TRAINING FOR VISUAL INSPECTION: CONTROLLED STUDIES AND FIELD IMPLICATIONS

Colin G. Drury
and
Anand K. Gramopadhye
State University of New York at Buffalo
Department of Industrial Engineering

1.0 ABSTRACT

The three factors most affecting visual inspection performance were derived from a generic task analysis of inspection. For each of these three, possible training interventions were found from the literature on industrial inspection. Direct tests of these interventions were made through five experiments on a computer-based simulator for aircraft visual inspection. One experiment is presented, showing how the size of the area seen by an inspector in a single visual fixation can be trained to improve search performance. Implications of the results of this controlled study for the training of aircraft inspectors are given.

2.0 INTRODUCTION

Training continues to be a major way in which airlines and other operators seek to improve human reliability in aircraft maintenance and inspection. While training systems for complex diagnostic tasks are currently being developed (Johnson, 1990), a parallel effort by the FAA's Office of Aviation Medicine is aimed at improving visual inspection training. This paper uses the generic task description of visual inspection (Drury, et al., 1990) to structure a search for ways in which visual inspection training can be enhanced. Table 1 shows this generic task description of aircraft inspection, where it can be seen that some of the tasks (1. Initiate, 2. Access, 5. Respond) are mainly procedural. The interventions to improve these are less likely to be training interventions than physical changes to job cards, access equipment, recording forms, etc. Even where training is used, there is considerable literature in existence on learning lists of procedures. Of the other tasks, one (6. Repair) is beyond the scope of visual inspection training as it is carried out by mechanics rather than inspectors, while the final task (7. Buy-back Inspect) recapitulates all other steps. This leaves the key tasks requiring training interventions as 3. Search and 4. Decision Making. Both have already been established as potentially the most error-prone in inspection (Drury, 1989). Within the area of Decision Making are two types of decision tasks, one involving reasoning and the other involving perception.

Table 1 Generic Task Description of Incoming Inspection, with Examples from Visual and NDT Inspection

TASK DESCRIPTION	VISUAL EXAMPLE	NDT EXAMPLE
1. Initiate	Get workcard. Read and understand area to be covered.	Get workcard and eddy current equipment. Calibrate.
2. Access	Locate area on aircraft. Get into correct position.	Locate area on aircraft. Position self and equipment.
3. Search	Move eyes across area systematically. Stop if any indication.	Move probe over each rivet head. Stop if any indication.
4. Decision Making	Examine indication against remembered standards, e.g., for dishing or corrosion.	Re-probe while closely watching eddy current trace.
5. Respond	Mark defect. Write up repair sheet or if no defect, return to search.	Mark defect. Write up repair sheet or if no defect, return to search.
6. Repair	Drill out and replace rivet.	Drill out rivet. NDT on rivet hold. Drive out for oversize rivet.
7. Buy-back Inspect	Visually inspect marked area.	Visually inspect marked area.

Training is aimed at reducing both search errors (all misses) and decision errors (misses and false alarms). From a review of the various training interventions available, it becomes apparent that some interventions are better suited to some component tasks. The following section presents part of this review as a research rationale which will lead to specific experimental tests of training interventions. This review covers three areas which are critical to inspection performance: search, decision-making and perception.

3.0 WHAT IS KNOWN ABOUT SEARCH, DECISION AND PERCEPTION TRAINING

3.1 VISUAL SEARCH

When an inspector searches an area for defects, small sub-areas are fixated in turn. The eye remains stationary for about 0.3 seconds during each fixation, at which time an area around the fixation point is processed for visual signs of defects. The area which can be processed in one fixation is known as the Visual Lobe, and is an important determinant of search performance. Larger visual lobes lead to more defects being detected within any given time frame. The visual lobe is measured by repeatedly flashing search areas onto a screen, and asking the inspector whether a target (defect) was present. By placing targets at different points around the center of the fixation, the size of the visual lobe can be mapped out by determining the boundary between those seen and those not seen. However, visual lobe size is not the only aspect of search determining performance. The inspector must have a strategy of how to move the visual lobe across the search area without missing any points, but with efficiency.

