach. Once the figure and display were positioned, the distance from the user's eye to the center point on the screen was measured using the vector measuring tool in Jack. The distance is 20 inches rounded to the nearest inch. Figure 3 shows the figures and positions used to approximate the viewing distance.

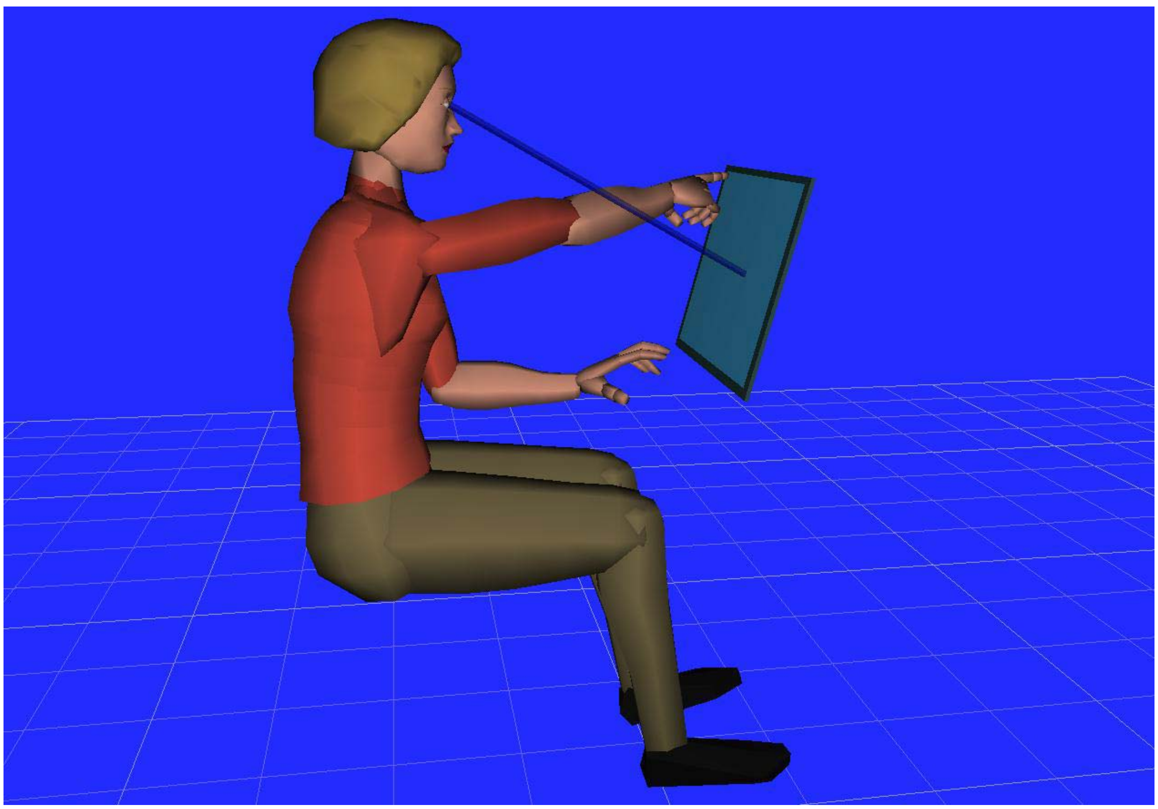

Figure 3. Determination of viewing distance using Jack 4.0.

Based on this analysis, a viewing distance, D, of 20 inches was used for this study. Coincidentally, Sanders and McCormick (1993) and military standard 1472F (MIL-STD-

1472F) (US Department of Defense, 1999) suggest 20 inches as a nominal reading distance for visual display terminals.

MIL-STD-1472F recommends a minimum size of 20 minutes of arc subtended visual angle (arc min.) for distinguishing targets of complex shape, without regard to the effect of color coding. If the viewing distance is 20 inches, then the symbol height will need to be a minimum of 0.116 inches to subtend 20 minutes of arc. MIL-STD-2525B recommends symbols sizes subtending 40 arc minutes (0.233 inches when viewed at a distance of 20 inches). Because the OCU will be activated by touch, larger symbols are more desirable. In an early touch screen study, Beaton and Weiman (1985) found that when determining the size of touch key targets, only vertical size was a significant factor in determining the number of errors. Horizontal size and key separation were not. Targets with a vertical size of 0.4 inches resulted in the smallest number of errors. MIL-STD-1472F indicates a larger size than this -- 0.59 inches for push buttons where the button will not be depressed below the panel; however this standard was developed for physical rather than touch screen buttons. Colle and Hiszem (2004) found that participants preferred and performed better with keys 20mm (0.787 inches) square for a kiosk touch screen. Performance differences were insignificant for a larger size. 0.787 would take up a significant amount of space on the OCU map display. In the Colle and Hiszem study, mean percent error for single digit entry was quite low (less than 3 percent) for even the smallest key size 10mm (0.39 inches) at all spacing distances. Smaller key sizes were not investigated. Based on these results, the vertical size of the symbols investigated in the present study ranged from a minimum of 0.116 to a maximum of 0.4 inches.

Once the critical parameter of vertical size was determined, other parameters were set using prior research findings or standards. For this study, these other parameters such as contrast, luminance, width to height ratio, and format were controlled.

Determination of Symbol Composition

Numerous methods such as matching, appropriateness ranking, comprehension and recognition testing have been developed for creating symbol sets (Easterby and Zwaga, 1978; Gittins, 1986; Wogalter, DeJoy, and Laughery; 1990; Blankenburger and Hahn, 1991; and Horton, 1994), however U.S. Department of Defense systems must use, MIL-STD-2525B which specifies the basic format for military symbols. The basic rationale is that a common symbology across systems reduces error and training requirements. Figure 4 shows the basic components of a military symbol as defined in the standard. Frame size for surface (ground) equipment symbols have a width to height ratio of 1 to 1.

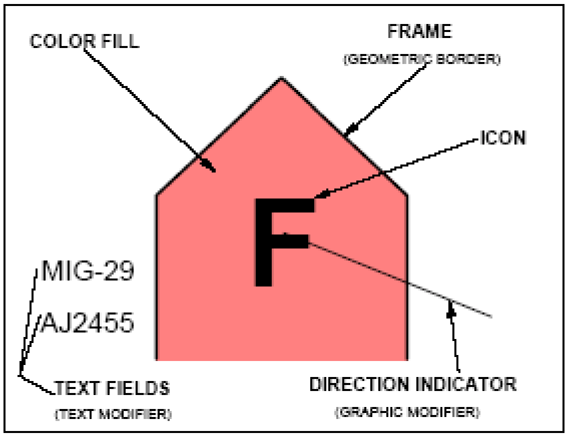

Figure 4. Components of MIL-STD-2525B symbol.

According to MIL-STD-2525B, "In general, medium to large object sizes (i.e., subtending 30-40 arc minutes) are recommended; however, implementers should conduct usability testing to determine the optimum size(s) at which warfighter performance is most effective." The aim of the present study was to conduct just such usability testing.

MIL-STD-2525B does not address specifically unmanned ground vehicles; however, the standard can be applied to aspects of an ARV symbol. The frame shape will be round to indicate surface ground equipment. The frame will be a solid line to indicate that the symbol is showing the present location of the vehicle. The fill color will be cyan (RGB values 0, 255, 255) to indicate that it is a friendly asset. No specific icon for a UGV is specified however, the icon for a ground vehicle is . Many of the icons in the standard include bold text characters added to the center of the icon if a standard shape has not been identified to represent that class of equipment. For example, "A" is used to represent an armored vehicle. For this study, an "R" was used to indicate that the vehicle is robotic. The direction of the movement is indicated by an arrow originating at the icon and projecting past the frame. A text field just outside the frame of the symbol indicates vehicle identification.

Icons versus Symbols and 2-D versus 3-D

Symbology that differs from MIL-STD-2525B has been proposed by various display developers. Most notable are three dimensional symbols and icons that resemble small pictures of the equipment represented. However, research has not established the value of these symbol types. Smallman, St. John, Oonk, and Cowen (2000) found that naming of conventional two-dimensional (2-D) military symbols was faster than that of realistic three-dimensional icons. In a follow-up study (Smallman, St. John, Oonk, and Cowen, 2001), reported that explicit analogue coding is more important than increased dimensionality (3-D views of battle spaces and icons) for speed of search tasks. They suggested creating 2-D symbicons by combining the interior of conventional MIL-STD-2525B symbols with the discriminable, shaped outline of realistic icons. They found that the conventional military symbol interior best encodes the platform information (affiliation, platform identity code) while the icon shape provides rapid identification of platform category. These new symbicons have not been incorporated into the military standard and the present study used symbols following MIL-STD-2525B.

