CVI*PLUS*: A MULTI-MEDIA, COMPUTER-BASED SOLUTION TO THERMAL SIGHT TRAINING

Jean L. Dyer
U.S. Army Research Institute Infantry Forces Research Unit
Fort Benning, GA
Barbara O Kane
Night Vision and Electronic Sensors Directorate
Fort Belvoir, VA

ABSTRACT

Combat vehicle identification (CVI) training materials for thermal sights have been lacking since the Army's fielding of thermal sights for anti-armor weapons in the 1970s. The night fratricide incidents in Desert Storm/Desert Shield can be attributed, in part, to inadequate thermal signature training. The paper covers training effectiveness research on a computer-based, multi-media training program, called CVI*Plus*, aimed at providing thermal signature training to support most of the Army's current and future thermal sights.

An assumption underlying program development was that the dynamic nature of thermal imagery and the uniqueness of thermal cues demand actual, not simulated, imagery to train skills adequately. Consequently, the training data base is digitized, high-resolution, thermal images of combat vehicles, collected specifically for the program. Night and day, black-hot and white-hot, thermal images of US and nonUS vehicles at eight aspect angles at four ranges are included. Visible images of each vehicle are shown as well. The version of the program available for research included pre- and posttests, a library of all images, and interactive training and testing exercises.

Three training experiments were conducted to determine the program's effects, determine effective training strategies, and identify needed improvements. The first experiment examined part-task training issues. The second addressed the effectiveness of fixed-pace training with knowledge of results feedback versus self-paced training with knowledge of performance feedback. The third focused on training at near versus far ranges. Within each experiment, the extent to which skills transferred to imagery not included in the training exercises was also examined. The findings supported changes to the instructional design of the program.

ABOUT THE AUTHORS

Dr. Jean L. Dyer is a research psychologist at the Army Research Institute at Ft. Benning, Georgia. She is currently specializing in night operations training, focusing on night equipment and leader training. Other areas of expertise include anti-armor weapon system training, small-unit training, training effectiveness analyses of training devices, and unit performance at the combat training centers. Dr. Dyer received her Ph.D. in measurement, evaluation, and research design from Michigan State University.

Dr. Barbara O'Kane is an engineering psychologist at the U.S. Army Night Vision and Electronic Sensors Directorate (NVESD) at Ft. Belvoir, Virginia. She is currently responsible for the Perception Studies Group at NVESD. This group performs thermal imagery training development, and conducts studies to support thermal system designs, compression, and image enhancement, as well as user interfaces with machine vision. Dr. O'Kane received her Ph.D. in psychophysics from Brandeis University.

CVI*PLUS*: A MULTI-MEDIA, COMPUTER-BASED SOLUTION TO THERMAL SIGHT TRAINING

Jean L. Dyer
U.S. Army Research Institute Infantry Forces Research Unit
Fort Benning, GA
Barbara O'Kane
Night Vision and Electronic Sensors Directorate
Fort Belvoir, VA

When the first-generation, thermal night sights for the TOW and DRAGON were fielded in the 1970s, no training materials were developed for target acquisition skills. Also lacking were materials on the characteristics of this new technology, e.g., fundamental concepts which enable soldiers to better interpret and understand the dynamic imagery seen in thermal sights under diverse atmospheric and combat conditions. Some excellent training materials were produced by the Night Vision Laboratory (Orentas, Zegel, & Gonzalez, 1991; Palmer, D'Agostino, & Lillie, 1982) and the Army Research Institute (Rollier, Champion, Roberson, & Graber, 1988) to address these problems. However, they were interim, partial solutions. Consequently, training deficiencies have continued to the present including the absence of training materials for the most recent, 2d-generation sights.

The training program for 2d-generation sights described in this paper was sponsored by the Army's Product Manager for Forward Looking Infrared (PM-FLIR) and executed by the Night Vision and Electronic Sensors Directorate (NVESD). The Army Research Institute (ARI) at Fort Benning, GA conducted the training effectiveness research reported here, simultaneously with program development.

THE THERMAL TRAINING PROGRAM

Imagery

The CVI*Plus* training program includes actual thermal imagery of vehicles, not simulated imagery typical of gunnery training devices. Actual imagery was deemed critical, given the task at hand, even though the number of vehicles was restricted due to the cost of imagery collection. High image fidelity and the ability to display the dynamic nature of thermal imagery were viewed as critical to transferring skills to field situations.

Imagery Collection. Imagery was collected using an 8-12 micron, calibrated thermal imager (Agema Model T1000). Each 12-bit digital frame used was created by averaging 16 frames. Frame averaging was used to effectively improve the imager's sensitivity to more closely match tactical systems. The imager was positioned 400 meters from the vehicles for ground-to-ground collection, and collected in the wide and narrow fields of view (FOV). Vehicles were exercised 20 minutes prior to imagery collection. They were then positioned in eight static orientations with respect to the imager; engines were idling. The eight vehicle orientations were: 0, 30, 90, 150, 180, 210, 270, and 330 degrees. The imagery was collected in April and August in the US temperate climate.