Visual lobe training has been studied by Leachtenaver (1978) for photo-interpreters, who found that practice on a search task increased visual lobe size. However, practice on a visual lobe measurement task may also increase lobe size and transfer this increase to search performance.

Search strategy training is an under-represented area in the literature. From the literature it is seen that systematic search is always more efficient than random search, so that a useful assumption is that the searcher is always trying to be systematic. One training objective should be to ensure systematic search, i.e., search in which all areas are fixated, and none are refixated during a single scan. The major difference between systematic and random search is whether or not an area is refixated. The only logical reason for an inspector to refixate an area before a total scan is completed is that the searcher does not remember whether or not that area has been fixated already. Hence, it is seen that it is necessary to provide a memory-aid to the subject to indicate the points of previous fixations to avoid refixations. This could be done by training the inspectors to use feedback from eye movements, either be continuously (on-line) or in a discrete manner at the end of a search task.

Feedback from eye movements can be provided regarding both the number of fixations and the interfixation distance. Literature suggests that these parameters are correlated with an inspector's efficiency in locating possible defects. Providing this sort of feedback would be expected to result in the inspector developing a more efficient search strategy.

3.2 DECISION MAKING - REASONING

Wickens (1984) states that training for decision making can be provided in the following ways:

- Make the decision maker aware of the nature of limitations and biases. Training operators to consider alternative hypotheses might reduce the likelihood of cognitive tunnel vision.
- Provide comprehensive and immediate feedback so that the operators are forced to attend to the degree of success or failure of their rules.
- Capitalize on the natural efforts of humans to seek causal relationships in integrating cues when correlations between variables are known beforehand. Hence, providing information to the operator so as to emphasize the co-relational structure would help in entertaining particular hypotheses.

3.3 DECISION MAKING - PERCEPTION

In perception of fault indications, there are typically many characteristics which must be taken into account. For example, a dent's size, depth, and position in the airframe all affect its criticality. Similarly, corrosion's criticality depends upon its location, severity, and area. The inspector gradually learns what combinations of characteristics make a fault critical, partly from written rules, but also partly from experience. This experience is known in the training literature as the development of a "schema" for that fault.

The schema is a general body of knowledge about the characteristics of a fault that does not contain a strict listing of its defining features (e.g., features which must all be present for a particular instance to be termed a category). According to Posner and Keele (1968) the development of a schema consists of two components:

- a general representation of the prototypical fault, i.e. the basic form from which all the forms are derived;
- abstract representation of the variability around this prototype.

Research in schema formation suggests that the nature of mental representation which people use to classify stimuli into categories is not a strict list of the characteristics of the prototype but also contains information concerning the variability around the template. This is suggested by Posner and Keele (1968) who found that exposure to a variety of instances of a schema induced better performance than repeated exposure to a single instance. Other data shows that assignment is not made by relating each new instance to a central prototype but rather relating it to the previous fault to which it is most similar and then assigning each new instance to the same category as that previous fault.

From the above discussion, it is thus seen that to help in the development of the schema the training provided should be of variable instances of the fault rather than a single instance of a prototypical member, or rules defining the features which would classify the members by criticality. The amount of variability provided in the training should be similar to that existing in the real setting.

4.0 WHAT IS KNOWN ABOUT TRAINING INTERVENTIONS

From the above discussion, training for visual search would be expected to result in reduced search errors (faults missed entirely) and reduced search time. Similarly, training for decision making and perception would be expected to result in a reduction both in misses and false alarms. Although training can be used to improve visual inspection performance, specific training schemes are not associated with factors that determine improvement in visual inspection performance. Hence, *adhoc* training schemes are developed that guarantee improvements for a particular task without consideration whether such a training scheme could be extended to a similar task or a different task, or whether the training is optimizing the use of instructor and trainee time. Hence, the first step in the development of a rational training scheme is to identify the factors that affect visual inspection performance. The next step is to determine which of the functions of the inspection task are trainable. This, in turn, will establish the sensitivity of the inspection parameters to training.

For any training scheme to be effective it should minimize both search errors and decision errors. Thus, referring to the generic task analysis of visual inspection, it is observed that intervention strategies developed at various stages of the inspection process can be hypothesized to change the inspection parameters, resulting in improved performance. In order to improve visual inspection performance, it is necessary to develop training schemes which predict improvements in each of the factors: search, decision making and perception. In the following section various training interventions are briefly described. These will need to be matched to the three factors above in order to derive valid, generalizable interventions.