Mental workload in the primary task

Due to resource limitations, scenario based manipulations of mental workload were impractical. Mental workload can however be manipulated in a primary task using several methods. Speed stress involves changing the rate of signal presentation from one or more sources (Cain & Hendy, 1998; Gawron, 2000; and Knowles, Garvey & Newlin, 1953). This method poses challenges in measuring objective performance measures such as time and accuracy. Another method involves changing the load stress by increasing the number of signal sources (Chiles & Alluisi, 1979 and Gawron, 2000). A slight variation on this method is to change the number of distracter symbols that must be attended to when searching for a target stimulus. This method was proposed in a study intended to determine mental workload thresholds (Cain & Hendy, 1998). For the present ARV OCU study, a MIL-STD-2525B symbol similar in shape to the candidate ARV symbol was used as a distracter to increase the difficulty of the task. That symbol represents a friendly force, armored ground vehicle symbol. The ARV and armored vehicle symbols differ only in the text character used for the icon – "R" for the ARV and "A" for the armored vehicle. This increases the importance of the size of the distinguishing character.

Mental workload measures

Several types of measures have been developed to assess mental workload. Gawron (2000) and Damos (1991) provide extensive surveys of the measures and discuss the relative merits of many of them. The measures can be divided into two types – objective and subjective. Objective measures can be further divided into three types – primary, secondary and physiological measures. Primary task measures assess mental workload by examining a participant's ability to perform a main (most important) task. Usually the speed or accuracy of performing the task is measured. The assumption is that when the participant's mental workload increases, performance on the task suffers. However, Hart (1989) and Hockey (1997) have suggested that participants may employ strategies to cope with stress and workload and it is only when those strategies fail that performance suffers. This dissociation between mental workload and performance has been demonstrated (Yeh & Wickens, 1988). Therefore, depending on mental workload levels and possibly history, a primary task measure may not give a complete indication of the robustness of a design. Introduction of a secondary task has been shown to be invasive unless it is a natural part of the real word setting (Hart & Wickens, 1990). Physiological measures such as number of eye blinks, heart rate, or EEG can be intrusive and may confound results. The mental workload state must also be inferred from the physiological measure. Subjective measures are a better alternative.

Subjective measures provide an integrated summary of mental workload from the perspective of the participant usually through rating scales developed for this purpose. Widely used subjective mental workload measures include the National Aeronautics and Space Administration task load index (NASA TLX) (Hart & Staveland, 1988), Subjective Workload Assessment Technique (SWAT) (Reid, Potter, and Bressler, 1989), and Modified Cooper-Harper (Wierwille & Casali, 1983). SWAT is perhaps the least intrusive of the three, requiring only three digits from a choice of 1, 2 or 3 to be elicited from the participant. Due to its relatively low level of intrusiveness, SWAT was selected for the present study.

Objective Performance Measures

Reaction time and accuracy are traditionally used to assess human performance. Reaction times have been shown to be sensitive to many factors including the complexity of the stimulus and the complexity of the required response. Reactions have been categorized as simple, recognition or choice depending on the number and type of stimuli and the response required once the stimulus occurs (Welford, 1980). Noticing a change in the movement of a symbol and lifting a finger off a keyboard can be considered a recognition reaction since there is only one appropriate response to make immediately after the stimulus is recognized. However, touching a symbol that has stopped moving from among a screen full of several moving symbols can be considered a choice reaction – the stimulus and the response are more complex. It also should be noted that the modality of the stimulus and response (e.g. visual, auditory, or psychomotor) may affect the reaction time. For example, simple reaction times to auditory stimuli are faster than to visual stimuli (Welford, 1980); however the effect appears to be related to stimulus intensity (Kohfeld (1971). Several models have been produced to describe choice reaction time. Possibly the best known is the Hick-Hyman Law (Hick, 1952 and Hyman, 1953) which is based on information theory and says that response time increases linearly as the number of different stimuli increases according to log(N), where N is the number of stimuli alternatives. According to the Hick-Hyman Law, as the number of symbols to scan and monitor increases, reaction time will increase.

Humans trade off speed and accuracy when attempting to perform reaction tasks – the faster one performs; the more errors one makes. This trade off is well described

(Wickelgren, 1977). Models have been developed that describe the time and accuracy involved in motions such as reaches. Fitts' Law (Fitts, 1954) and variations thereof (e.g. Welford, 1960; Kvalseth, 1980, MacKenzie and Buxton, 1992; Drury and Hoffman, 1992) generally relate the size of the objects that must be touched (targets) and the distance between them to movement time. According to these models, it is more difficult to accurately make motions into a smaller target zone at greater speed.

Accuracy refers to how close a measurement comes to the correct value and error refers to a mismatch between the correct value and the value measured. In human performance research, errors are often counted per error type. Human error can be classified simply as errors of commission (doing the wrong thing), omission (doing nothing when a response is required), sequential error (performing tasks out of sequence), time error (performing at the wrong speed, too late or early), or extraneous act (Swain and Guttman, 1983). Wickens and Hollands (2000) have overlaid previous classifications onto an information processing model to categorize error as mistakes, slips, mode errors, or lapses. Mistakes are errors resulting from having an inappropriate intention and carrying it out. Slips are errors in execution when the intention is correct. Mode errors involve performing an action that would be appropriate in one context in a context where it is the wrong thing to do. Mistakes, slips, and mode errors are errors of commission. Lapses are errors of omission caused by forgetting to perform a task.

Research Objective

The primary purpose of this research was to determine which size symbols should be used on an operator control unit to indicate the (moving) status of unmanned ground vehicles. This determination is important because of the conflicting design motivation to reduce the size of symbols so that they cover less of the map but are large enough to read and touch particularly in high workload situations. Military standards for symbols discourage use of color, shape, or flashing to indicate a dynamic change in status. This study established the time based performance when using a symbol that conforms to these standards.

Based on the reviewed literature, the following hypotheses were tested:

- Time to recognize and respond to a change in ARV status will be less for symbols with greater height.
- As workload (number of distracter symbols) increases, recognition and response times will increase.
- There will be an interaction between workload and symbol height.

Methodology

Experimental Design

A two factor, within subjects design (3x3) was used. Participants were a random effect. Factor A, symbol size, and Factor B, number of distracters, were considered fixed effects.

Independent Variables

Factor A included three sizes of symbols to represent the ARVs. The symbols had an on-screen vertical height of 0.116 inch, 0.233 inch, and 0.400 inch, subtending 20 minutes, 40 minutes, and 69 minutes of arc when viewed at a distance of 20 inches. The symbols were sized proportionally (i.e. maintain a width to height ratio of 1:1) to achieve the different levels of vertical height. Figure 5 shows the ARV symbol that was used.

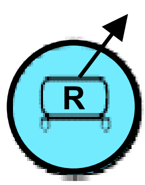

Figure 5. ARV symbol.

Factor B included three amounts of distracter symbols intended to vary mental workload. The number of distracter symbols present in addition to the 4 ARV symbols was 12, 8, and 0 for high, medium, and low workload. These levels were pilot tested to confirm their ability to manipulate subjective mental workload. The distracter symbols were increased in size to match that of the ARV symbols so that differences in saliency are not based on size differences. Figure 6 shows the distracter symbol that was used.

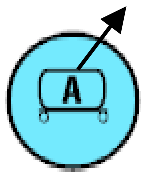

Figure 6. Distracter symbol.

Dependent Variables

Several dependent variables were measured. Two were based on reaction time. Recognition time, measured the time to react to a change in vehicle status. Recognition time was defined as the elapsed time between the ARV status change (stimulus onset) and the participant lifting an index finger off the space bar. Participants were required to rest their hands on the workstation keyboard with an index finger depressing the space bar until they were ready to touch the screen. The second reaction time variable, response time, measured the time to touch the ARV symbol on the screen. Response time was defined as the elapsed time between the participant lifting a finger off the space bar and touching the correct ARV symbol such that the touch registered. The time clock was stopped when the participant lifted their finger off the correct symbol. Participants were all bare handed and were allowed to use any part of their finger to touch the screen (e.g. tip, nail, or pad)

Both recognition time and response time were collected in case uncontrolled differences in location of the ARV symbol and participants' physical ability masked the effects of workload and symbol size. Response time may be more prone to this masking but because recognition time involves only the small motor component of lifting a finger off the space bar, it was considered less prone.

Accuracy was measured in two ways. First, errors during the recognition timing were defined as those when the participant lifted their finger off the space bar before the target ARV symbol had stopped moving. These errors of commission (usually mistakes but occasionally slips) were recorded automatically by the simulation and show up as a negative recognition time. Although participants were instructed not to lift their finger until they were absolutely sure a symbol had stopped moving, these errors were measured to confirm that participants followed instructions not to trade accuracy for speed.

Extra touches of the correct ARV symbol required to register with the simulation system were counted as a second accuracy measure. The data collector received auditory feedback via headphones when a touch was not registered correctly. Each trial continued until the participant correctly recognized and reported the change in vehicle status. Errors of commission when the participant attempted to touch the wrong symbol were noted (although, due to the accuracy instruction, these errors did not occur).

Subjective experiences of mental workload were measured using SWAT, and this measurement provided a validity check to confirm the mental workload manipulation.

