STINFO COPY

AFRL-HE-WP-TR-2006-0157

Encoding and Displaying ATR Designations in SAR Images

Pe'erly Setter
Joel Norman
Maya Lipkin-Goldberg
Hadas Marciano

Ergonomics and Human Factors Unit
Institute of Information Processing
and Decision Making
University of Haifa
Haifa, Israel

September 2006

Interim Report for September 2005 to September 2006

20070413109

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Cognitive Systems Branch
Wright-Patterson AFB OH 45433-7604

NOTICE

AFRL-HE-WP-TR-2006-0157 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR

//SIGNED//

DANIEL G. GODDARD
Chief, Warfighter Interface Division
Air Force Research University

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* September 2006	**2. REPORT TYPE** Interim	**3. DATES COVERED** *(From - To)* September 2005 - September 2006

4. TITLE AND SUBTITLE
Encoding and Displaying ATR Designations in SAR Images

5a. CONTRACT NUMBER FA8650-04-D-6405
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER 62202F

6. AUTHOR(S)
Pe'erly Setter, Joel Norman, Maya Lipkin-Goldberg, Hadas Marciano

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER 7184X03C

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Ergonomics and Human Factors Unit
Institute of Information Processing and Decision Making
University of Haifa
Haifa, Israel

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Cognitive Systems Branch
Wright-Patterson AFB OH 45433-7604

10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/HECS

11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-HE-WP-TR-2006-0157

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.
Cleared as AFRL/WS-06-2855 on 16 Jan 2007.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Automatic target recognition (ATR) may be integrated into advanced imagery exploitation capabilities. The design of the imagery analyst's interface can be expected to determine both the usability/operability of the aid and its acceptance by the user community. This study reports on exploratory development and evaluation of such an interface.

15. SUBJECT TERMS
Automatic Target Recognition (ATR), Synthetic Aperture Radar (SAR), Target Direction

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR	30	Gilbert G. Kuperman
					19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

THIS PAGE LEFT INTENTIONALLY BLANK

Contents

THIS PAGE LEFT INTENTIONALLY BLANK

1. Executive Summary

The present study both compares two methods of providing analysts with ATR (Automatic Target Recognition) confidence ratings, and examines whether providing these ratings yields superior performance to ATR designations without added confidence information. Twelve analysts participated. They were presented with SAR (Synthetic Aperture Radar) images each of which contained one of three types of ATR designations. Two of the designation types included ATR confidence ratings and a third did not. The two ATR designations types that included confidence levels specified the confidence of the ATR system (three levels: less than .70, about .80, and more than .90) either by surrounding the SAR item with three different **shapes** or by placing a **number** next to the item. The third designation type did not give confidence information, but simply surrounded those SAR items designated by the ATR as targets with an ellipse. All the designations were in partially transparent red. Each SAR image contained between 10 and 18 items, 5 to 12 of these being targets (T62, BMP2, or, BTR60) and the others distractors (ZIL131 or D7). The analysts were also given a post-experimental questionnaire to assess their subjective opinions of the three designations.

Hit Rates (HR) and False Alarm Rates (FAR) and the signal detection statistic d' were calculated and analyzed[1]. These measures of performance did not yield any major differences between the three designation types, with the exception of slightly (but borderline significant) fewer FARs for the ellipses. The post-experimental questionnaire indicated that the subjective feelings of the analysts just barely favored presenting ATR

[1] Note that throughout this report, HRs, FARs, and d's refer to the performance of the participants in the study and not to the performance of the ATR system.

confidence designations, and that of the two modes of presenting the confidence information they preferred the shapes over the numbers.

As very minor differences were found between the three designation types, the question was raised whether the ATR confidence ratings were, perhaps, not heeded by the participating analysts. To check this all instances where targets were designated with an ATR confidence level were combined (numbers and shapes) and HRs and FARs were calculated for the three confidence levels. It was seen that both HRs and FARs decreased with decreasing confidence levels, as had been shown in our earlier study (Setter, Norman, & Marciano, 2004), indicating that the analysts were indeed aware of the ATR confidence designations. Finally, it was argued that while the study appeared to indicate that ATR confidence designations do not benefit analysts' performance, there are good reasons to assume that in a "real life" situation they will.

2. Introduction

Human image analysts are often aided by ATR (Automatic Target Recognition) systems that designate (cue) which item in a display the ATR system identifies as a target. Such analysts utilize these ATR target designations (also labeled cuing or aiding) in deciding whether an element in the image is a target or not. ATR systems not only designate targets, but are also capable of giving the analyst an assessment of confidence that the designated element is indeed a target. The present study both compares two methods of providing the analysts with ATR confidence ratings, and examines whether providing the analysts with these ratings yields superior performance to ATR designations without added confidence information.