4.1 VISUAL LOBE TRAINING

The visual lobe is a very important determinant of search performance, with a larger visual lobe require fewer fixations than those with a smaller visual lobe. A large visual lobe or peripheral acuity may account for superior search performance. We still need to know how a large visual lobe can affect search performance and how people can be trained so as to increase the size of the visual lobe. If the above questions are answered, this would then result in astrategy for improving the visual lobe size. The more general question which arises is: how does lobe size training generalize across tasks (e.g., targets and backgrounds). Will the visual lobe training on a given target type result in an improved search performance for a different target type? Thus, it is essential to identify whether such a cross-over effect exists. If it does, then it is sufficient to train the person on one target type. If not, then it is essential to identify various target subsets, say ***T1, T2,*** within which cross-over does occur so that inspectors can be provided visual lobe training on a single target belonging to each target subset.

4.2 FEEDBACK TRAINING

A person needs rapid and accurate feedback in order to correctly classify a defect, or to know the effectiveness of a search strategy. Every training program should begin with frequent feedback and gradually delay this feedback until a level of proficiency has been reached. Additional feedback beyond the end of the training program will help to keep the inspector calibrated. Logically, the following feedback could be provided:

- Feedback regarding the correctness of classifying defective items into categories
- Feedback of search strategy from monitoring eye movements
- Feedback of fixation times from the eye movement search.

The first is known to be essential to learning in perceptual tasks. It provides the novice information regarding the critical difference between a defective item and a perfect item, thus helping to develop a mental template which has the internal characteristics of the defective item. We are, however, still unsure as to what has improved. For example, has learning resulted in a new internal conceptual model of the task (i.e., is the inspector using only certain dimensions of the fault to classify it)?

It has been shown that an important difference between the best and the poorest search performance is the length of the sweeps between eye fixations during a search task. Thus, there exists a difference between how a novice and an expert move their eyes across the visual field. Gould (1973), in a visual inspection study of circuit chips, found that most of the eye fixations occur within a definite boundary, which is the area most likely to contain the targets. It is demonstrated that eye movements in a visual search scenario occur based on knowledge of the location of faults and on the probability of them occurring. The question that needs answering is: does feedback information regarding the eye movements help improve the scanning strategy? If so, then providing such feedback information would aid the inspectors by allowing them to identify areas not covered or areas where they spend excessive time, and helping them develop a strategy to cover the entire area more effectively.

4.3 FEEDFORWARD TRAINING

The novice inspector has little knowledge of the type of faults, probability of faults and occurrence of faults, so that visual search would be expected to be inefficient. Providing feedforward information should result in an improved search strategy because the uncertainty is reduced; by the inspector knowing both where to look and what to look for. The inspector can use the information to achieve a more systematic search strategy, guided by the knowledge of the fault characteristics. Kleiner (1983) suggests that experienced inspectors make use of feedforward information that complements their sensitivity to the fault. If the fault is one that is not easily detected, then the inspector relies heavily on the information provided. Inspection tasks that will most likely benefit from the addition of feedforward information include those in which it is critical for the fault to be detected, those in which the fault is particularly difficult to detect, and those in which the product may contain rare, detrimental and easily overlooked, faults.

4.4 ATTRIBUTE TRAINING

If a fault has, say, four different attributes contributing to its criticality, then the inspector must be trained on each of these attributes, to allow the inspector to set a response criterion for each attribute. The training should be generalizable in the sense that the inspector should be able to classify the items as defective if the items are faulty on one or more of the attributes. The inspector could be trained on which attributes to match against the standard first, based on the probability of the item being faulty on the attributes, and the ease with which the matching occurs. Experience and training of the inspectors determine how defect attributes are arranged in order.

A similar training scheme has been proposed by Salvendy and Seymour (1973) for developing industrial skills. Here, separate parts of the job are taught to criterion, and then successively larger sequences of the job are integrated. Kleiner (1983) used such progressive part training very effectively in inspection.