Control Variables

The number, shape and color of ARV symbols on map were held constant for all trials. There were 4 symbols representing ARVs on the map display. The rate of movement of the symbols was set to 1 pixel per second. The time during which the change in status occurred was determined by calling a random function that returned a time between 3 and 18 seconds to control for temporal certainty because reaction time has been shown to be sensitive to this factor (Welford, 1980). The same LCD flat panel with touch screen overlay was used for all trials. The distance of each participant from the display was maintained at 20 inches by fixing the position of the keyboard, display and chair across all trials for each participant. A viewing distance of 20 inches was confirmed by measuring the distance between the bridge of the participant's nose and the center of the display. Trials were conducted in the same windowless room so that lighting levels were consistent. The same type of background map was used for all trials so that contrast between the map and symbols was controlled. Participants began all trials with the index finger of their dominant hand depressing a red marker on the space bar of a keyboard. The marker was positioned in line with the center of the display. Figure 7 shows the marker and keyboard.

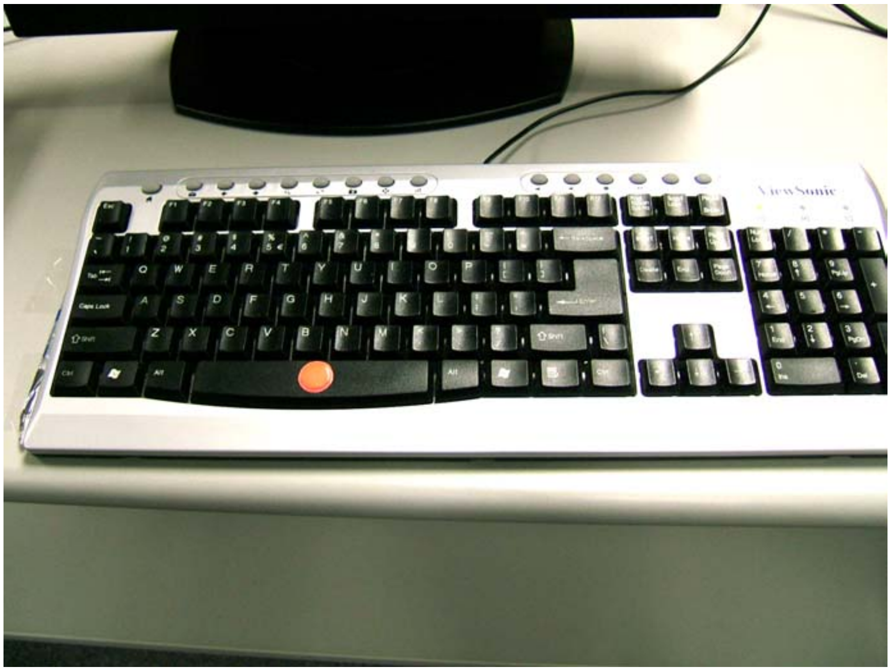

Figure 7. Starting position marker and keyboard.

Participants

The experiment was conducted using 12 civilian volunteer participants (six males and six females). To qualify as a participant, volunteers needed to be between age 20-35 with visual acuity corrected to at least 20/20 vision in one eye and 20/100 in the other. A 20/30 acuity means the person being tested can successfully identify alphanumeric characters at 20 feet that a person with normal vision can see at 30 feet. These vision requirements match those for current military occupational specialties assigned to aerial unmanned vehicles or proposed as operators of unmanned ground vehicles (US Department of the Army, 1999). Initial screening of participants for visual acuity was based on self report with follow up confirmation using a Snellen wall chart. Prior computer experience was used to screen candidate participants because computer experience cannot be assumed for potential users of the OCU. Participants were drawn from a pool of volunteers in the Boulder, CO area who use a personal computer as part of their work duties or academic study. Participation was voluntary and the participants were not compensated directly.

Instrumentation

The experiment was conducted in the usability laboratory at Micro Analysis and Design, Boulder, CO. Software to control presentation of the experimental conditions and data collection (reaction time, response time and error) was developed by Micro Analysis and Design according to specifications supplied by ARL. The software was derived from the concept OCU graphical user interface shown in figure 2. For purposes of this study, the symbols were not allowed to overlap. The panel on which the symbols were viewed was an 18.1 inch liquid crystal display (LCD) manufactured by Landmark Technology with a 3M MicroTouch touch screen overlay. The display was set to a resolution of 1280 x 1024.

Procedure

Informed Consent

The author distributed an informed consent form to each participant, reviewed the details of the form, and answered all questions (Appendix A). Participants wishing to continue with the experiment signed the informed consent form. No participants elected not to sign the informed consent form but if they did, they would have been withdrawn from the experiment. Participants who signed the consent form continued to the SWAT card sort.

Familiarization

Each participant was given an orientation briefing on the experimental apparatus and its function. Then they were given 5 familiarization trials to learn the experimental procedure. The number of familiarization trials was selected based on similar procedure in the reaction time literature (Welford, 1980) and to avoid fatigue and learning effects. The size of the symbols and number of distracters used in familiarization was different from the experimental levels to avoid learning effects. A symbol size, 0.3 inches, intermediary between the medium and largest symbol size was used. The number of distracter symbols was 2, intermediary between the low and medium workload levels. Participants also practiced entering a 3 digit SWAT score after the 5 familiarization trials.

SWAT Card Sort

Each participant sorted a deck of SWAT cards to establish a subjective workload scale for their ratings. This involved each participant ranking, from lowest to highest, 27 combinations of three levels of three workload subscales. Each combination was represented on one SWAT card. Participants were required to read SWAT card sorting instructions before they began sorting. The instructions are included as Appendix B. The participant was allowed a 15 minute break following the card sort to control for fatigue effects during the touch screen phase.

Treatment Presentation Order

The number of the ARV symbol in which the change in status occurred was not randomized because the symbols appeared in different locations on the map for each trial and opportunities for anticipating which ARV had a change in status are minimal. The time during which the change in status occurred was determined by calling a random function that returned a time between 3 and 18 seconds. Varying the time at which the change in status occurred minimized opportunities for correctly anticipating when to lift off the space bar.

The order of treatment presentation is critical in mitigating the potential order effects of workload treatment. The order of presentation for the three levels of symbol height was partially counterbalanced within each workload level using a partial Latin square arrangement as shown in Table 1 (Keppel, 1991).

During the experiment, each participant experienced the number of distracters for one level of workload and completed all three levels of symbol height before changing workload level. Participants completed 4 observations per treatment (symbol height by workload level combination). The number of observations was determined based on the reaction time literature (Welford, 1980) and to avoid fatigue effects. The participant initiated each repetition by pressing the marker on the space bar. SWAT scores were collected only once per treatment. Participants entered a 3 digit (1, 2, or 3 for each digit) SWAT score using a mouse and graphical user interface after each treatment (not after each repetition).

Table 1. Experimental design structure

S#	B1 = low workload			B2 = medium workload			B3 = high workload		
1	A1	A2	A3	A1	A3	A2	A2	A1	A3
2	A2	A3	A1	A3	A1	A2	A3	A2	A1
	B1 = low workload			**B3 = high workload**			**B2 = medium workload**		
3	A1	A2	A3	A1	A3	A2	A2	A1	A3
4	A2	A3	A1	A3	A1	A2	A3	A2	A1
	B2 = medium workload			**B1 = low workload**			**B3 = high workload**		
5	A1	A2	A3	A1	A3	A2	A2	A1	A3
6	A2	A3	A1	A3	A1	A2	A3	A2	A1
	B2 = medium workload			**B3 = high workload**			**B1 = low workload**		
7	A1	A2	A3	A1	A3	A2	A2	A1	A3
8	A2	A3	A1	A3	A1	A2	A3	A2	A1
	B3 = high workload			**B1 = low workload**			**B2 = medium workload**		
9	A1	A2	A3	A1	A3	A2	A2	A1	A3
10	A2	A3	A1	A3	A1	A2	A3	A2	A1
	B3 = high workload			**B2 = medium workload**			**B1 = low workload**		
11	A1	A2	A3	A1	A3	A2	A2	A1	A3
12	A2	A3	A1	A3	A1	A2	A3	A2	A1

A1 = .116 in. symbol height — B1= 0 distracter symbols

A2 = .223 in. symbol height — B2 = 8 distracter symbols

A3 = .400 in. symbol height — B3 = 12 distracter symbols

Results

Data Analysis

All 12 participants completed all of the trials. Inferential statistics performed using SAS/STAT® software were used to determine if statistically significant differences exist for the Factor A main effect, Factor B main effect, and the interaction effect. A mixed model analysis was conducted with a significance level of $p \leq 0.05$ for each of the dependent variables. Factors A and B were considered fixed effects. The workload blocks (factor B period) and the nested symbol size blocks (factor A period) were included in the analysis. Table 2 shows the general format of the summary table for each of the dependent variables including the appropriate degrees of freedom and equations for the computed F-values.

Table 2. Test of Fixed Effects Table

Within-Subject Design				
SOURCE	**DF**	**SS**	**MS**	**F**
Subject (S)	11	SS_S		
Workload (W)	2	SS_W	MS_W	MS_W / MS_{SxWxBP}
Factor B Period (BP)	2	SS_{BP}	MS_{BP}	
Error S x W x BP	20	SS_{SxWxBP}	MS_{SxWxBP}	
Symbol Height (H)	2	SS_H	MS_H	MS_H / MS_{HxWxS}
H x W	4	SS_{HxW}	MS_{HxW}	MS_{HxW} / MS_{HxWxS}
Factor A Period (AP)	2	SS_{AP}	MS_{AP}	
Error H x W x S	64	SS_{HxWxS}	MS_{HxWxS}	
Total	107	SS_{Total}		

If the analysis of variance revealed significant differences for either of the main effects or the interaction effect, a post-hoc analysis using least squares means (LSM) was conducted to isolate which treatment combinations produced the indication. Results of the analyses are detailed in the following sections.