Imagery Processing. Raw imagery was processed in several steps for the training program. It was cropped to fit the window size of the program and to create a suitable vehicle-to-background ratio. The images were then scaled to simulate views at four ranges by using a sensor convolution kernal employing the narrow FOV for the close ranges and the wide FOV for the far ranges (i.e., applying the appropriate sensor spatial degradation for each range). Generic ranges, labeled Close-up, Ranges 1, 2, and 3, were used to keep the program unclassified and because no particular thermal sight was being simulated. The 12-bit images were displayed in an 8-bit mode by a process that ensured the hot spot features were visible. In the field, the soldier can adjust the sight's brightness and contrast controls to bring out the hot-spot features. But in the CVI*Plus* program only one setting was available. Therefore this setting had to show the features clearly to represent field capability.

Program Features

Program Modules. The computer program is Windows-based. Soldiers progress at their own rate. The version available for research purposes had five modules: Sensor Controls, Vehicle Basics, Image Library, Training Exercises, and Tests. Tests paralleled the Training Exercises; special pre- and posttests were included as well. The Image Library, Training Exercises, and the Tests were used in the training effectiveness research. Program descriptions presented here reflect the research version available, not the current version.

Imagery Database. The program contained 14 vehicles: M1 Abrams, M60A3, T72, M551 Sheridan, M2 Bradley, BMP-2, LAV-25, BTR-80, ZSU 23-4, M998 HMMWV, M113, ZIL-131, 2.5 ton truck, and 5 ton truck. For each vehicle, night and day thermal images in black-hot and white-hot, at eight aspects, and at four ranges were available. The Close-up range provided a distinct view of the heat plant of each vehicle; the farthest range, a view typical of what a gunner sees at tactical engagement ranges. Visible imagery was displayed only at the Close-up range.

The Image Library is what the name implies, a storehouse or reference for all vehicle images. A pull-down menu allows the user to select vehicles. Two- and three-vehicle displays can be selected. Side-by-side displays allow comparisons of different vehicles or of different images of the same vehicle (see Figure 1). Buttons on the tool bar control the imagery conditions that are displayed: day/night/visible, black-hot/white-hot, and range. At the bottom of the screen are two (three) rows of smaller images presenting the eight aspects of the two (three) vehicles in the screen's center. These images control the aspects in the central display. During the research, the Image Library was used by the instructor to present thermal cues to groups of soldiers, as the prototype program did not have an instructional module on cues.

Training Exercises. Two training exercises, Signature Challenge and Vehicle ID, were used in the experiments. Both required soldiers to name vehicles, but their training features differed.

Signature Challenge (SC) was a timed exercise, requiring the soldier to discriminate between as few as two or as many as eight vehicles. Responses were limited to the vehicles within a vehicle set. After each trial, the soldier was given either confirmatory or corrective feedback on his response. Sets of vehicles were sequenced to meet training needs. Gates or criteria and vehicle exposure time were established for each vehicle set. A soldier could not progress to the next set until the gate was achieved. A response could not be changed, once made. Images were presented randomly within each set.

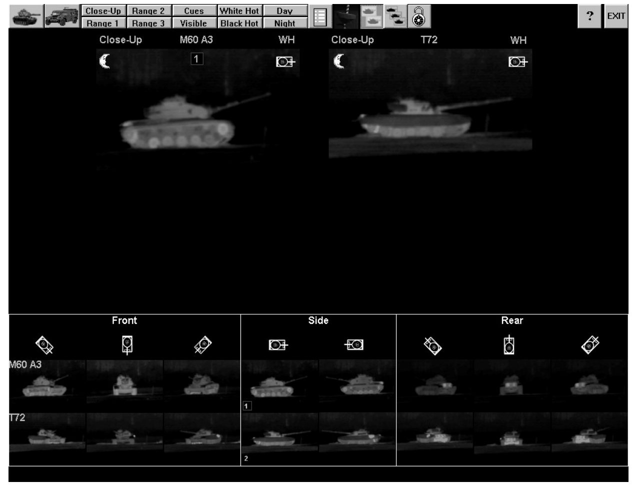

Figure 1. Image Library.

Vehicle ID (Veh ID) required the soldier to name the vehicle and its aspect. It was not timed, and responses could be changed before proceeding. Feedback on the correctness of the responses was given after each trial. When the soldier erred in identifying the vehicle or aspect, the image of the vehicle named was shown beside the original display, providing corrective visual feedback. Images were presented randomly. In contrast to SC, all 14 vehicle names were presented as response options; there was no gating capability; and vehicle sets could not be sequenced.