4.5 SCHEMA TRAINING

It is essential that the subject develop a valid mental template (schema) of the fault. The key to the development of a schema is that it should provide for successful extrapolation to novel situations which are still recognizable instances of the schema. We need to know how schemas are developed, and how aircraft inspectors can be trained to develop schemas. What sort of training (rule based or knowledge based) should be provided to the inspectors for effective development of such schemas?

The effects of two methods of training need to be evaluated in schema development: "active training" and "passive training". In active training, the inspector is presented with various instances of the fault and no- fault, and has to classify them as defective/non-defective. Feedback is provided regarding the correctness of classification. In contrast, passive training is where the inspector is merely presented with various instances of the faults without requiring an active response. Both have been found to be effective, with active training particularly useful for older inspectors.

5.0 TESTING TRAINING INTERVENTIONS FOR VISUAL INSPECTION

As part of a longer-term study of training for visual inspection, a series of tests of intervention strategies was undertaken under controlled conditions. Each test was aimed at determining whether a particular intervention had an impact on improved performance. Because these needed controlled conditions, often with many repetitions of similar faults, actual airframes and inspectors were not logically possible. For example, the hundreds of cracks and dents required for the visual lobe training would never be available to an inspector. Thus, a visual inspection simulator was developed, using a SUN workstation computer to reproduce the essential aspects of the visual inspection task. Four fault types can be generated on a search screen which shows a large array of rivets on the fuselage of an aircraft. The search screen can be moved along the fuselage to search the whole of the pre-defined fuselage section. Also, a single screen can be presented briefly to the subject to allow visual lobe measurements.

The five experiments run on this simulator are as follows:

<u>Experiment 1: Visual Lobe Training.</u> This experiment tests for the possible cross-over effects on the size of visual lobe measured for different fault types.

<u>Experiment 2: Feedback Training.</u> This compares a control group and three feedback groups, having on- line and off-line feedback of both cognitive factors and performance factors.

<u>Experiment 3: Feedforward Training.</u> Again, a control condition is used as a baseline against which to compare rule-based feedforward, knowledge-based feedforward and combined feedforward.

<u>Experiment 4: Attribute Training.</u> Training for decision making using attributes training, i.e. providing the trainee with several levels of severity and complexity, is compared to a control condition where narrative descriptions are provided of the fault attributes.

<u>Experiment 5: Schema Training.</u> Schema development is encouraged by exposing trainees to a wide variety of schema instances (corrosion levels and patterns) in both active and passive schemes.

As these controlled studies were designed to test the feasibility of the training interventions, economical but valid designs were used. In this way, areas which should have an impact on future training schemes for aircraft inspectors could be rapidly assessed. When the results of these experiments are analyzed, the specific implications for airline inspection training can be developed. At this stage, all experiments have been conducted and analyzed, but because of space limitations only the first is described here.

6.0 TYPICAL INTERVENTION TEST: VISUAL LOBE PRACTICE

The objectives of this experiment were to determine the relationship between visual lobe size and search performance, relate changes in lobe size to search performance, and evaluate the effectiveness of lobe training. In particular, the experiment measured whether crossover effects exist in visual lobe training. It used two types of rivet fault (cracks and loose rivets) and two types of area fault (corrosion and dents) to determine whether visual lobe training on one fault would generalize to other faults of the same or different classes.

6.1 METHOD

Twenty-four subjects were used for this study, randomly assigned to four different groups, Rivet Training, Area Training, Neutral Training, No Training. All subjects were tested for 20/20 vision and color blindness.

Rivet Training Group: Subjects assigned to this group initially performed the visual search task on all the four fault types (randomly ordered) followed by visual lobe training on rivet fault: rivet crack. The visual lobe training consisted of performing the visual lobe task five times. The training session was followed by the search task on the four fault types.

Area Training Group: Subjects assigned to this group also initially performed the visual search tasks on all four targets (ordered randomly). They followed this by visual lobe training on one area fault and dent. The visual lobe training consisted of performing the lobe task five times. The training session was followed by the search task on all four fault types.

Neutral Training Group: Subjects assigned to the group performed the visual search task in a similar manner to subjects in the previous groups. However, this was followed by visual lobe training on a neutral target, a computer-generated character. This training session was followed by a similar visual search task.