Recognition time

The effect of symbol size on recognition time was significant ($\underline{F}_{2,64}$=4.99, $\underline{p}$=0.0097). As symbol size increased, recognition time decreased. The post hoc test indicated the smallest symbol size was significantly different from the medium and from the largest size, but the difference between the medium and largest was not significant. No main effect was found for workload (number of distracters) although the probability was close to criterion ($\underline{F}_{2,20}$=3.19, $\underline{p}$=0.0626). The post hoc test indicated a significant difference between low and high workload levels. There was no significant interaction between symbol size and workload. Figure 8 shows the mean and standard error for recognition time as a function of symbol size at each level of workload. Symbol size is shown categorically because it was treated as a fixed effect.

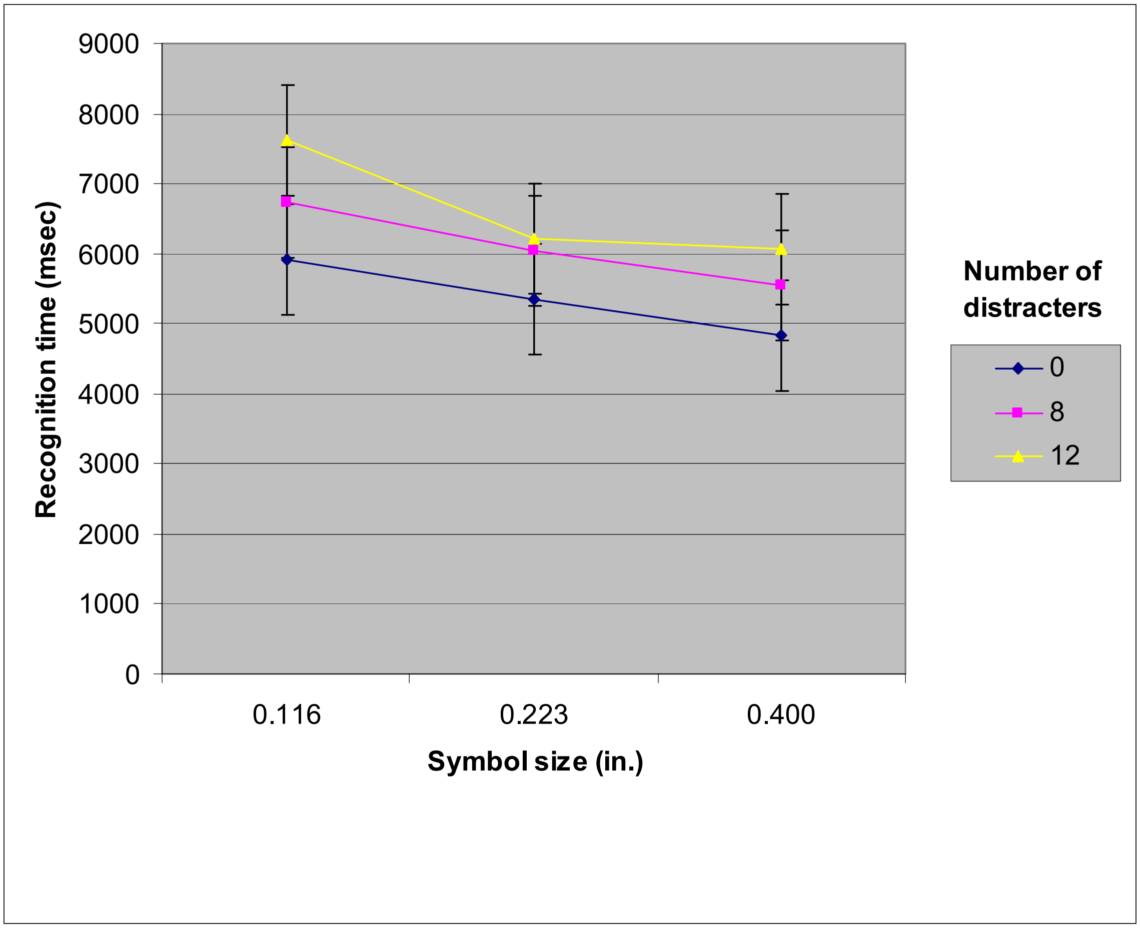

Figure 8. Mean and standard error for recognition time as a function of symbol size at each level of workload.

Response time

The effect of symbol size on response was significant ($\underline{F}_{2,64}$=41.52, $\underline{p}$=<0.0001). As symbol size increased, response time decreased. The post hoc test indicated the smallest symbol size was significantly different from the medium and from the largest size, but the difference between the medium and largest was not significant. No main effect was found for workload (number of distracters). There was no significant interaction between symbol size and workload. The order of presentation of symbol size was statistically significant ($\underline{F}_{2,64}$ $\underline{p}$=0.0414) indicating that the partial counter balancing scheme to prevent order effects may have been ineffective. Figure 9 shows the mean and standard error for response time as a function of symbol size at each level of workload. Symbol size is shown categorically because it was treated as a fixed effect.

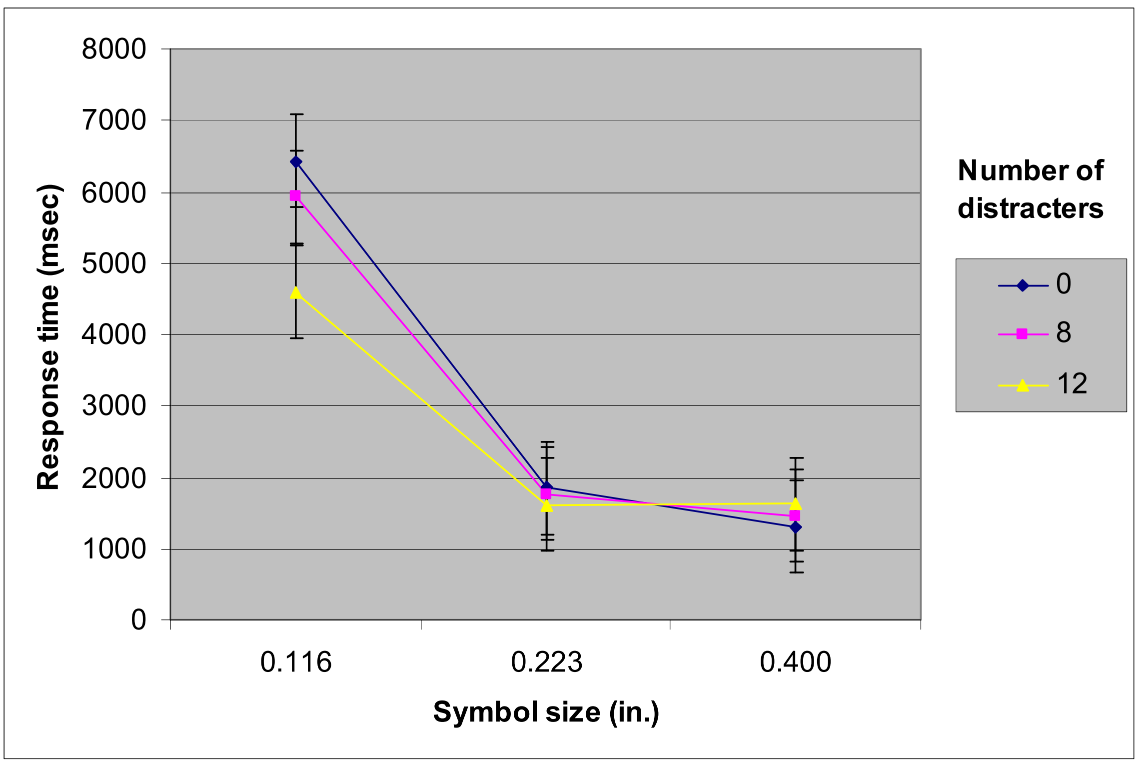

Figure 9. Mean and standard error for response time as a function of symbol size at each level of workload.

Errors committed reacting too early

Out of 432 trials (12 participants x 9 conditions x 4 repetitions) there were only 8 instances when a participant lifted their finger off the space bar before one of the ARV symbols had stopped moving. There was no observable pattern to these errors, no significant effects were indicated, and the 8 instances were deleted from the data for analysis of recognition and response times.

Errors committed attempting to touch the correct symbol

Analysis of the extra touches required for the system to register that the correct symbol was touched indicates a main effect for symbol size ($\underline{F}_{2,64} = 62.51$, $\underline{p}<0.0001$). As symbol size increased, errors decreased. The post hoc test indicated that the number of extra touches was statistically different between the smallest symbol size and the medium and between the smallest and largest. The difference between the medium and largest symbol size was not significant. There was no significant effect for workload or the interaction between symbol size and workload. Figure 10 shows the mean and standard error for number of extra touches as a function of symbol size at each level of workload. Symbol size is shown categorically because it was treated as a fixed effect.