Summary feedback at the end of a vehicle set in both SC and Veh ID was total % correct. Percent correct for each vehicle was not given. In Veh ID, the % correct for aspect was also given. Lastly, there was no capability for exiting either the SC or the Veh ID exercises for self-study and returning to the exit point.

Tests. The SC and Veh ID tests paralleled the training formats, but there was no feedback after each trial nor at the end of a vehicle set. The pre- and posttests were not timed, and had a visible and thermal component. Each component had 30 images, three displays of ten vehicles, at the closest range. Thermal images were night, white-hot. The pre- and posttests were identical; vehicles and their aspects in the visible and thermal tests were also identical. Images were presented randomly. The names of all 14 vehicles were displayed as choices.

PURPOSE OF THE EXPERIMENTS

Three experiments were conducted to determine the program's effects, determine effective training strategies, and identify needed improvements. The first examined part-task training issues. The second addressed the effectiveness of fixed-pace training with knowledge of results feedback versus self-paced training with knowledge of performance feedback. The third focused on training at near versus far ranges.

Research Paradigm

The research paradigm in each experiment was the same (see Table 1). In each, only six or eight of the 14 vehicles were trained. Thus, across the experiments, the pre- and posttests provided a comparison of the program's general effects on vehicles trained to those not trained.

Table 1. Sequence of Events in Each Experiment

Pretests	Thermal and visible components. Same in each experiment.
Group instruction	Instruction via overhead projection of the Image Library. Varied with the experiment.
Training Exercises	Training conditions varied with the experiment. Soldiers randomly assigned to training conditions within each experiment. Used SC and Veh ID training exercises.
Transfer Tests	Transfer tests varied with the experiment. No transfer imagery included. Used SC and Veh ID tests.
Posttests	Same as pretests.

Group instruction was conducted by a military instructor who used the Image Library via an overhead projection system. The two-display screen showed the thermal and visible image of each vehicle to allow soldiers to link thermal with visible features. Thermal cues were identified as each aspect was displayed. Vehicles with similar thermal cues were then compared. Thermal cues were based on the heat plant of the vehicle (engine, exhaust), heat created by friction (suspension), and other cool and hot spots on the turret and hull, as described in O'Kane, Biederman, Cooper, and Nystrom (1997). The instruction was consistent with Biederman's (1987) theory of recognition-by-components, as it focused on the components seen from most aspects and at most distances. The vehicles, ranges, and other conditions displayed varied with the experiment.

Experiment 1: Part-Task Training

Learning to identify a large number of vehicles is a difficult skill that cannot be achieved in a single training session. Some form of part-task training is necessary. Experiment 1 was our initial examination of how to divide a pool of vehicles into sets and to create an effective training sequence with these sets. Questions raised in earlier aircraft recognition research (Gibson, 1947, Whitmore, Cox, & Friel, 1968) reflected similar concerns on sequencing many aircraft within a training program.

Experiment 1 compared part-task training schedules. The critical question in part-task training is how well training on components transfers to the whole task. Fractionated training schedules (Proctor & Dutta, 1995) break a task into components that are performed concurrently, as opposed to segmented schedules, which involve parts that must be performed sequentially. Fractionated schedules apply to the training of vehicle identification, as we considered the whole task to be that of identifying all images in the program, with task components referring to subsets of images from the total pool.

We compared two pure-part training schedules. In a pure-part schedule, each part is trained separately, with the final step being training on the whole task. Eight vehicles were trained. One pure-part strategy had four sets of two vehicles (Diad). The other (Triad) had two sets of three vehicles and one pair. The part-task vehicle set sequence was counter-balanced within each schedule. Each schedule ended with training on all eight vehicles. All aspects of each vehicle in each set were displayed. Training was conducted with night, white-hot imagery at Range 1, using the SC training exercises. Pass criterion was no more than four errors: 79% for two vehicles, 85% for three vehicles, and 93% for all vehicles.

The transfer test, using Veh ID, had 64 displays of the eight vehicles. Two transfer conditions were examined using flank and oblique aspects. Range was increased for 38% of the images. Day thermal imagery was used for another 38%. Front and rear views were the no transfer condition (24%).

Soldiers (*N* = 24), from a light Infantry unit, were randomly assigned to the Diad and Triad conditions. Average time in service was 2.5 years, with half serving only one year.

The primary questions in Experiment 1 were:

- Which pure, part-task schedule is more effective in training vehicle identification?
- Which schedule produces the higher transfer to other exemplars of the vehicles?
- Do identification skills transfer more easily to variations in range or variations in the time of day?

Secondary questions of interest were:

- Does training in thermal imagery transfer to visible imagery?
- Does training in thermal imagery transfer to vehicles not trained?
- What vehicles tend to be confused with each other during training and testing?