Control (No Training) Group: Subjects assigned to this group performed similar visual search tasks. However, they did not undergo any visual lobe training. Subjects in this group performed a computer task for a duration equal to the time required for the completion of the visual lobe training session in the other groups. This was followed by the visual search task.

6.2 TASKS

Visual Search Task. The visual search task was the computer-simulated airframe visual inspection task. Subjects had to search for a single fault type in a given area. Visual search performance of the subjects was evaluated on four faults which were classified into two types:

1. Area Faults - 1) corrosion, and 2) dent
2. Rivet Faults - 1) rivet crack, and 2) loose rivets (indicated by streaks of dirt on the rivet edge).

The task was unpaced. During each of the four visual search tasks, the subjects had to search for one of the predefined faults. Subjects were instructed to work as rapidly as possible consistent with accuracy. Subjects verified their response by clicking on the fault with the mouse button. Once a fault was located in a given area, subjects inspected the next area.

Visual Lobe Task. The purpose of the lobe task was to determine the size of the visual lobe; i.e., how far into the periphery a subject could see in a single fixation. The basic procedure consisted of determining at what distance from the central fixation point the target was completely seen by the subject in a single fixation of the fault screen. The exposure duration was kept sufficiently short (0.33 s) to allow the subject a single fixation only. Subjects had to identify a single fault (a rivet fault in the case of the rivet training group, an area fault in the case of the area training group and a neutral fault (across) in the case of the neutral training group). The fault would appear on the horizontal center line of the target screen, at six equally spaced predetermined locations on the horizontal center line, three positions on either side of the central fixation point. No prior information concerning the position of the target was provided to the subjects. The subjects identified the position of the target, either to the left or to the right of the origin and accordingly pressed the key "Q" and "P" to register their response. Subjects were requested to avoid guessing and register responses only if they were sure as to the position of the targets. The fault screen alternated with a fixation screen, consisting of crosswires at the central fixation point exposed for a period of 2 seconds. The purpose of the fixation screen was to help the subjects fixate in the center of the screen after each viewing of the target screen.

6.3 RESULTS

To determine whether the visual lobe increased in size during the training, an ANOVA was conducted for the lobe size for the three groups (1, 2 and 3) receiving lobe training. Over the five training trial significant effects of group ($F(2,15) = 11.05$, $P < 0.0011$), training trial ($F(4,60) = 13.46$, $P < 0.0000$) and their interaction ($F(8,60) = 1.75$, $P < 0.1046$) were found. To test whether the visual lobe training transferred to the visual search task, Analyses of Variance (ANOVAs) were performed on the mean search times for each fault type. These analyzes showed no main effects of groups, but highly significant group X trial interaction. Figure 1 shows these group X trial interactions, where it can be seen that the two faults trained in the visual lobe training had the largest improvement. For the faults not trained by visual lobe training, the improvement was greater where there was more similarity to the visual lobe fault. Neutral training had a smaller amount of transfer, while no training, i.e., spending equivalent time on other computer tasks, had no beneficial effect.