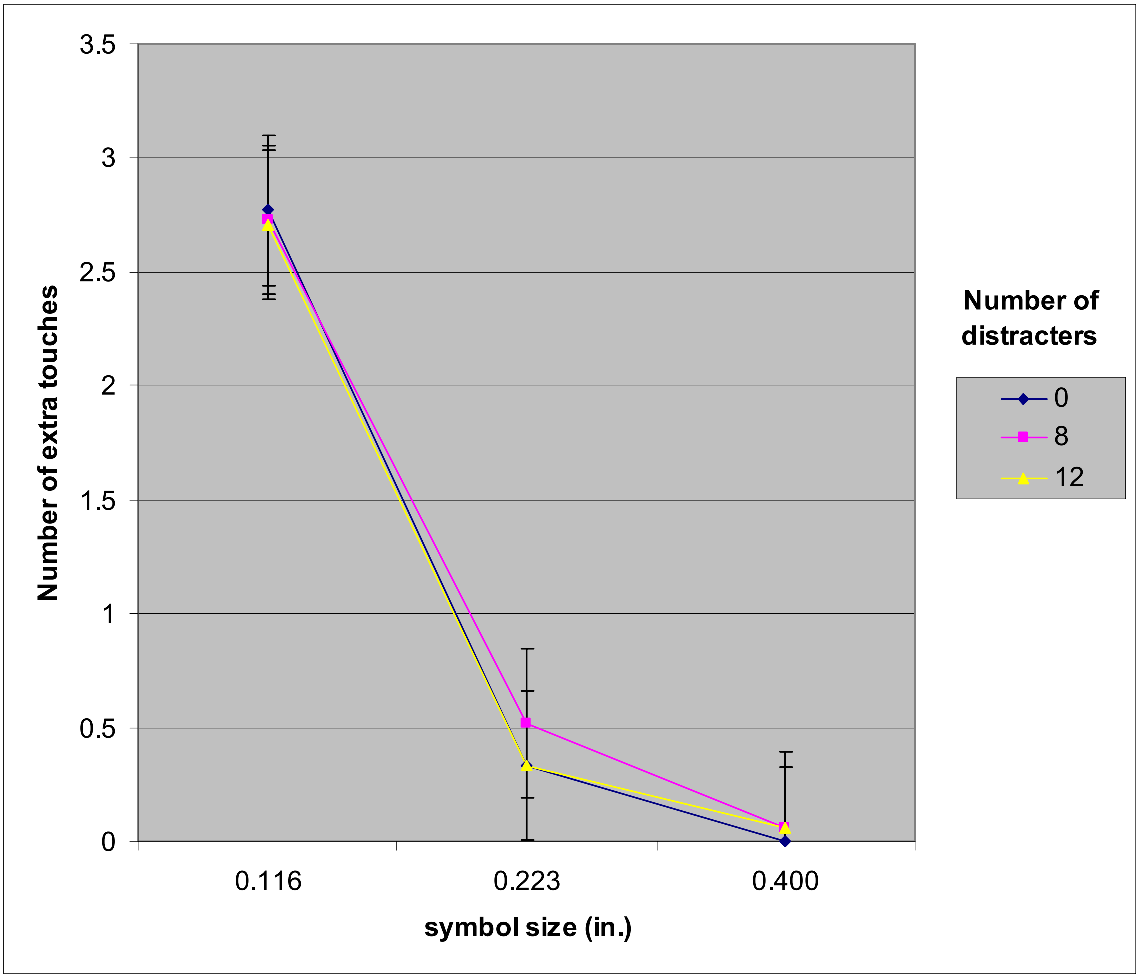

Figure 10. Mean and standard error for number of extra touches as a function of symbol size at each level of workload.

SWAT

The SWAT analysis program (Reid, Potter, and Bressler, 1987) was used to analyze all of the card sorts for the 12 participants and then to scale their scores. Using the SWAT analysis program, a Kendall's coefficient of concordance ($\underline{W}$ = 0.84, $\underline{p}$<0.01) demonstrated significant agreement among the 12 participants with respect to their rank ordering of the 27 of the SWAT cards (i.e. combinations subscale levels). As a result and in accordance with the SWAT, the card sort data from all participants was combined during the SWAT scale development procedure to form one overall interval scale for the group of participants. The analysis of variance was performed on the scaled scores. The main effect for symbol size was significant ($F_{2,64} = 39.59$, $p<0.0001$). As symbol size increased, overall SWAT scores decreased. Pairwise comparisons indicate that all levels of symbol size were significantly different from each other. The main effect for number of distracters was significant ($F_{2,20} = 10.28$, $p=0.0008$). As the number of distracters increased, overall SWAT scores decreased. The post hoc test indicated that the difference between low (0 distracters) and medium workload (8 distracters) was statistically different and the difference between low (0 distracters) and high workload (12 distracters) was also significant. The difference between 8 and 12 distracters is not significant. There was no significant interaction between symbol size and workload. Figure 11 shows the mean and standard error for SWAT score as a function of symbol size at each level of workload. Symbol size is shown categorically because it was treated as a fixed effect.

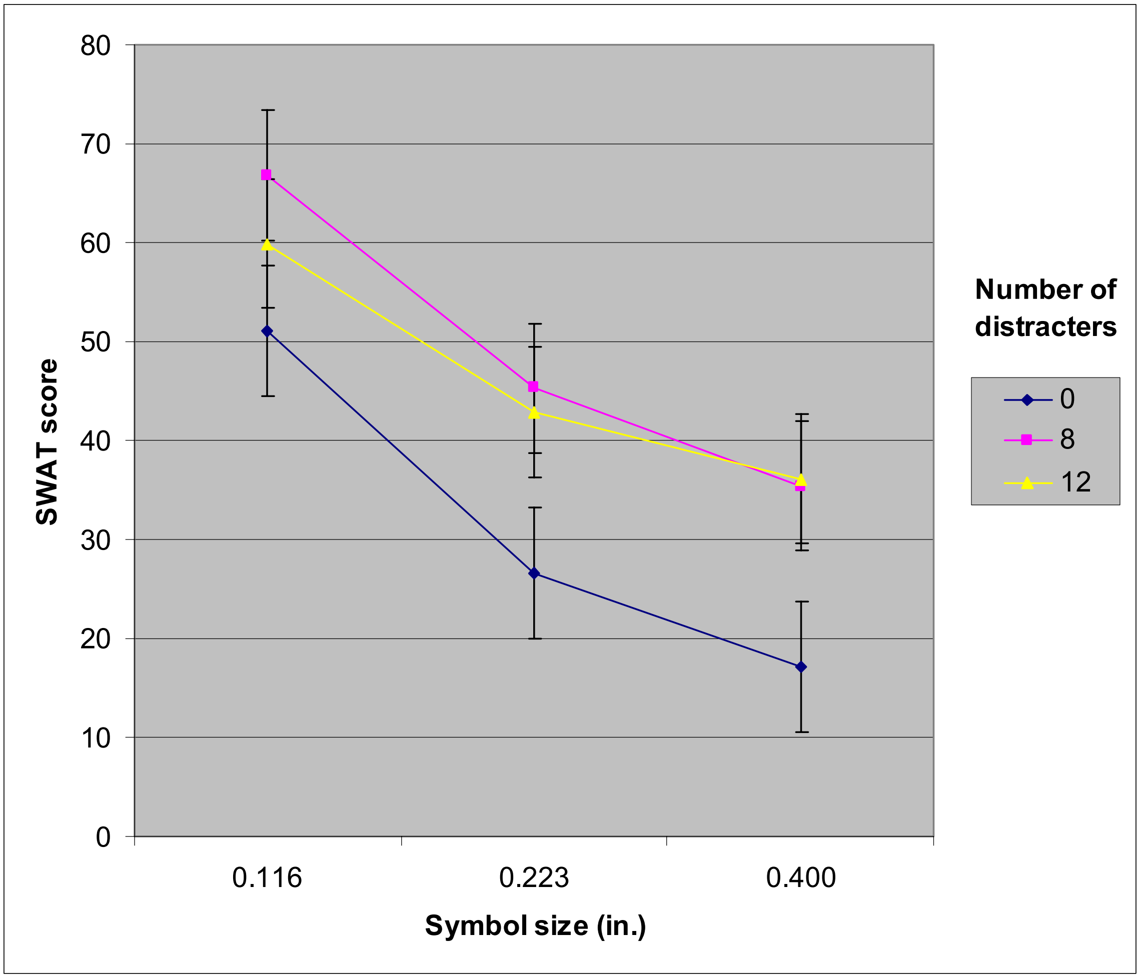

Figure 11. Mean and standard error for scaled SWAT score as a function of symbol size at each workload level.

