UNCLASSIFIED

Defense Technical Information Center
Compilation Part Notice

ADP010770

TITLE: Human Factors in Aircraft Inspection

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Aging Aircraft Fleets: Structural and Other Subsystem Aspects [le Vieillissement des flottes d'avions militaires : aspects structures et autres sous-systemes]

To order the complete compilation report, use: ADA390841

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, ect. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP010764 thru ADP010778

UNCLASSIFIED

Human Factors in Aircraft Inspection

Colin G. Drury
State University of New York at Buffalo
Department of Industrial Engineering
342 Bell Hall
Buffalo, NY 14260, USA
716-645-3624, fax 716-645-3302
drury@buffalo.edu

Abstract: Inspection of both airframes and engines is a key activity in maintaining continuing airworthiness. Unless structural defects are detected at the appropriate time, structural failure may result. The reliability of the inspection system must be known in order to schedule safe inspection intervals. However, inspection reliability necessarily includes human inspector reliability so that knowledge of human inspection performance is vital to safety. This paper describes models of the major functions of the human inspector, and applies these within a framework of inspection reliability. From these models, and field experiments on inspectors a set of factors known to affect inspection reliability is derived. These can be used to define good practices necessary to continuously improve inspection performance.

1. Introduction: Inspection plays a critical role in airworthiness assurance. It is used as the detection system for required maintenance procedures and as a final check that the maintenance has been performed correctly. Inspection failure at either stage can compromise public safety. A critical defect may remain undetected and thus unrepaired, or on aircraft with a procedural error (e.g. a missing lock-wire) may be released for service.

These issues have been demonstrated in dramatic fashion in aircraft accidents. In 1988 an Aloha Airlines B-737 aircraft suffered fuselage failure from undetected multi-site damage. In addition to aircraft structures, inspection errors have caused engine failures, for example the JT8-D failure on takeoff on a Delta flight from Pensacola in 1998. In both instances the inspection technique was technically capable of detecting the defect (a crack) but the overall system of technology-plus-human inspector failed. These incidents focused attention on the role of the human inspector in the technology-plus-inspector system.

For many years (see Swain, 1990) human factors engineers had been quantifying human reliability using techniques derived from system safety. Fault tree analysis (FTA) and Failures Modes and Effects Analysis (FMEA) had been employed to determine how failures in the human components of a system affected overall system reliability. This set of techniques was first applied to aircraft inspection by Lock and Strutt (1985), who used their detailed task description of inspection to derive potential systems improvements.

Two parallel lines of research also impact on improving human reliability in inspection. First, for many years it has been traditional to measure inspection system reliability in terms of the probability of detecting defects with specified characteristics under carefully controlled conditions. This set of techniques is used to define the inspection system capability, particularly for non-destructive inspection. The second research thread has been the on-going study of human factors in industrial and medical inspection. Early realization that industrial inspectors were not perfectly reliable led to many hundreds of studies aimed at modeling and improving inspection performance.

This paper covers the modeling and improvement of aviation inspection performance, treating human factors as an explicit aspect of inspection capability. Parts of the text that follow are modified from a recent report on one inspection technique, Fluorescent Penetrant Inspection (FPI), published in Drury (1999).

2. NonDestructive Inspection (NDI) Reliability: Over the past two decades there have been many studies of human reliability in aircraft structural inspection. All of these to date have examined the reliability of Nondestructive Inspection (NDI) techniques, such as eddy current or ultrasonic technologies.

From NDI reliability studies have come human/machine system detection performance data, typically expressed as a Probability of Detection (PoD) curve, e.g. (Rummel, 1998). This curve expresses the reliability

Paper presented at the RTO AVT Lecture Series on "Aging Aircraft Fleets: Structural and Other Subsystem Aspects", held in Sofia, Bulgaria, 13-16 November 2000, and published in RTO EN-015.

of the detection process (PoD) as a function of a variable of structural interest, usually crack length, providing in effect a psychophysical curve as a function of a single parameter. Sophisticated statistical methods (e.g. Hovey and Berens, 1988) have been developed to derive usable PoD curves from relatively sparse data. Because NDI techniques are designed specifically for a single fault type (usually cracks), much of the variance in PoD can be described by just crack length so that the PoD is a realistic reliability measure. It also provides the planning and life management processes with exactly the data required, as structural integrity is largely a function of crack length.

A typical PoD curve has low values for small cracks, a steeply rising section around the crack detection threshold, and level section with a PoD value close to 1.0 at large crack sizes. It is often maintained (e.g. Panhuise, 1989) that the ideal detection system would have a step-function PoD: zero detection below threshold and perfect detection above. In practice, the PoD is a smooth curve, with the 50% detection value representing mean performance and the slope of the curve inversely related to detection variability. The aim is, of course, for a low mean and low variability. In fact, a traditional measure of inspection reliability is the “90/95” point. This is the crack size that will be detected 90% of the time with 95% confidence, and thus is sensitive to both the mean and variability of the PoD curve.