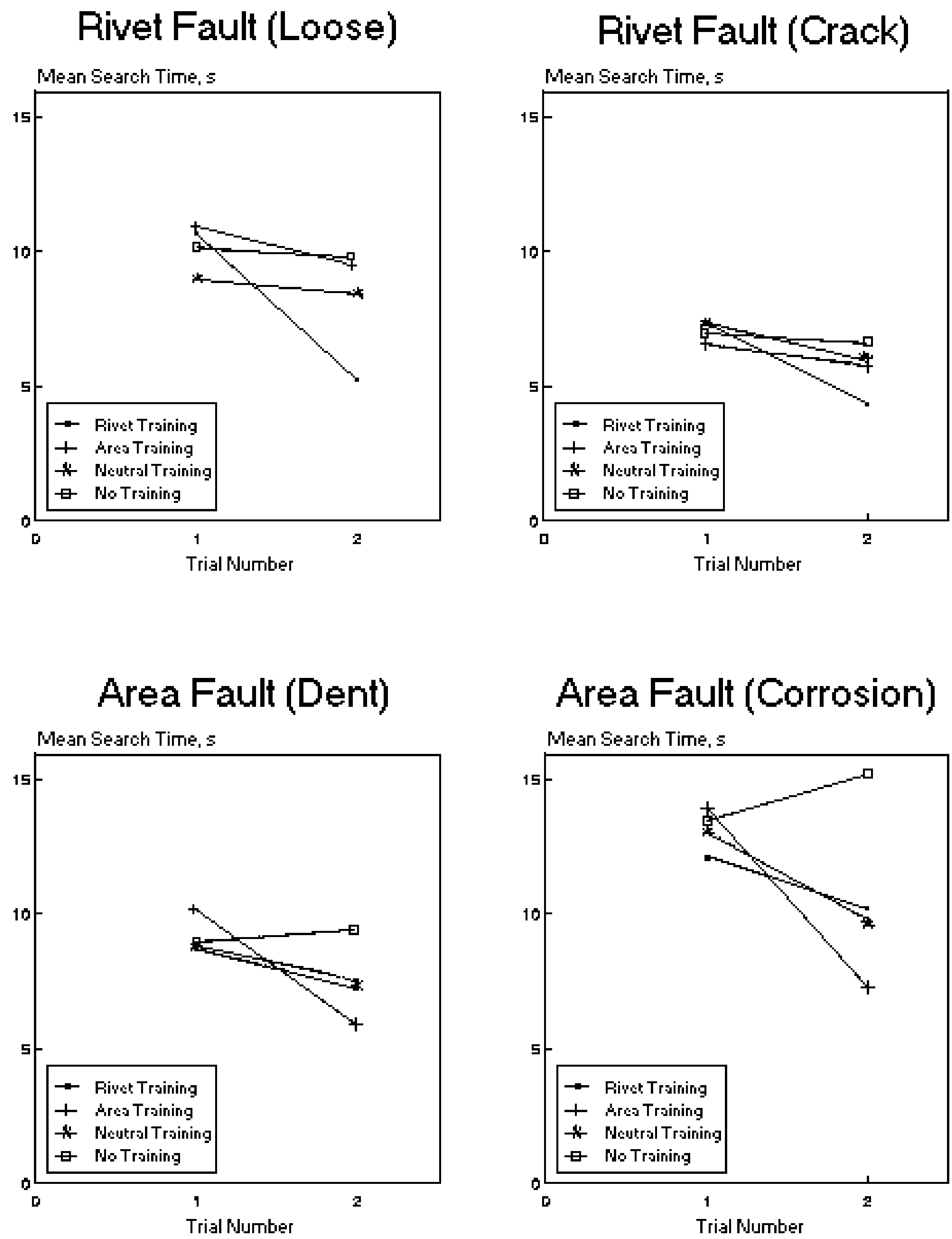

Figure 1 Search Performance Before and After Visual Lobe Training

Similar results can also be seen when the changes in visual lobe size during training are related to the changes in search time after training. Table 2 relates the dependence of search time for each fault type to the increases in lobe size, using the coefficient of determination (r^2) as the measure of dependence. There was a direct transfer from the fault used in visual lobe training to that fault in visual search, with a smaller transfer to the other fault in the same group (rivet or area). The neutral fault visual lobe training transferred only to one area fault.

Table 2 Dependence (r2) of Percent Changes in Search Time on Percent Changes in Visual Lobe Size for Each Group

Percent Increase in Visual Lobe Size For:	Percent Decrease in Search Time For:			
	Loose Rivet	Rivet Crack	Corrosion	Dent
Rivet Training Group	0.75	0.36	0.01	0.21
Area Training Group	0.09	0.00	0.68	0.85
Neutral Training Group	0.16	0.05	0.74	0.00

6.6 DISCUSSION AND CONCLUSIONS

Providing training, even just repeated practice, in rapidly detecting a fault in peripheral vision, does indeed increase the size of the area in which that fault can be detected in a single glimpse, i.e., the visual lobe. This increased visual lobe is not merely a result of increased familiarity with the experimental visual lobe task, as it transfers to a more realistic inspection task, visual search. Thus, even such a basic measure of inspection performance as the visual lobe can be trained to improve. For each fault type there was a 20-30% increase in lobe size over just five practice trials. This transferred to the search task with percentage changes in overall visual search time of:

Rivet Training Group 1 30%

Area Training Group 2 32%

Neutral Training Group 3 18%

No Training Group 4 -4%

There is a close correspondence between the training on actual faults (rivet and area training groups) and improvement in search times, and even some improvement for training on a neutral fault, i.e., one which did not appear in any search tasks. No training, as expected, produced no effect.