Although analysis of the subjective workload subscales against performance is often reported when using other workload measures such as NASA TLX (Hart and Staveland, 1988), such analysis has been cautioned against for SWAT. Boyd (1983) reported that the SWAT subscales may not be orthogonal – ratings across the 3 SWAT subscales were not independent when the dimensions were varied independently within a task. Likewise, Colle and Reid (2005) in their discussion of research relevant to a mental workload redline or cutoff mention that the time demand and other SWAT subscales are not independent either as a consequence of participant response strategy or because the demand felt by the participant on the subscales is not orthogonal. Given this prior research, the SWAT subscales were not analyzed separately for the present study.

Discussion

Given the scientific literature on reaction time, results for the objective performance measures (recognition time, response time, and extra touches) were as expected for symbol size. As the symbol size increased; recognition time, response time, and extra touches decreased. Significant differences were seen in these measures between the symbols subtending 20 and 40 minutes of arc and between symbols subtending 20 and 69 minutes of arc. Relatively small changes in symbol size (20 and 40 minutes subtended arc) resulted in mean response times that were roughly 3 times greater for the smaller size. The number of extra touches required was 5 to 8 times greater depending on the number of distracters. Also, as expected, subjective workload increased as symbol size decreased with differences seen between all symbol size levels. The perceived workload was between 1.4 to 1.9 times greater for the smallest to medium sized symbols depending on the number of distracters. It is not unexpected that a significant difference was not seen between the medium and largest symbol sizes for the objective measures but was observed for subjective mental workload. The workload caused by the differences in these symbol sizes may not have been significant enough to result in a performance differences but subjective mental workload measures are more sensitive to these differences (Hart and Wickens, 1990; Yeh and Wickens, 1988; and Hockey, 1997). These differences in perceived workload are nonetheless important since the toll of coping with prolonged elevated mental workload may lead to fatigue and eventually error (Hockey, 1997).

Results for objective performance measures were not as expected for the mental workload manipulation. It was expected that as the number of distracter symbols increased, recognition time, response time, and extra touches would have increased. No such significant differences were observed although the main effect for recognition time was close to criterion. This measure may have been less masked by differences in the reach strategies of the participants across trials. The subjective mental workload measure did show a significant difference as a result of the mental workload manipulation. Again, it is possible that this difference in mental workload was not severe enough to cause a significant performance difference. Under more severe or prolonged mental workload conditions, the difference may have a greater effect on performance. It is interesting to note that participants commented that the symbol size itself had a greater effect on their impression of mental workload than did the number of distracters. The implication from these results is that the size of touch screen symbols may be an important source of mental workload.

Because both factors involved visual modality, it was expected that there would be an interaction between workload and symbol size for each of the dependent measures but this was not observed. This is likely due in part to the lack of a significant effect for the mental workload manipulation as an independent variable.

Reid and Colle (1998) have reported that across a range of studies, SWAT scores of about 40 may indicate a critical level of mental workload or overload. The mean SWAT scores for the present study were close to 40 for the medium sized symbol and close to 60 for the smallest sized symbol. These scores seem particularly high considering the participants were performing a relatively simple search and pointing task. The high scores may be explained by a failure to expose participants to the full range of the experimental conditions ahead of time. Participants practiced with a symbol size between the medium and large size and with the number of distracters between the low and medium workload levels. Reid and Colle (1998) and Colle and Reid (1998) recommend providing participants experience with the complete difficulty range during practice trials to control for context effects. In hindsight, the lack of context may have inflated the SWAT scores with participants trying to use the full range of the scale. However, the lack of context should not have affected the relative comparison of subjective workload rating.

Participants reported that it would be easier to scan the screen if they were seated further away. Attempting to scan a screen at close range is analogous to scanning a larger area – greater eye movement is required to cover the area to be scanned. Participants also commented that the largest symbols were too big to determine easily if the symbol was moving. It cannot be determined from this study if this was due to the absolute size of the symbols, the distance they moved relative to their size (note that all symbols moved the same distance regardless of size), or their size relative to the screen.

Conclusions

Summary

Results suggest that symbols smaller than those recommended for keypads may be sufficient for interactive map displays. For static (non-vibrating) platforms with barehanded operator controllers, MIL-STD-2525B conforming symbols that subtend 40 minutes of arc at a distance of 20 inches from the operator control unit touch screen may be sufficiently large to ensure adequate performance under low to moderate mental workload conditions. Performance for larger symbols was not significantly different. Results also suggest that the size of touch screen symbols may be a source of perceived mental workload and that small changes in symbol size may multiply response time and number of errors.

These results may have implications beyond design of military systems. Visual displays operated via touch screen are increasingly common in consumer products and non-military remotely controlled vehicles such as those used in undersea operations, mining, space exploration, and explosive demolition. Touchscreen displays particularly those operated while performing other tasks (e.g. using an in-vehicle navigation device while driving an automobile) may present significant sources of workload depending on the size and number of symbols on the display. Results of this study indicate that the extra time and touches required to select symbols may be several times greater for smaller symbols. This may result in more time attending to the display instead of the driving task. Time pressure may not be as critical when controlling non-military remotely controlled vehicles however the increased mental workload imposed by small and numerous symbols may contribute to fatigue and eventually increased errors during critical operations.

Limitations of Research

Several variables were controlled in this study and attempts to extrapolate results beyond those controls may not be valid. Only one type of touch screen and resolution were tested. Due to resource limitations, only three symbol heights and three numbers of distracters were used. Nothing can be concluded for sizes between those tested. The distracter symbol was the same for all conditions. Symbols with different characteristics such as contrast, color, and shape may produce different results. The background map and mapped area were the same for all conditions. Other maps may vary in the contrast between map and symbol and density of terrain features. These differences may change the difficulty in distinguishing among symbols, tracking them or recognizing changes in their status. Only four ARV symbols were used. While this matches the proposed number of UGVs that a Soldier may be asked to control, different numbers of symbols to be tracked and touched may have produced difference results.

These results apply to static, non-moving platforms (i.e. the participant and the touch screen were not in a vibrating environment). If the workspace for the human operator and touch screen were moving, it may not be as easy to read fine detail of the symbols (Lewis and Griffin, 1980) or touch them as quickly or accurately (McDowell, Rider, Truong, and Paul, 2005) due to vibration of the display or experienced at parts of the operator's body such as the head, eyes, arm, and hand.

Application of touch screens and graphical symbols has become ubiquitous for personal data assistants (PDAs) and other hand held devices. These devices are not always located directly in front of and centered on the user when operated. For example, these devices may be mounted in the center of an automobile console or on the leg of a Soldier. Viewing the symbols at an angle other than perpendicular to the user's line of sight may reduce the perceived width of the symbol. Attempting to accurately touch symbols at angles other than those studied in the present experiment may be more difficult to due the complexity of the motion or parallax effects. PDA screens are much smaller than the 18.1 inch OCU screen used in the present study and the symbols studied will take up more space relative to the PDA screen. Given the informal participant feedback about larger symbol sizes discussed earlier, results for smaller screens may differ.

Future Research

Follow on studies investigating OCU symbols for touch screens should consider factors such as workspace and operator vibration, gloved touches, and more complex operator control unit tasks. Despite benefits to training and comprehension of standardizing symbols, it can be argued that the OCU symbols do not need to follow standards strictly. Symbols better suited to controlling ARVs could be used for the OCU but when the location of the ARVs is sent to other displays on other platforms, the symbols could be translated into the MIL-STD-2525B set. If this translation concept is accepted, then symbology that increases the saliency of the entity state should be explored. These concepts include use of flashing symbols, redundant shape coding using frame shapes such as those suggested by Smallman et al (2000, 2001), and “zoomable” symbols that increase in size as the user’s finger approaches them (Albinsson and Zhai, 2003).

Because differences were found between symbols subtending 20 and 40 arc minutes, symbols heights in this range should be investigated further. A mix of distracter symbols with systematically varied graphical parameters such as width to height ratio, contrast, color, and luminance should be investigated. Because symbols viewed off angle (i.e. not perpendicular to the user’s line of sight) may appear smaller, viewing angle should also be manipulated in future studies.

Follow on experiments should be performed with participants exposed to the full range of the experimental conditions ahead of time to improve benchmarking subjective mental workload levels. Another measure such as NASA TLX should be used in follow on research to confirm the subjective mental workload and determine the demand type imposed by changes in touch screen symbol size.

References

Achille, L.B., Schmidt-Nielsen, A., & Sibert, L.E. (1992). Dual-task performance as a function of presentation mode and individual differences in verbal and spatial ability (NRL report 9372). Washington, DC: Naval Research Laboratory.

Albinsson, P. and Zhai, S. (2003). High precision touch screen interaction. In Proceedings of the conference on human factors in computing systems. On line at http://www.almaden.ibm.com/u/zhai/papers/TouchScreen.pdf.

Andrews, M. A., Schmidt, D. R., & Killion, T. (September – October 2001). Enabling the objective force. In Army acquisition, logistics & technology (pp. 9-11). Fort Belvoir, VA: Department of the Army.

Badler, N., Phillips, C., & Webber B. (1993). Simulating humans, computer graphics animation control. Oxford University Press, New York

Beaton, R. J. and Weiman, N. (1984). Effects of touch key size and separation on menu-selection accuracy (Tech. Report HFL-500-01). Beaverton, OR: Tektronix Laboratory, Human Factors Research Department.