Several studies have examined the effects of aiding human decision processes with automatic systems in such environments as hospital intensive care units, nuclear power plants, and aircraft cockpits. The general consensus from these studies is that the addition of automatized, but invariably imperfect, decision aids causes what is sometimes labeled "automation bias" (e.g., Maltz & Shinar, 2003; Merlo, Wickens, & Yeh, 1999; Skitka, Mosier, & Burdick, 1999). This bias is seen in two types of prevalent errors: 1) Omission errors, where the human operator fails to detect a target that has not been specified by the automatic system (these are "misses" in signal detection jargon, and are the complement of "hits", exemplified by the HR in the present study); and 2) Commission errors, where the human operator accepts an erroneously designated target by the automatic system (these are "false alarms" when a target is designated). It is assumed that these errors stem from "over-trust" in the automatic systems (e. g., Maltz, 2005). What is more, the general finding is that operators rely more heavily on automatic cueing systems the more reliable that cueing is seen to be (e.g., Maltz & Meyer, 2001; McFadden, Giesbrecht, & Gula, 1998).

Similar conclusions have been found in studies directly related to the general question of interest in this report, namely the effects of ATR cueing on SAR analysts' performance. In one such study, See, Davis, and Kuperman (1997) presented 12 participants with SAR images. The task was to search for the same target in each SAR image and mark it with the aid of the mouse. The participants were also requested to give confidence ratings on their responses. These researchers compared performance on cue aided and unaided presentations. In the aided block of trials, four boxes surrounded SAR items, and these were absent in the unaided block. Four dependent variables served in the comparison of performance on aided and unaided presentations. Three of these, Percent of Correct Localizations, the Signal Detection Measure of Sensitivity d', and Response Time, did not yield any significant differences between aided and unaided presentations, the performance being remarkably similar. However, the fourth dependent variable, the confidence ratings, did yield a significant main effect of aiding, with overall higher confidence ratings for the aided presentations. A second independent variable was also manipulated in this study, the difficulty of the image scanned (amount of clutter). A significant interaction was found between this variable and the aiding variable in the case of the confidence ratings, indicating that the aiding increased the participant's confidence particularly in the case of the more difficult images. These researchers also carried out further analyses which showed that operator performance and confidence was influenced by the reliability of the cuing, such that if all the cues were false alarms performance was worse than when no cues were given. The opposite effect was also found, better performance when the cues appeared to be valid.

In another study, Setter, Norman, and Marciano (2004) examined the effects of ATR reliability on analysts' performance. Trained analysts performed a SAR target identification task with stimulus materials very similar to those used in the present study. Their performance with ATR cueing was compared to performance with no ATR cueing. In the ATR cued image sets, the analysts were informed that the ATR reliabilities (proportion of correct target designations) were 0.80, 0.50, or 0.33. Comparing cued and uncued image sets indeed yielded higher hit rates (HRs) on the image set with the high reliability (0.80), and inferior performance on the image set with the low reliability

(0.33), compared to the 0.50 reliability. However, the same was true for the false alarm rates (FARs); these also yielded higher values with the high reliability set and lower with the low reliability set. These findings indicate that the ATR reliability changed the analysts' criterion, but not their identification performance. This was confirmed in a Signal Detection analysis where the d' measure of sensitivity values were found not to differ with changes in ATR reliability.

Taken together, the findings of the latter two studies lead to similar conclusions. First, that under the conditions studied, the main effect of ATR target designation aiding is on the subjective confidence levels of the participants in the experiments. This was directly measured in the See et al. (1997) study, and can be inferred from the Setter et al. (2004) study where the participant's criterion shifted to a more lax one with higher reliabilities, indicative of greater confidence. This is similar to the "over-trust" discussed above. What is more, both studies indicate that the effects of aiding on participant confidence are a result of the perceived reliability of that aiding. On the other hand, both studies indicate that the addition of ATR target designations for aiding the identification process does not improve performance, at least in a simplistic interpretation of performance (see more on this point in the Discussion section).

In the present study SAR images were used. These images contained between 10 and 18 SAR elements, some of which were targets and some of which were distractors. The analysts were presented with three blocks of 24 SAR images each, each block using a different ATR target designation method. These consisted of two ways of designating

ATR confidence, shapes or numbers, and simple ATR designations without confidence information. The study focused on comparing analysts' performance on these three types of designations.