From previous visual search studies it is known that visual search follows a speed/accuracy tradeoff curve, so that what has been measured here as search times, can also be interpreted as search accuracy in a given, fixed time. Thus, this experiment has demonstrated the value of training in increasing the inspector's ability to receive and interpret peripheral visual information. The benefits of a simple, simulator- based study in rapidly determining the feasibility of new training techniques has also been demonstrated. A study based on actual faults on a real aircraft structure would have been impossible as single glimpses cannot be repeated without the inspector learning the true identity of each fault. A study using hardware to simulate the faults would be extremely cumbersome, with hundreds of fuselage samples identical apart from fault location being required.

7.0 IMPLICATIONS FOR TRAINING PRACTICE

Each of the experiments run on the simulator has produced positive results, so that more is now known about the basic training interventions for aircraft visual inspection. Successive experiments have shown how the essential factors in visual inspection (visual lobe, search strategy, decision making and perception) can be improved with the appropriate intervention. The next step is to relate these to specific steps which can be taken to improve specific inspector training schemas.

As was noted at the outset, visual inspection is not easy to train. How can the trainer move the trainee's eyes and direct the trainee what to see? In the specific area of visual lobe size improvement, how can the area taken in during one fixation be enlarged? Experiment 1 has shown that if the visual lobe is enlarged, better search performance does result. Thus, lobe size improvements are worth pursuing. It has also been shown that practice in detecting targets away from the direct line of sight in very short time intervals trains people to be alert to their peripheral vision, hence increasing lobe size. The transfer results show that the closer the trained fault is to the actual target characteristics, the greater will be the training benefit.

To take advantage of these results, visual lobe training can be given on a computer system, or even using color slides of actual defects, each defect being exposed very briefly for the inspector. With such practice, the increases in lobe size can be expected to transfer to actual defects on actual aircraft. The more similar the training materials are to the real defects, the greater will be the training benefit, but even visual lobe training on a neutral target provided some improvement in search performance.

Rapid exposure of defects (off-axis) to the inspector is, thus, a valid method of improving the conspicuity of defects, i.e., visual lobe size. This effect should generalize well across many defect types. What is still not known is how to use visual lobe training efficiently. It has been shown to be effective, but how often to give it during training, and how often to retrain are still research objectives, but the basic effectiveness has now been established which makes it worthwhile to proceed with refining the technique.

Experiments 2-5 have been completed and analyzed, with the task now remaining being to interpret their results into practical training interventions in a similar manner.

8.0 REFERENCES

Drury, C. G. (1989). The information environment in aircraft inspection. In *Proceedings of the Second International Conference on Human Factors in Aging Aircraft*, Biotechnology, Inc., Falls Church, VA.

Drury, C.G., Prabhu, P. and Gramopadhye, A. (1990). Task Analysis of Aircraft Inspection Activities: Methods and Findings, In *Proceedings of the Human Factors Society 34th Annual Conference*, Santa Monica, California, 1181-1185.

Gould, J. D. (1973). Eye movements during visual search and memory search. *Journal of Experimental Psychology, 98*, 184-195.

Johnson, W.B. (1990). Application of new technology for aviation maintenance training: an industry status report. In *Proceedings of the Human Factors Society 34th Annual Meeting*, 1171-1175.

Kleiner, B. (1983). *The development of an effective industrial training system.* Unpublished Master's Thesis, State University of New York at Buffalo.

Leachtenaver, J. C. (1978). Peripheral acuity and photointerpretation performance, *Human Factors, 20 (5)*, 537-551.

Posner, M. I. and Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Salvendy, G. and Seymour, J. W. D. (1973). *Prediction and Development of Industrial Work Performance.* New York: John Wiley and Sons.

Wickens, C.D. (1984). *Engineering Psychology and Human Performance.* Scott, Foresman & Co., Glenview, Il.