Blankenburger, S. & Hahn, K. (1991). Effects of icon design on human-computer interaction. International Journal of Man-Machine Studies, 35, 363-377.

Boyd, S. (1983). Assessing the validity of SWAT as a workload measurement instrument. In Proceedings of the human factors society 27th annual meeting. (pp. 124-128). Santa Monica, CA: Human Factors Society.

Cain, B & Hendy, K. C. (1998). A protocol for validating task network simulation procedures. TTCP technical report number TTCP/HUM/98/0004. Washington, DC: The Technical Cooperation Program.

Chiles, W. D. & Alluisi, E. A. (1979). On the specification of operator and occupational workload with performance measurement methods. Human Factors, 21(5), 515-528.

Colle, H.A. & Hiszem, K.J. (2004). Standing at a kiosk: effects of key size and spacing on touch screen numeric keypad performance and user preference. Ergonomics, 47(13), 1406 – 1423.

Colle, H. A. and Reid, G. B. (1998). Context effects in subjective mental workload ratings. Human Factors, 40, 591-600.

Colle, H.A. and Reid, G.B. (2005). Estimating a mental workload redline in a simulated air-to-ground combat mission. International Journal of Aviation Psychology, 15(4), 303-319.

Dahn, D. & Gacy, M. (2002). Human machine interface design for control of robotic systems. In Proceedings of the 23rd army science conference. Online at www.asc2002.com

Damos, D. L. (Ed.) (1991). Multiple task performance. Washington, DC: Taylor & Francis.

Drury, C. G. & Hoffmann, E. R. (1992). A model for movement time on data-entry keyboards. Ergonomics, 35, 129-147.

Dudenhoeffer, D.D., Bruemmer, D.J., & Davis, M.L. (2001). Modeling and simulation for exploring human-robot team interaction requirements. In B.A. Peters, J.S. Smith, D.J. Medeiros, and M.W. Rohrer, (Eds.), Proceedings of the 2001 winter simulation conference.

Easterby, R., & Zwaga, H. (Eds.). (1978). Developing effective symbols for public information. In Information design: the design and evaluation of signs and printed material (pp. 277-297). New York: John Wiley & Sons.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47, 381-391.

Gawron, V. J. (2000). Human performance measures handbook. Mahway, NJ: Lawrence Erlbaum.

Gittins, D. (1986). Icon-based human-computer interaction. International Journal of Man-machine Studies, 24, 519-543.

Griffin, T. (January-February 2004). Unmanned ground vehicles. In Army acquisition, logistics & technology: Future combat systems (pp. 42-43). Fort Belvoir, VA: Department of the Army.

Gordon, C., Bradtmiller, B., Churchhill, T., Clauser, C., McConville, J., Tebbetts, I., & Walker, R. (1989). 1988 anthropometry survey of U.S. Army personnel: Methods and summary statistics (Technical Report Natick/TR-89/044). Natick, MA: U.S. Army Natick Research, Development and Engineering Center.

Hart, S. G. (1989, April). Crew workload management strategies: a critical factor in system performance. Paper presented at the Fifth International Symposium on Aviation Psychology, Columbus, OH.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (task load index) results of empirical and theoretical research in P. A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139-183). Amsterdam: North-Holland.

Hart, S. G. & Wickens, C. D. (1990). Workload assessment and prediction. In H. R. Booher (Ed.), MANPRINT: An approach to systems integration (pp. 257-296). New York: Van Nostrand Reinhold.

Hick, W. E. (1952). On the rate of gain of information. Quarterly Journal of Experimental Psychology, 4, 11-26.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload. Biological Psychology, 45, 73-93.

Human Factors and Ergonomics Society. (2002). Human factors engineering of computer workstations (Draft standard BSR/HFES 100). Santa Monica, CA: Author.

Horton, W.K. (1994). The icon book: visual symbols for computer systems and documentation. New York: J. Wiley.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. Journal of Experimental Psychology, 68, 612-614.

Keppel, G. (1991). Design and analysis: a researcher's handbook, (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

Knowles, W. B., Garvey, W. D., & Newlin, E. P. (1953). The effect of speed and load on display control relationships. Journal of Experimental Psychology, 46, 65-75.

Kohfeld, D. L. 1971. Simple reaction time as a function of stimulus intensity in decibels of light and sound. Journal of Experimental Psychology, 88, 251-257.

Kramer, A. F. (1991). Physiological metrics of mental workload: a review of recent progress. In D. L. Damos (Ed.). Multiple task performance (pp. 279-328). Washington, DC: Taylor & Francis.

Kvalseth, T. O. (1980). An alternative to Fitts' law. Bulletin of the Psychonomic Society, 16, 371-373.

Lewis, C.H. & Griffin, M.J. (1980). Predicting the effects of vertical vibration frequency, combinations of frequencies and viewing distance on the reading of numeric displays. Journal of Sound and Vibration, 70, 355-377.

McDowell, K., Rider, K., Truong, N., & Paul, V. (2005). Effects of ride motion on reaction times for reaching tasks. In Proceedings of the SAE world congress. Warrendale, PA: SAE International.

Reid, G.B. & Colle, H.A. (1988). Critical SWAT values for predicting operator workload. In Proceedings of the human factors society 32nd annual meeting. (pp 1414-1418). Santa Monica, CA.

Reid, G.B., Potter, S.S., & Bressler, J.R. (1987). User's guide for the Subjective Workload Assessment Technique (SWAT) (AAMRL-TR-87-023). Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Sanders, M. & McCormick, E. (1993). Human factors in engineering and design. New York: McGraw-Hill.

Schipani, S. (2002). An evaluation of operator workload during partially autonomous vehicle operations. In Proceedings of the PerMIS conference. Gaithersburg, MD: NIST.

Smallman, H.S.; Oonk, H.M.; St. John, M.; & Cowen, M.B. (2000). Track recognition using two-dimensional symbols or three-dimensional realistic icons (Technical Report 1818, March). San Diego, CA: SPAWAR.

Smallman, H.S.; Oonk, H.M.; St. John, M.; & Cowen, M.B. (2001). Searching for tracks imaged as symbols or realistic icons: A comparison between two-dimensional and three-dimensional displays (Technical Report 1854, April). San Diego, CA: SPAWAR.

Swain, A.D. & Guttman, H.E. (1983). Handbook of human reliability analysis with emphasis on nuclear power plant applications (NUREG CR-1278). Washington, DC: Nuclear Regulatory Commission.

UGS Corp (2005). Classic Jack [Computer software]. http://www.ugs.com/products/tecnomatix/human_performance/jack/classic_jack.shtml.

US Army Research Laboratory. Robotics alliance overview. Retrieved on line http://www.arl.army.mil/alliances/roboview.htm 4 June 2003.

US Department of the Army. (14 April 2003). Operational requirements document (ORD) for the future combat systems (FCS), change 3. Appendix G. unit of action (UA) future combat system (FCS) operational mode summary/mission profile (OMS/MP) 29 March 2003.

US Department of the Army. (31 March 1999). Military occupational classification and structure. Pamphlet 611-21.

US Department of Defense. (January 1999). Military Standard 2525B: Common Warfighting Symbology. Defense Information Systems Agency, Reston, VA retrieved on line at http://symbology.disa.mil/symbol/history.html.

US Department of Defense. (August 1999). Military Standard 1472F: Human Engineering Design Criteria for Military Systems. Defense Information Systems Agency, Reston, VA retrieved on line at http://hfetag.dtic.mil/docs-hfs/mil-std-1472f.pdf

Wickens, C.D. & Hollands, J.G. (2000). Engineering psychology and human performance. (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years progress. Ergonomics, 3, 189-230.

Welford, A.T. (Ed.). (1980). Reaction times. New York: Academic Press.

Wickelgren, W.A. (1977). Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica, 41, 67-85.

Wierwille, W. W. & Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. In Proceedings of the 27th annual meeting of the human factors society. (pp. 129-133). Santa Monica, CA: Human Factors Society.

Wogalter, M. S., DeJoy, D. M., & Laughery, K. R. (Eds.). (1999). Warning and risk communication. Philadelphia, PA: Taylor & Francis.

Yeh & Wickens. (1988). Dissociation of performance and subjective measures of workload. Human Factors, 30, 111-120.

Appendix A. Volunteer Agreement Affidavit

VOLUNTEER AGREEMENT AFFIDAVIT:
ARL-HRED Local Adaptation of DA Form 5303-R. For use of this form, see AR 70-25 or AR 40-38

The proponent for this research is:	**U.S. Army Research Laboratory** **Human Research and Engineering Directorate** **Aberdeen Proving Ground, MD 21005**

Authority:	Privacy Act of 1974, 10 U.S.C. 3013, [Subject to the authority, direction, and control of the Secretary of Defense and subject to the provisions of chapter 6 of this title, the Secretary of the Army is responsible for, and has the authority necessary to conduct, all affairs of the Department of the Army, including the following functions: (4) Equipping (including research and development), 44 USC 3101 [The head of each Federal agency shall make and preserve records containing adequate and proper documentation of the organization, functions, policies, decisions, procedures, and essential transactions of the agency and designed to furnish the information necessary to protect the legal and financial rights of the Government and of persons directly affected by the agency's activities]
Principal purpose:	To document voluntary participation in the Research program.
Routine Uses:	The SSN and home address will be used for identification and locating purposes. Information derived from the project will be used for documentation, adjudication of claims, and mandatory reporting of medical conditions as required by law. Information may be furnished to Federal, State, and local agencies.
Disclosure:	The furnishing of your SSN and home address is mandatory and necessary to provide identification and to contact you if future information indicates that your health may be adversely affected. Failure to provide the information may preclude your voluntary participation in this data collection.