Experiment 2: Fixed-Pace with Knowledge of Results versus Self-Paced with Knowledge of Performance

The two major training modules within the program, SC and Veh ID, offered an opportunity to compare two training strategies. A timed, fixed paced strategy with knowledge of results (Fixed-KR; correct or incorrect) was implemented with SC. A self-paced mode with corrective, visual feedback (Self-KP) was implemented with Veh ID. We viewed the corrective, visual feedback as a variation of knowledge of performance (KP).

The Fixed-KR strategy corresponded to the flash technique of aircraft recognition training examined by Gibson (1947). But our exposure times were 7 and 10 sec, compared to the less than 1 sec exposures cited by Gibson. With Veh ID, soldiers could pace themselves and profit from the corrective visual feedback. Research in motor learning has shown that KP can be more effective than KR alone (Proctor & Dutta, 1995). In a group instructional mode, individualized, corrective visual feedback is not possible. And this training strategy was not used historically (Gibson, 1947; Warnick & Smith, 1989; Whitmore et al.,1968). On the other hand, this type of feedback is easily provided by a computer.

In Experiment 2, three tanks and three APCs were trained. Three vehicle sets formed the Fixed-KR training: tanks, APCs, and lastly all vehicles. As Veh ID had no sequential, vehicle set capability, soldiers were exposed to only the all-vehicle set in the Self-KP condition. Pass criterion was 85% for the part-task sets; 91% for the all-vehicle set. For Veh ID, the research staff ensured soldiers attained the gate, as it was not computer-controlled. In each vehicle set, all aspects of each vehicle were displayed at Range 2 in night, white-hot imagery.

The transfer test, using Veh ID, had 48 displays. Two transfer conditions were investigated: night and day black-hot imagery, each constituting 38% of the images. Flank and oblique aspects were used. Front and rear views were the no transfer condition (24%).

Soldiers (*N* = 35) from the Bradley Leader Course (BLC) and a National Guard (NG) unit were randomly assigned to the two training conditions. Half the BLC students had three years or less of service, compared to 25% of the NG soldiers.

The primary questions were:

- Is a training strategy that incorporates timed trials and provides knowledge of results more or less effective than a self-paced training strategy that provides knowledge of performance via corrective, visual feedback?
- Is transfer greater to vehicles displayed in black-hot day or black-hot night imagery? Is transfer affected by the training strategy?
- Does practice in deciding aspect improve aspect scores in a transfer condition?

Secondary questions were the same as in Experiment 1.

Experiment 3: Training at Near and Far Ranges

The last experiment compared the effects of training at near versus far ranges. The point has been made in previous research (Cockrell, 1979; Warnick & Smith, 1989) that training on cues seen primarily at close distances may be counter-productive in learning critical cues detectable at tactical ranges or under degraded conditions. On the other hand, there is a strong argument for the need to train thermal imagery at close distances, so individuals thoroughly understand the heat plant and structural characteristics that generate the vehicle's thermal signature.

Two tanks, two tracked APCs, and two wheeled APCs were trained. For each vehicle type, one vehicle was US, the other nonUS. Soldiers practiced at one of two range bands: Near or Far. The Near range band included the two nearest ranges (Close-up and #1); the Far band, the two farthest ranges (#2 and #3).

Soldiers first had a single iteration of Veh ID training where they were exposed to the cardinal aspects of each vehicle (24 displays). They then went to a series of eight gated, SC exercises. US/NonUS vehicles were paired by type into three sets; presented first at the closer distance within the range band, and then at the farther range. Vehicles were regrouped in the last two sets; all

US presented first, then all nonUS. The final two sets were at the farthest distance within the Near or Far range band. The pass criterion was 85%. Within each SC vehicle set, all aspects of each vehicle were presented; 16 images in each of the first six sets; 24 in the last two sets. All imagery was night, white-hot.

Both SC and Veh ID tests were used for transfer. Soldiers who trained on near ranges were tested at far; those trained on far were tested at near. Each test had 24 transfer images, oblique aspects only, and 12 no transfer images, front and rear views.

Soldiers (*N* = 67), from a mechanized Infantry unit, were randomly assigned to the Near and Far practice conditions. Sixty percent had three or fewer years in service.

The primary questions were:

- Is the amount of training required to identify vehicles at near range the same or different from the amount of training required to identify vehicles at farther ranges?
- Does skill in identifying vehicles at near ranges transfer to farther ranges? Does skill in identifying vehicles at far ranges transfer to nearer ranges?

Secondary questions were the same as in Experiments 1 and 2.