3. Method

3.1. Participants

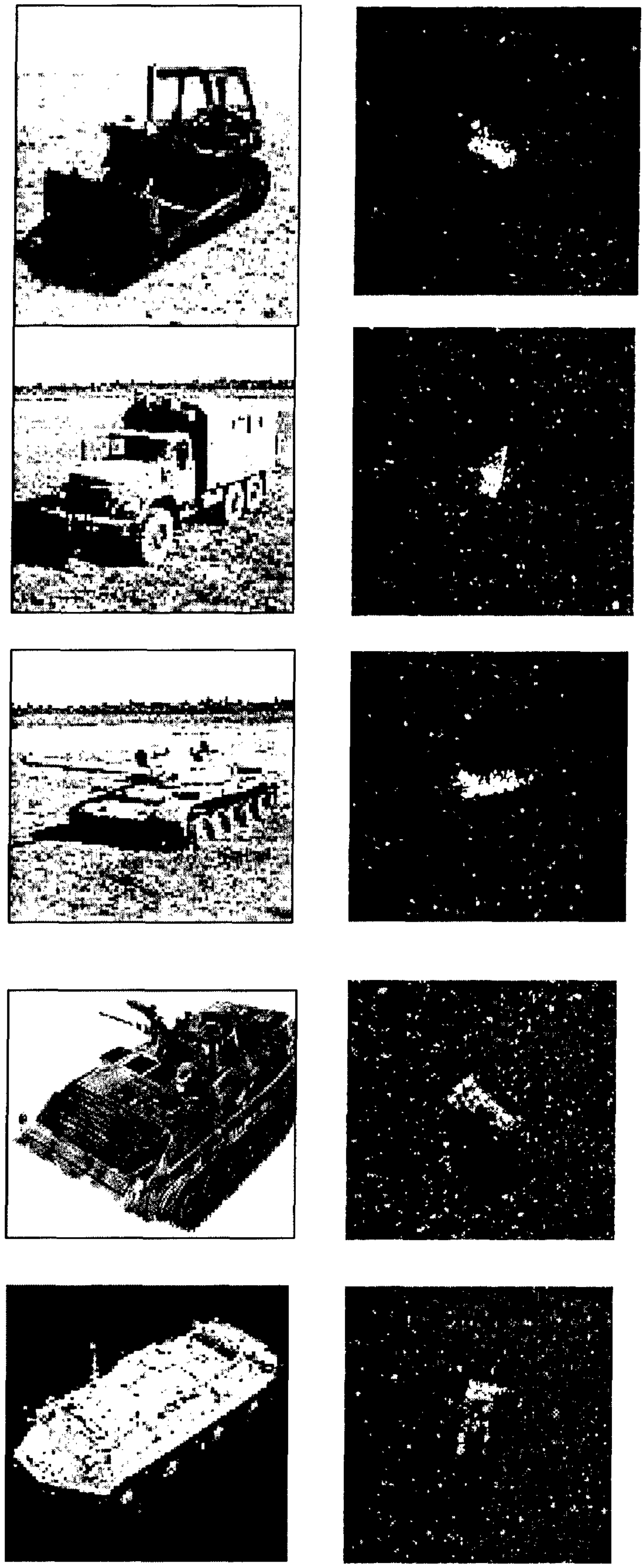
Twelve trained analysts participated in the study, 5 women and 7 men.

3.2. Apparatus

The experiment was carried out on 4 IBM Thinkpad Notebook portable PC computers (two with 14" screens and two with 15" screens), all utilizing 1024X768 pixel resolution.

3.3. Images

The 72 images were created from items in the MSTAR SAR Database. Three SAR items served as targets: T62, BMP2, and BTR60, and two items served as distractors: ZIL131 and D7 (truck and bulldozer) (see Figure 1 for photographs and SAR images of single exemplars of these targets and distractors). Each image contained between 10 and 18 items, 5 to 12 of these being targets the others distractors (see examples in Figures 3, 4 and 5). The items were inserted into MSTAR SAR backgrounds with the aid of Photoshop 7.0 ME graphic software. The arrangement of the items in the images was not random, but in accord with known combat doctrines. Many other precautions were taken to make the images appear very authentic, such as making sure that all the shadows, those of the targets, the distractors, and the background were in the same direction.



Distractor – D7

Distractor – ZIL131

Target – T62

Target – BMP2

Target – BTR60

Figure 1: Examples of the five vehicles that served as targets and distractors in the present study: photographs and SAR images of single vehicles. (The experimental SAR images contained between 12 and 18 such vehicles).