Part A • Volunteer agreement affidavit for subjects in approved Department of Army research projects
Note: Volunteers are authorized medical care for any injury or disease that is the direct result of participating in this project (under the provisions of AR 40-38 and AR 70-25).

Title of Research Project:	The Effects of Symbol Size and Workload Level on Status Awareness of Unmanned Ground Vehicles	
Human Use Protocol Log Number:	ARL-20098-04038	
Principal Investigator(s):	Mr. John F. Lockett, III	Phone: 410-278-5875 E-Mail: jlockett@arl.army.mil
Associate Investigator(s)	none	Phone: E-Mail:
Location of Research:	Micro Analysis and Design, Boulder, CO	
Dates of Participation:	5 test days between 3-9 April 2005	

Part B • To be completed by the Principal Investigator

Note: Instruction for elements of the informed consent provided as detailed explanation in accordance with Appendix C, AR 40-38 or AR 70-25.

Purpose of the Research

The primary purpose of this research is to determine which size symbols should be used on an operator control unit to indicate the status of unmanned ground vehicles. This determination is important because of the conflicting design motivation to reduce the size of symbols so that they cover less of the map but are large enough to read and touch particularly in high workload situations.

Procedures

You are being asked to participate in a laboratory experiment. Your visual acuity will be confirmed by asking you to read from an eye chart. You will be asked to sort a group of cards according to the relative importance of three different workload factors. You will then be asked to view a computer display with different sized and numbers of symbols and use a touch screen device to indicate changes in the symbols. After several trials, you will be asked for your impression of workload according to the cards that you sorted. Your participation in this study will last approximately 2 hours.

Benefits

Your participation in this study will help researchers determine the size of symbols for unmanned ground vehicles that will appear on operator control units.

Risks

Risks associated with participation in this study are expected to be no more than those associated with using a personal computer equipped with a touch screen for 2 hours.

Confidentiality

All data and information obtained about you will be considered privileged and held in confidence. All data will be recorded using a volunteer identifier code and a separate file with your consent form and the Principal Investigator will keep your assigned volunteer identifier code in a locked cabinet. Complete confidentiality cannot be promised, particularly if you are a military service member, because information bearing on your health may be required to be reported to appropriate medical or command authorities. In addition, applicable regulations note the possibility that the U.S. Army Medical Research and Materiel Command (MRMC-RCQ) officials may inspect the records.

Compensation

None

Disposition of Volunteer Agreement Affidavit

The Principal Investigator will retain the original signed Volunteer Agreement Affidavit and forward a photocopy of it to the Chair of the Human Use Committee after the data collection. The test administrator will provide a copy to the volunteer.

Consent

I do hereby volunteer to participate in the research project described in the table above. I have full capacity to consent and have attained my 18th birthday. The implications of my voluntary participation, duration, and purpose of the research project, the methods and means by which it is to be conducted, and the inconveniences and hazards that may reasonably be expected have been explained to me. I have been given an opportunity to ask questions concerning this research project. Any such questions were answered to my full and complete satisfaction. Should any further questions arise concerning my rights or project related injury, I may contact the **ARL-HRED Human Use Committee Chairperson at Aberdeen Proving Ground, Maryland, USA by telephone at 410-278-5800 or DSN 298-5800.** I understand that any published data will not reveal my identity. If I choose not to participate, or later wish to withdraw from any portion of it, I may do so without penalty. I may at any time during the course of the project revoke my consent and withdraw without penalty or loss of benefits. However, I may be requested to undergo certain examinations if, in the opinion of an attending physician, such examinations are necessary for my health and well being.

Your signature below indicates that you: (1) are at least 18 years of age, (2) have read the information on this form, (3) have been given the opportunity to ask questions and they have been answered to your satisfaction, and (4) have decided to participate based on the information provided on this form.

<table>
<tr><td colspan="2">Printed Name Of Volunteer (First, MI., Last)</td></tr>
<tr><td>Social Security Number (SSN)</td><td rowspan="2">Permanent Address Of Volunteer</td></tr>
<tr><td>Date Of Birth
(Month, Day, Year)</td></tr>
<tr><td>Today's Date
(Month, Day, Year)</td><td>Signature Of Volunteer</td></tr>
<tr><td colspan="2">Signature Of Administrator</td></tr>
</table>

Contacts for Additional Assistance

If you have questions concerning your rights on research-related injury, or if you have any complaints about your treatment while participating in this research, you can contact:

Chair, Human Use Committee U.S. Army Research Laboratory Human Research and Engineering Directorate Aberdeen Proving Ground, MD 21005 (410) 278-0612 or (DSN) 298-0612	**OR**	Office of the Chief Counsel U.S. Army Research Laboratory 2800 Powder Mill Road Adelphi, MD 20783-1197 (301) 394-1070 or (DSN) 290-1070

Appendix B. SWAT Card Sort Instructions

SWAT CARD SORT INSTRUCTIONS FOR PARTICIPANTS

During this experiment, you will be asked to quantify the mental workload required to complete the tasks. Mental workload refers to how hard you work to accomplish a task or group of tasks. The workload imposed on you at any one time consists of a combination of three dimensions. The Subjective Workload Assessment Technique (SWAT) defines these dimensions as (1) time load, (2) mental effort load, and (3) psychological stress load.

Time Load

Time load refers to the fraction of the total time that you are busy. When time load is low, sufficient time is available to complete all of your mental work with some time to spare. As time load increases, spare time drops out and some aspects of performance overlap and interrupt one another. This overlap and interruption can come from performing more than one task or from different aspects of performing the same task. At higher levels of time load, several aspects of performance often occur simultaneously -- you are busy and interruptions are very frequent.

Time load may be rated on the following three point scale.

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.
2. Occasionally have spare time. Interruptions or overlap among activities are very frequent.
3. Almost never have spare time. Interruptions or overlap among activities are very frequent or occur all the time.

Mental Effort Load

Mental effort load is an index of the amount of attention or mental effort required by a task regardless of the number of tasks to be performed or any time limitations. When mental effort load is low, the concentration and attention required by a task is minimal and performance is nearly automatic. As the demand for mental effort increases due to task complexity or the amount of information which must be dealt with in order to perform adequately, the degree of concentration and attention required increases. High mental effort load demands total attention or concentration due to task complexity or the amount of information that must be processed.

Mental effort load may be rated using the following three point scale.

1. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.
2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability or unfamiliarity. Considerable attention is required.
3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

Psychological Stress Load

Stress load refers to the contribution to total workload of any conditions that produce anxiety, frustration, or confusion while performing a task or tasks. At low levels of stress, one feels relatively relaxed. As stress increases, confusion, anxiety, or frustration increase and greater concentration and determination are required to maintain control of the situation.

Psychological stress load may be rated on the following three point scale.

1. Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.
2. Moderate stress due to confusion, frustration, or anxiety noticeably add to workload. Significant compensation is required to maintain adequate performance.
3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

In order to develop your individual workload scale, information is needed regarding the amount of workload you feel is imposed by various combinations of the dimensions described above. You will be asked to rank order the workload associated with each of the combinations.

You will be given a set of 27 cards with the combinations from each of the three dimensions. Each card contains a different combination of levels of time load, mental effort, and psychological stress. Your job is to sort the cards so that they are rank ordered according to the level of workload represented on each.

When completing your card sort, please consider the workload imposed on a person by the combination represented on each card. Arrange the cards from the lowest workload condition through the highest condition. You may use any strategy you choose in rank ordering the cards. One strategy that has proven useful is to arrange the cards into a number of preliminary stacks representing "high", "moderate", and "low" workload. Individual cards can be exchanged between stacks if necessary and then rank ordered within stacks. Stacks can then be recombined and checked to ensure they represent your ranking of lowest to highest workload. However, the choice of strategy is up to you and you should choose the one that works best for you.

There is no correct order. The order should be what, in your best judgment, best describes the progression of workload from lowest to highest for a general case rather than any specific event. When performing the card sort, use the descriptors printed on the cards. Please remember not to sort the cards based on a particular task (such as flying an airplane). Sort the cards according to your general view of workload and how important you consider the dimensions of time, mental effort, and psychological stress load to be.

The card sort is being done so that a workload scale may be developed for you. This scale will have a distinct workload value for each possible combination of time load, mental effort load, and psychological stress load. For example, if your rating for time load, mental effort load and psychological stress load were all "1", then the overall workload would be minimal. On the other hand, if they were all "3", then workload would be maximized.

The sorting will probably take 30 minutes to an hour. Please feel free to ask questions at any time. Thank you for your cooperation.