RESULTS

General Effectiveness: Pretests and Posttests

Pre- and posttest scores assessed the general effectiveness of the program, as they generalized across the soldier samples and the experiments. Three effects were replicated. First, FLIR scores increased only for the vehicles that were trained, a mean of of 34%. Visible scores also increased for the same vehicles, a mean of 22%. For vehicles that were not trained, FLIR scores increased 6%; visible, 3% (see the interactions in Figure 2). FLIR and visible scores are also depicted in Figure 2 to point-out the similarity in these scores.

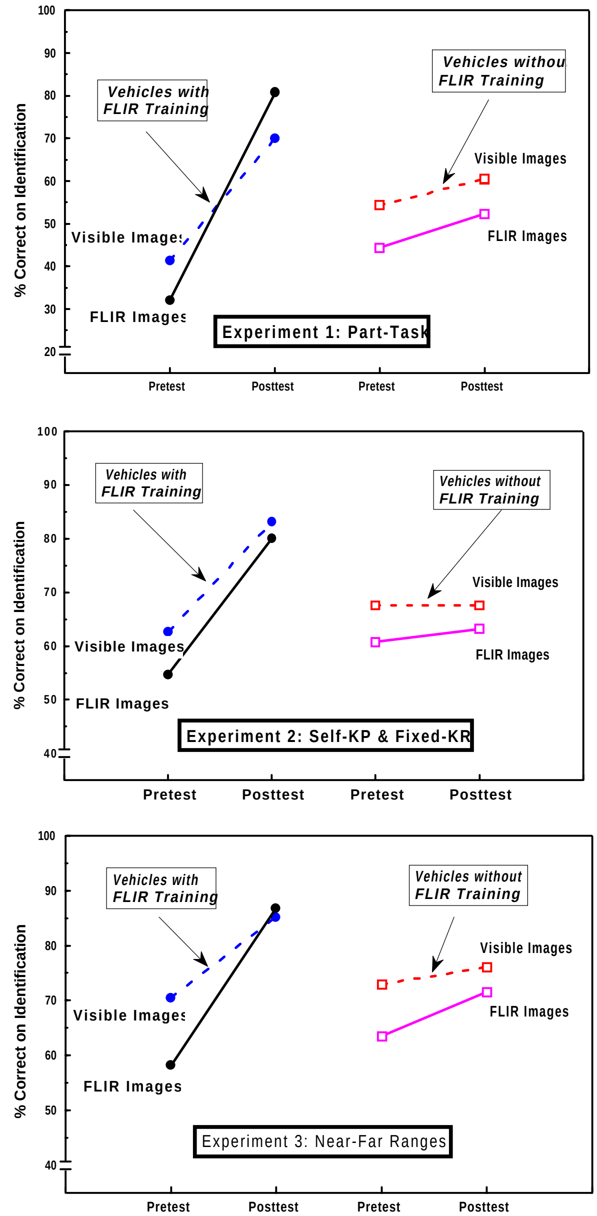

Figure 2. Changes from pre- to posttest for vehicles trained and vehicles not trained. (Exp1 $F(1,15) = 37.06$, $p < .0000$. Exp2: $F(1,31) = 31.91$, $p < .0000$. Exp3: $F(1,63) = 61.72$, $p < .0000$.)

A second finding was that the slowest response times were to the FLIR images on the pretest. On the posttest in Experiments 2 and 3, the posttest times were twice as fast, decreasing from 15 seconds to 7.7 sec. Response time to the visible images decreased by only 1 sec to a mean of 6 sec. Even though reaction time to the FLIR images did not equal that to the visible, the findings reflect a greater confidence in interpreting and understanding the FLIR imagery after training.

The third finding was that, although scores increased with training, the rank order of soldiers remained relatively similar from pre- to posttest (see Table 2).

Table 2. Pretest and Posttest Correlations

	Pre-FLIR	Post-Vis	Post-FLIR
Pre-Vis	.91	.70	.49
	.93	.90	.88
	.86	.83	.70
Pre-FLIR		.73	.49
		.86	.84
		.70	.85
Post-VIS			.76
			.85
			.85

Note. First entry is Experiment 1, then 2, then 3. All correlations significant at $p < .05$.

Experiment 1 Results: Part-Task

Soldiers in the Diad condition required significantly fewer trials, 40% less, to reach criterion on all part-task vehicle sets than those in the Triad condition, $F(1,22) = 8.27$, $p<.0084$. However, identification scores on the first training session of the all-vehicle set did not differ for the Diad and Triad conditions. For both, cross-set confusions and interference occurred when all vehicles were presented. Vehicle scores ranged from 44% correct for the ZSU to 78% for the M1, despite soldiers reaching the criteria of 79% and 85% on the prior part-task vehicle sets. When examining vehicle difficulty for all soldiers and mean time to respond to each vehicle, the fastest responses were associated with the highest scores, the slowest responses with the lowest scores, $r = -.99$. Soldiers responded faster when they knew a vehicle.