3.4. Procedure

The experiment was carried out in a fairly large lecture hall, with the four notebook computers placed on separate tables in different parts of the hall. This allowed the testing of four analysts at a time. The experiment began with a training session. Its aim was to acquaint the analysts with the three targets and two distractors. The analysts received a set of instructions (see Appendix 1) together with rather large black and white photographs of the 5 items (targets and distractors) and next to them examples of their SAR images (see Figure 1). These pictures were available to the analysts throughout the training session. The training session consisted of four 25-trial blocks. On each trial the analysts viewed a single SAR item (one of the three targets or one of the two distractors) under which two clickable buttons appeared, with the word "target" on the right and the word "distractor" on the left (in Hebrew). When the analysts pressed the wrong button, a feedback tone informed her/him of the error. An example of a training session trial appears in Figure 2.

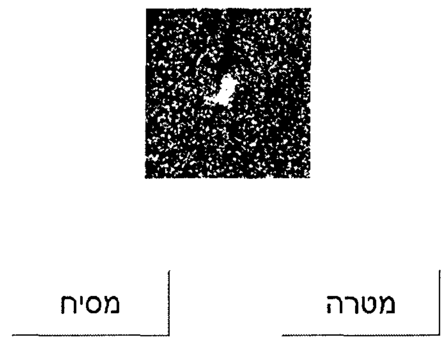


Figure 2: An example of a training session trial. The image is of a single SAR image of a vehicle, and the task is to click on the appropriate button; target (on the right) or distractor (on the left).

Following the training session the analysts were presented with the instructions for the main experiment (see Appendix 2) where their task was to determine which of the 12 to 18 SAR items in the image were targets and to mark them by moving a cursor with the aid of the mouse to the target and clicking on the left mouse button. In each image some of the items were designated by the ATR as targets and others were not. In all the images about 80% of the designated items were targets and 20% were distractors. Likewise of all the targets that appeared in the image, the system designated about 80% of them.

Three types of ATR designation were presented to each analyst in separate blocks of 24 images each (within subject design). The order of these three was counter-balanced over the 12 participants. Two of the ATR designations included ATR confidence information and the third did not. All designations appeared in transparent red on the black and white image (see Figures 3, 4 and 5).

- **Numbers**. The items designated as targets received one of the following notations: "<7" meaning less than 70% ATR confidence; "~8" meaning about 80% ATR confidence; and ">9" meaning more than 90% ATR confidence (see Figure 3).
- **Forms**. The items designated as targets received one of the following notations: A circle around the item meaning less than 70% ATR confidence; a triangle around the item meaning about 80% ATR confidence; and a square around the item meaning more than 90% ATR confidence (see Figure 4). The logic behind the specific choice of shapes is that more angles in the shape indicate a higher confidence.
- **Ellipses**. These designations did not give confidence information. Those items designated as targets by the ATR were encircled by an ellipse (see Figure 5).

The exact choice of which target and which distractor should be designated by the ATR was randomly determined by the computer program.