There were large individual differences in the number of sessions required to reach criterion on the part-task training sets. The maximum number of sessions ranged from 3 to 10 with a mean of 7 per set. The correlation between scores on the initial session of a vehicle set and number of sessions to reach criterion averaged -.68. These findings suggest that practice without remediation was not effective for those with initial minimal skills.

Unfortunately, the number of sessions required to reach criterion on the all-vehicle set and the ability to transfer skills could not be examined because of unexpected computer problems and time constraints. Training for 14 of the 24 soldiers had to be terminated early during the experiment so they could complete the posttests. Only four completed the transfer test.

Experiment 2 Results: Self-KP and Fixed-KR

Overall, the findings supported the conclusion that the self-paced training with knowledge of performance, via visual corrective feedback, was a better training strategy than the fixed-pace training with knowledge of results feedback. Yet, the part-task training did benefit soldiers in the Fixed-KR condition when they began the all-vehicle set. On the first session for the all-vehicle set, Fixed-KR scores (82%) were significantly higher than the Self-KP scores (68%), $F(1,30) = 9.56$, $p<.0052$. However, there were no significant differences between the two groups in the number of sessions required to reach criterion on the all-vehicle set. The gate on this set was changed from 91% to 85% because of unanticipated time constraints. For soldiers who had difficulty reaching criterion, those in the Self-KP group tended to have a more steady increase in performance with time, despite having lower pretest FLIR scores and lower initial scores on the set than the Fixed-KR group. This may be because soldiers were able to pinpoint their confusions via the corrective visual feedback displays.

On transfer, those with Self-KP training scored significantly higher on vehicle identification than those with Fixed-KR training, 88% vs 79%, $F(1,22) = 5.99$, $p<.0227$. No differences occurred for the no-transfer conditions. Practice in determining aspect via the Veh ID training exercises (Self-KP) resulted in significantly higher aspect scores on the transfer test as compared to soldiers (Fixed-KR with SC) who did not have the benefit of such training, 84% vs 71%, $F(1,22) = 12.00$, $p<.0022$.

A significant interaction with the training conditions occurred on the pre-posttest scores. FLIR scores for the Self-KP group increased more (18%) than FLIR scores for the Fixed-KR group (5%), $F(1,21) = 5.15$, $p<.0337$.

For both training conditions, identification scores were significantly higher for night, black-hot imagery than for day, black-hot imagery, 88% vs 79%, $F(1,22) = 9.07$, $p<.0064$. Higher scores for the night, black-hot images may reflect that transfer was to a different polarity only. With the day, black-hot imagery, the hot spots (thermal cues) also changed. For example, with the sun's heat, the shape of the turret is typically more evident and components of the suspension system often appear cooler.

Finally, the need to provide a means of exiting a training exercise when a learning plateau is reached was demonstrated with both training strategies. Some soldiers had particular difficulty in reaching the 91% criterion (see Figure 3).

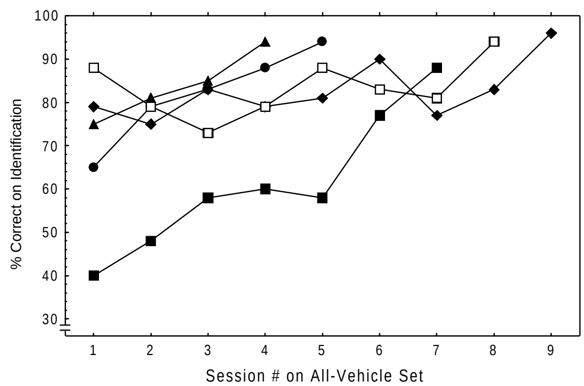

Figure 3. Learning curves for five soldiers on the all-vehicle set in Experiment 2.

Experiment 3 Results: Near-Far Ranges

During the initial Veh ID training exercise, where only the cardinal aspects were presented, the soldiers exposed to Near imagery scored higher (78%) than those exposed to Far imagery (65%), $F(1,54) = 14.71$, $p<.0003$. Identification scores for flank views were higher (83%) than for the front and rear (60%), $F(3,162) = 44.98$, $p<.0000$. Front and rear aspects were confused with each other, e.g., the T72 and M1, BMP with T72 and M2, LAV and BTR. Also, when the front and rear aspects were presented, great disparity in scores occurred among the vehicles, ranging from 27% to 83% correct. Another index of the difficulty of the far images was time to respond. Soldiers averaged 23 sec for near images; 32 sec for far images.

During SC training, soldiers took twice as many sessions per vehicle set to master vehicles at the Far versus the Near range band, $\chi^2(2) = 25.33$, $p< .0000$ (see Figure 4). Also of interest is that when the vehicles were regrouped from tanks and APCs to US and nonUS, soldiers easily discriminated the US vehicles, but not the nonUS. It was hypothesized that the similarity among the nonUS vehicles (low profile, relatively cool, similar front and rear views) contributed to confusions among the nonUS vehicles, while the thermal signatures of the US vehicles were more distinct and therefore easier to discriminate.