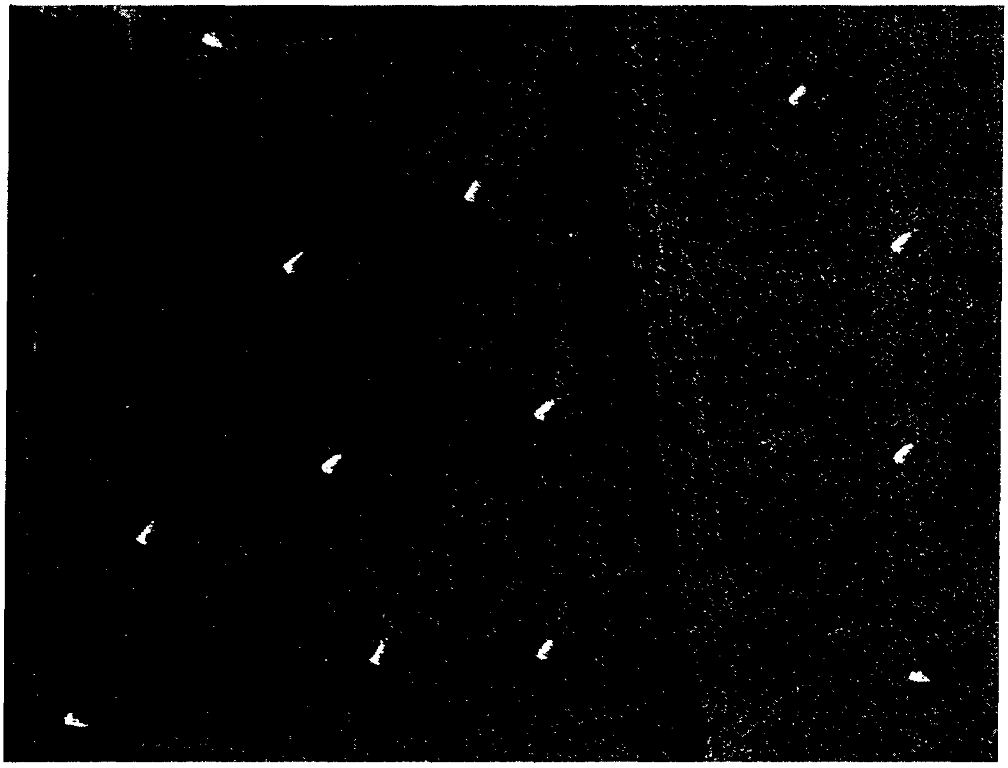


Figure 3: An example of a trial containing ATR designations that appear as numbers:
"<7" - less than 70%;
"~8" - about 80%;
">9" - more than 90%

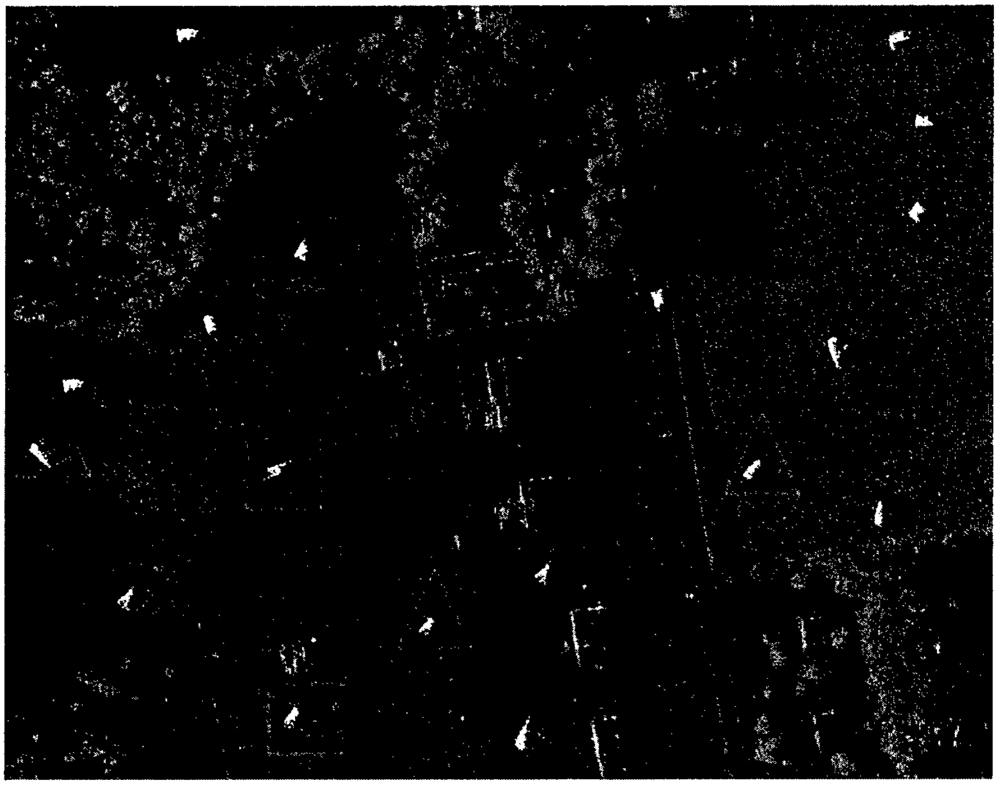


Figure 4: An example of a trial containing ATR that appear as shapes:
Circle - less than 70%;
triangle - about 80%;
square - more than 90%

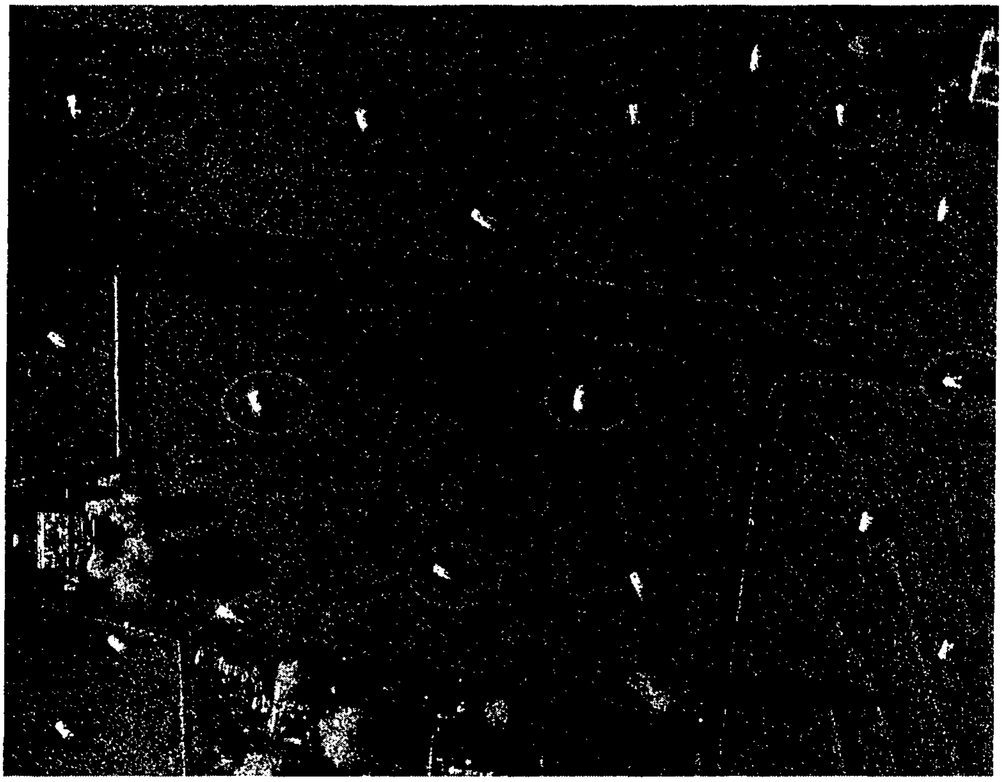


Figure 5: An example of a trial containing ATR designations that do not present ATR confidence information.

The three 24 image sets were presented with each of the three designations types and were counterbalanced over the 12 participants. In other words, each designation type was paired with one of the 3 image sets for four participants.

The participant analysts were instructed to examine each SAR image and determine which of the SAR items was one of the three targets, T62 or BMP2 or BTR60, and mark those targets with the aid of the computer mouse. They were explicitly told to scan all the SAR items and mark those three targets whether or not they had been designated by the ATR. They marked the assumed targets with the aid of a press on the left button of the mouse, which overlaid the target with a red X. Once they had marked the target they could not change that marking. Once the analysts were certain that they had marked all the targets in the image, they pressed the Enter key to advance to the next image.

Following the experiment the analysts were informed of how many points they had accumulated in the experiment. The point total was simply the number of overall hits less the number of overall false alarms. This point system served as a motivational factor as the analysts compared their results with those of their peers.

At the end of the experiment the analysts were given a short debriefing questionnaire, which included five questions (translated here from the Hebrew):

1) What are your feelings about the interpretation process? List all comments that might be relevant.

2) Do you feel that there is an advantage to the addition of ATR designations during the interpretation process?

3) Do you think that it is easier to work with the addition of ATR confidence (Numbers and Shapes) or without them (Ellipses)?

4) Which of the two ATR confidence designations was more convenient: Numbers / Shapes (circle one)?

5) Did you pay attention to the level of confidence of the ATR as you interpreted the images?

4. Results

The central aim of the study was to compare analysts' performance using the three ATR designation types. The overall performance with these three designation types is presented in Table 1. The table presents hit rates (HR), false alarm rates (FAR), and the Signal Detection statistic for these rates, d'. It should be noted that the d' values in the table are means over the 12 participants, and not the d' values that the mean HRs and FARs would yield.

Table 1. Mean Hit Rates (HRs), False Alarm Rates (FARs), and d' for the three designation types

Designation Type	Hit Rate	False Alarm Rate	d'
Numbers	0.65	0.34	0.94
Forms	0.66	0.31	1.01
Ellipses	0.63	0.28	1.10

As can be seen in the table the differences between the three designation types are slight. Indeed, Analyses of Variance (ANOVAs) indicated that the differences between the three designation types were not significant for HRs, $F(2,11) = 0.51$, ns., nor for d', $F(2,11) = 0.70$, ns. In the case of the FARs, a trend could be seen where $F(2,11) = 3.29$, $p<0.561$. A post hoc Duncan's Multiple Range Test (alpha = 0.05) indicated that the FAR for the ellipses was significantly lower than that for the numbers.

The three ANOVAs above were carried out on all the participants' markings; those on ATR designated items as those on non-designated items. In a separate analysis we examined the HRs and FARs in only those cases where the ATR had designated the item. Here, too, we found little effect of the type of designation. Table 2 parallels Table 1, but only includes the items that were designated by the ATR.

Table 2. Mean Hit Rates (HRs) and False Alarm Rates (FARs) for the three designation types (only designated items)

Designation Type	Hit Rate	False Alarm Rate
Numbers	0.68	0.39
Forms	0.69	0.40
Ellipses	0.67	0.36

Here too, there are only small differences for the three designation types. ANOVAs confirmed this obvious lack of differences; $F(2,11) = 0.31$, ns. for HR and $F(2,11) = 0.57$, ns. for FAR.