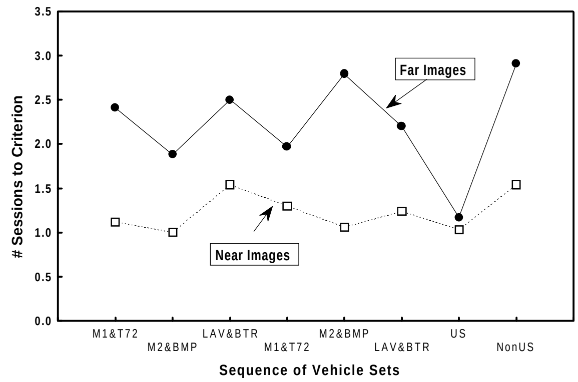

Figure 4. Sessions to criterion for far and near practice conditions in Experiment 3.

Soldiers did transfer skills to another distance when the oblique aspects were presented, but the dominating factor was vehicle distance, not training experience. As expected, the easiest transfer was from the far to the near imagery (85%). More difficult was the transfer from near to far imagery (71%), $F(1,63) = 22.21$, $p<.0000$. But ease of transfer also varied with the vehicle, $F(5,315) = 3.09$, $p<.0096$ (see Figure 5). At the near range, the only vehicle falling below the overall mean of 85% was the BMP. At the far range, there was a greater spread of scores. Identification scores were above 80% for only the M1 and the M2; the BMP's score was the lowest (53%). For the no transfer aspect angles, front and rear, proficiency was maintained better for near than far images.

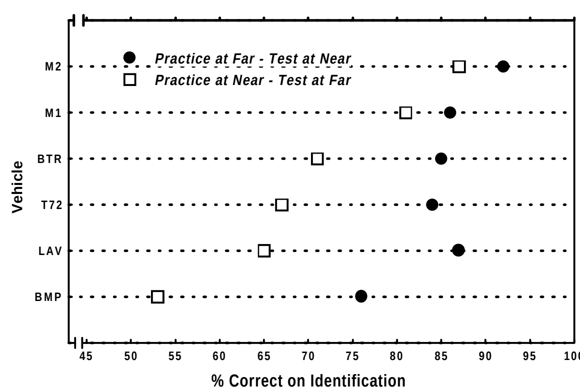

Figure 5. Vehicle scores on transfer as a function of range band presented in Experiment 3.

Although one would expect the no transfer scores to be higher than transfer, this was not the case for the aspects examined. The no transfer scores were significantly lower than the transfer scores; 69% vs 76% for SC, $t(66) = 2.52$, $p<.0141$, and 75% vs 80% for Veh ID, $t(66) = 2.61$, $p<.0111$. These findings reflect the greater difficulty of the front and rear aspects regardless of distance.

Vehicle confusion matrices also showed persistent misidentifications when the front and rear views were presented (e.g., LAV rear confused with M2, BMP front misidentified as T72 and M2; BTR front misidentified as BMP). The test format also influenced both transfer and no transfer scores, with the Veh ID test yielding significantly higher scores than SC: 80% vs 76% for transfer, $F(1,63) = 9.52$, $p<.0030$, and 75% vs 69% for no transfer. $F(1,63) = 9.48$, $p<.0030$. However, the relative order of vehicle difficulty was the same in both tests.

DISCUSSION AND CONCLUSIONS

In summary:

FLIR training improved scores on both the thermal and visible images of vehicles in the training, but not for other, non-trained vehicles.

Faster response times were indicative of greater skill.

Initial individual differences remained, despite substantial improvement in thermal skills.

Vehicle sets with many images led to learning plateaus and learner frustration. Shorter and multiple sets worked better.

Vehicle aspect affected vehicle identification.

Not all transfer conditions were of equal difficulty. Front and rear aspects were hardest, followed by obliques, with flanks the easiest. Images at far distances were harder than those at near. Day imagery was harder than night imagery.

Very high standards, above 90%, were not achieved within a short lesson, particularly for difficult imagery (far ranges, vehicles with similar thermal signatures).

Techniques, such as knowledge of performance with visual feedback displays, helped soldiers understand their confusions, made learning more efficient, and enhanced transfer.

Timed, drill-type exercises seemed to be better suited for "fine-tuning skills" and for soldiers entering the training with considerable expertise.

The automated gating procedure, with increasingly difficult sets of vehicles, was highly motivating to the soldiers. This technique was also essential in making soldiers appreciate the challenge of the vehicle recognition task.

Soldiers did learn to distinguish vehicles at far ranges. But starting training on distant vehicles was not an efficient strategy for those with limited initial skills.