The responses to the post-experimental questionnaire yielded the following results:

a) **What are your feelings about the interpretation process? List all comments that might be relevant.**

The comments of the 12 analysts were not especially enlightening, and the information they supplied overlapped that found in the responses to the subsequent questions.

b) **Do you feel that there is an advantage to the addition of ATR designations during the interpretation process?**

Five analysts responded with "yes". Two others also wrote "yes", but with additional comments. One added that it is advantageous only for detection but not for identification, and the other that it is not advantageous when there are many targets (items) crowded near to each other. Two analysts responded "no", and one wrote "not always". One wrote that "it depends on the reliability of the ATR" and another that he "did not refer to the designations".

c) **Do you think that it is easier to work with the addition of ATR confidence (Numbers and shapes) or without them (Ellipses)?**

Seven analysts wrote that it was easier with the added ATR confidence, while 3 thought that it was better without. One other analyst reported that "it is not unequivocal", and one that "the added ATR confidence was not very convenient".

d) **Which of the two ATR confidence designations was more convenient: Numbers / Shapes (circle one)?**

Nine of the analyst chose the shapes as more convenient, two the numbers, and one noted that the two were equivalent.

e) **Did you pay attention to the level of confidence of the ATR as you interpreted the images?**

Four analysts responded that they did, four that they did not, and four that they did but only part of the time.

5. Discussion

Overall, our results point to very minor differences between the effects of the three ATR designation types. The comparison of the two ATR confidence designation types used in the study, numbers vs. shapes, did not yield significant differences in identification performance. However, 9 of the 12 analysts expressed preference for the shapes over the numbers, and one said that the two designation types are equivalent.

Since no performance differences were found between the two ATR confidence designations, one might ask if the analysts actually did pick up these designations. Recalling that Setter, Norman, and Marciano (2004) found that hit rates and false alarm rates increased concomitantly with increases in ATR confidence ratings, we decided to determine if the same was true in the present study. We examined HRs and FARs for all the designated items with the two types of ATR confidence designations, over both types (numbers and shapes). The results appear in Table 3 below.

Table 3. Mean Hit Rates (HRs) and False Alarm Rates (FARs) as a function Of ATR designation confidence

ATR confidence	HR	FAR
>.90	.80	.52
~.80	.69	.37
<.70	.55	.30

As can be seen in Table 3, both the HRs and FARs increase systematically with increasing ATR confidence designations, and in both cases the effect is significant ($p<0.0005$ and $p<0.0001$, respectively). This finding mimics those of Setter et al. (2004) and indicates that the analysts did indeed pick up and were influenced by the ATR confidence levels of the ATR designations. It should also be noted that the methods of presenting the analysts with ATR confidence ratings were quite different in the two studies. In the Setter et al. (2004) study the ATR confidence level was presented to the analysts before each block of identification trials, and the analysts were told that this was "the reliability level" of the ATR system. In contrast, in the present study, specific items within a single SAR image received numerical or shape designations, and these could vary between designated items within a single SAR image. Yet both types of ATR reliability information influenced the analysts in a similar manner.

One further point that merits note is that the two types of ATR confidence designations, shapes and numbers, did not yield performance superior to that of the simple encircling of the items with an ellipse. This would appear to indicate that there is no advantage to adding ATR confidence designations to the display, but we feel that such a conclusion would be premature. It should be emphasized that the analysts in the present study had not had any previous experience with ATR confidence designations, and it is quite possible that with training and experience they could learn to benefit from such designations. This is also true of the two studies reviewed in the introduction, where the participants had not had any previous experience with ATR designations. What is more, recall that in our earlier study (Setter et al., 2004) there was a concomitant increase in HRs and FARs with increasing ATR confidence, and that we found similar results in a post hoc analysis of the findings of this study. Findings of this sort can be seen to imply that there is no true increase in performance level with higher confidence designations, but simply a manifestation of the "over-trust" attitude mentioned in the introduction. However, it should be noted that we created the artificial SAR images from a data bank of very similar SAR images of single vehicles. We had no way of knowing what confidence level a real ATR system would have assigned to each of these images. One can surely assume that in a true ATR system higher confidence levels will be assigned to "superior" SAR images. In other words, we would assume that the SAR images that receive higher ATR confidence levels will have a higher probability of being true targets. In such a case, the fact that the analysts viewing these high ATR confidence images lower their criterion cutoff point; i.e., become more lax in their willingness to call the image a target, will in the "real life" case yield superior performance. There will be an increase in the HRs without the concomitant rise in FARs, which is what we are seeking. In other words, the fact that the participant analysts tend to take the ATR confidence levels seriously and change their willingness to accept these ATR designations when the confidence is high will, in "real life" situations, improve actual performance.