Even after extensive training, some vehicle confusions remained, for certain aspects and at certain ranges.

To ensure expertise in vehicle recognition, training strategies must consider the similarities in thermal signatures regardless of the vehicle class (e.g., tank, APC, logistics).

The computer-based training format adapted well to individual differences. It was more efficient than the group-administered training in prior research.

Hard or persistent discriminations point to the need for special instruction on how to distinguish these vehicles. This instruction could be provided as an option during a training exercise, upon exercise completion, or both. Also, some of these confusions could have been reduced if an instructional module, which systematically pointed out each thermal feature of each vehicle, had been available in the version examined. Additionally, transfer results indicated that a special module on day-thermal cues would be beneficial.

The findings support Lintern's (1989) statement that high levels of skill on part-tasks do not always mean that individuals have developed resistance to interference from additional loads. In each experiment, we found that you cannot assume if soldiers learn to discriminate vehicles A and B and vehicles D and E, that they can automatically discriminate A and B from D and E. Nor will their level of expertise in discriminating the original vehicles be maintained when other vehicles are in the pool. Soldiers must learn to discriminate vehicles from their "lookalikes." A training strategy that deliberately presents mixes of similar vehicles and also considers the most likely confusable vehicle image conditions is needed.

The findings also have implications for test construction. Obviously, scores can be affected by the vehicle pool. With similar images/vehicles, scores are likely to be low. With distinct images/vehicles, scores are likely to be high. Instructors need to be intimately aware of these factors to create fair, yet rigorous tests.

Changes to the program have been made, subject to the resources available. These changes reflect the findings in these experiments as well as soldier feedback from beta sites. They include: (a) creation of vehicle sets with limited numbers of images, (b) vehicle sets that can be programmed in terms of difficulty, (c) feedback summaries for SC and Veh ID tailored to each vehicle, (d) the option of varying the pass criteria to correspond to the soldier's expertise and training progress, (e) addition of an instructional module on vehicle cues, and (f) provision for soldiers to access the Image Library at any point in the program for additional study. The data base has been expanded to 35 vehicles.

Future research should examine other part-task training schedules to determine the most efficient and effective ones. Research is also needed on how to best determine the distinguishing thermal cues of vehicles, on the impact of relating a vehicle's thermal signature to its heat plant and physical structure, and on how cues should be presented within a multi-media context.

Based on comments made by the soldiers in the experiments, the program was a success and was viewed as meeting a critical training need. The computer exercises made vehicle identification learning active, not passive. The highly perceptual nature of the vehicle identification task demands images for training. Words and a few images simply do not describe a thermal signature adequately. The capability of the computer to display multiple images for purposes of comparison, study, and training was critical. The use of actual thermal imagery in conjunction with computer-based, practical exercises that adapt to each soldier's learning rate and can continually challenge soldier expertise is clearly the way to make the difficult task of vehicle recognition more interesting, efficient, and effective.

REFERENCES

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review, 94(2), 115-147.

Cockrell, J. T. (1979). Effective training for target identification under degraded conditions (TP 358). Alexandria, VA: US Army Research Institute.

Gibson, J. J. (Ed). (1947). Motion picture testing and research (Report No. 7). Washington, DC: Army Air Forces, Aviation Psychology Program.

Lintern, G. (1989). The learning strategies program: Concluding remarks. Acta Psychologica, 71, 301-309.

O'Kane, B. L., Biederman, I., Cooper, E. E., & Nystrom, B. (1997). An account of object identification confusions. Journal of Experimental Psychology: Applied. 3, 21-41.

Orentas, R. P., Zegel, F. H., & Gonzalez, J. (1991). Desert Shield IR Guide II. Ft. Belvoir, VA: Center for Night Vision and Electro-Optics.

Palmer, J. E., D'Agostino, J., & Lillie, T. J. (1982). Infrared recognition and target handbook. Ft. Belvoir, VA: Night Vision and Electro-Optics Laboratory.

Proctor, R. W., & Dutta, A. (1995). Skill acquisition and human performance. Thousand Oaks, CA: Sage.

Rollier, R. L., Champion, D. F., Roberson, P. R., & Graber, J. G. (1988). Procedures and guidelines for thermal imaging (RP 88-39). Alexandria, VA: US Army Research Institute.

Warnick, W. L., & Smith, N. D. (1989). Summary of research on combat vehicle identification completed at ARI Fort Hood Field Unit, Fort Hood, Texas, 1980-1995 (RR 1540). Alexandria, VA: US Army Research Institute. (DTIC No. AD-A217 451).

Whitmore, P. G., Cox, J. A., & Friel, D. J. (1968). A classroom method of training aircraft recognition (HumRRO Technical Report 68-1). Washington, DC: George Washington University, Human Resources Research Office.