6. References

Maltz, M. (2005). Modeling the efficacy of automated aids in target acquisition under conditions of heavy workload. *Optical Engineering, 44*(8), 086201.1-086201.6.

Maltz, M. & Meyer, J. (2001). Use of warnings in an attentionally demanding detection task. *Human Factors, 43*(2), 217-226.

Maltz, M. & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human Factors, 45*(2), 281-295.

McFadden, S. M., Giesbrecht, B. L. & Gula, C. A. (1998). Use of an automatic tracker as a function of its reliability. *Ergonomics, 41*(4), 512-536.

Merlo, J. M., Wickens, C.D. & Yeh, M. (1999). *Effect of reliability on cue effectiveness and display signaling*. Technical Report ARL-99-4/FED-99-3.

See, J. E., Davis, I. & Kuperman, G. G. (1997). *Automatic target cueing and operator performance with enhanced APG-70 Synthetic Aperture Radar imagery*, AL/CF-TR-1997-0171, Armstrong Laboratory, Wright-Patterson Air Force Base, Ohio (ADA347534).

Setter, P., Norman, J. & Marciano, H. (2004). *The effects of ATR on analysts interpretation of SAR images*. Research Report, Ergonomics and Human Factors Unit, Institute of Information Processing and Decision Making, University of Haifa (in Hebrew).

Skitka, L. J., Mosier, K. L. & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991-1006.

7. Appendix 1

Instructions at the start of the experiment (translated from the Hebrew)

The purpose of this experiment is to examine the effects of different confidence levels of ATR (Automatic Target Recognition) on target identification. The ATR system is a computerized system that identifies targets. Ideally it should only designate targets and no other objects, but due to several reasons, it can make errors, designating distractors as if they are real targets and missing real targets. In other words, because of the inaccuracy of the system some of the targets in the experiment will not be designated and some of the distractors will be. In the experiment SAR images containing a variety of vehicles will be presented. Some of them will be targets and some distractors. Your task will be to distinguish between targets and distractors.

The **targets** in the experiment will be vehicles of the following types:

T62, BMP2, BTR60.

The **distractors** in the experiment will be vehicles of the following types:

ZIL131, D7

At the start of the experiment there will be a training phase where you will practice distinguishing between targets and distractors. On each practice trial you will see a SAR image of a single vehicle and your task will be to press on the "target" button if the image is of a target or on the "distractor" button if the image is of a distractor. The buttons will appear on the screen. When you make an error you will hear a tone notifying you that you erred.

Here are examples of the SAR images of the vehicles as they will appear in the training trials and to their left are normal images of the same vehicles.

(The analysts were presented with the images in Figure 1)

8. Appendix 2

Instructions for the Experiment
(translated from the Hebrew)

On each trial you will be shown a SAR image where some of the vehicles will be designated by the ATR system, and this means that the ATR system identified that vehicle as a target.

In this experiment you will receive varied information about the confidence level of the ATR system and your task will be to identify and designate the targets in the SAR image. The number of target and distractors will vary from image to image.

The experiment will comprise three parts:

- In one of the parts the ATR confidence will be presented by numbers: confidence lower than 70 by "<7", medium confidence of about 80 by "~8", and high confidence higher than 90 by ">9".
- In another part the ATR confidence will be presented by shapes surrounding the vehicle (the high confidence level by a square, the medium confidence level by a triangle, and the lower confidence level by a circle). The meaning of these shapes is that the more angles in the shape the higher the confidence.
- In a third part the vehicles will be designated by an ellipse surrounding them, but no ATR confidence levels will be presented.

(The three parts will not necessarily appear in the order above).

Your task will be to examine all the vehicles in the image (those designated and those not), and decide which of them is a true target and mark them with an X by pressing the left mouse button. **In other words, in each image you should mark all the vehicles which are targets, T62, BMP2, or BTR60 (whether designated by the ATR or not) and not mark the other vehicles.**

Take note: you have to mark the vehicle proper; otherwise the mark will not be picked up by the computer. Before you mark a vehicle you should make sure that it is indeed a target, since you will not be able to erase the mare once you have made it. Once you have marked all the targets in the SAR image press the ENTER key on the keyboard and you will be presented with the next SAR image. At the end of the experiment your performance will be evaluated and you will see a number that signifies how well you performed. It will be based on one point for each true target identification and minus one point for each time you identified one of the distractors as a target.