In NDI reliability assessment the model of detecting a signal in noise is one very useful model. Other models of the process exist (Drury, 1992) and have been used in particular circumstances. The signal and noise model assumes that the probability distribution of the detector’s response can be modeled as two similar distributions, one for signal-plus-noise (usually referred to as the signal distribution), and one for noise alone. (This “Signal Detection Theory” has also been used as a model of the human inspector, see Section 3.1). For given signal and noise characteristics, the difficulty of detection will depend upon the amount of overlap between these distributions. If there is no overlap at all, a detector response level can be chosen which completely separates signal from noise. If the actual detector response is less than the criterion or “signal” and if it exceeds criterion, this “criterion” level is used by the inspector to respond “no signal.” For non-overlapping distributions, perfect performance is possible, i.e. all signals receive the response “signal” for 100% defect detection, and all noise signals receive the response “no signal” for 0% false alarms. More typically, the noise and signal distributions overlap, leading to less than perfect performance, i.e. both missed signals and false alarms.

The distance between the two distributions divided by their (assumed equal) standard deviation gives the signal detection theory measure of discriminability. A discriminability of 0 to 2 gives relatively poor reliability while discriminabilities beyond 3 are considered good. The criterion choice determines the balance between misses and false alarms. Setting a low criterion gives very few misses but large numbers of false alarms. A high criterion gives the opposite effect. In fact, a plot of hits (1 – misses) against false alarms gives a curve known as the Relative Operating Characteristic (or ROC) curve which traces the effect of criterion changes for a given discriminability (see Rummell, Hardy and Cooper, 1989).

The NDE Capabilities Data Book (1997) defines inspection outcomes as:

		Flaw Presence	
		Positive	**Negative**
NDE Signal	**Positive**	True Positive No Error	False Positive Type 2 Error
	Negative	False Negative Type 1 Error	True Negative No Error

And defines

$$\text{PoD} = \text{Probability of Detection} = \frac{TruePositives}{TruePositives + FalseNegatives}$$

$$\text{PoFA} = \text{Probability of False Alarm} = \frac{FalsePositives}{TrueNegatives + FalsePositives}$$

The ROC curve traditionally plots PoD against (1 – PoFA). Note that in most inspection tasks, and particularly for engine rotating components, the outcomes have very unequal consequences. A failure to detect (1 – PoD) can lead to engine failure, while a false alarm can lead only to increased costs of needless repeated inspection or needless removal from service.

This background can be applied to any inspection process, and provides the basis of standardized process testing. It is also used as the basis for inspection policy setting throughout aviation. The size of crack reliably detected (e.g. 90/95 criterion), the initial flaw size distribution at manufacture and crack growth rate over time can be combined to determine an interval between inspections which achieves a known balance between inspection cost and probability of component failure.

The PoD and ROC curves differ between different techniques of NDI (including visual inspection) so that the technique specified has a large effect on probability of component failure. The techniques of ROC and PoD analysis can also be applied to changing the inspection configuration, for example the quantitative study of multiple FPI of engine disks by Yang and Donath (1983).

Probability of detection is not just a function of crack size, or even of NDI technique. Early work by Rummel, Rathke, Todd and Mullen (1975) demonstrated that FPI of weld cracks was sensitive to metal treatment after manufacture. The detectable crack size was smaller following a surface etch and smaller still following proof loading of the specimen. This points to the requirement to examine closely <u>all</u> of the steps necessary to inspect an item, and not just those involving the inspector.

3. Human Factor in Inspection: Human factors studies of industrial inspection go back to the 1950's when psychologists attempted to understand and improve this notoriously error-prone activity. From this activity came literature of increasing depth focusing an analysis and modeling of inspection performance, which complemented the quality control literature by showing how defect detection could be improved. Two early books brought much of this accumulated knowledge to practitioners: Harris and Chaney (1969) and Drury and Fox (1975). Much of the practical focus at that time was on enhanced inspection techniques or job aids, while the scientific focus was on application of psychological constructs, such as vigilance and signal detection theory, to modeling of the inspection task.

As a way of providing a relevant context, we use the generic functions which comprise all inspection tasks whether manual, automated or hybrid (Drury, 1992). Table 1 shows these functions, with an example from fluorescent penetrant inspection. We can go further by taking each function and listing its correct outcome, from which we can logically derive the possible errors (Table 2).

Table 1. Generic Task Description of Inspection Applied to Fluorescent Penetrant Inspection.

Function	Description
1. Initiate	All processes up to visual examination of component in reading booth. Get and read workcard. Check part number and serial number. Prepare inspection tools. Check booth lighting. Wait for eyes to adapt to low light level.
2. Access	Position component for inspection. Reposition as needed throughout inspection.
3. Search	Visually scan component to check cleaning adequacy. Carefully scan component using a good strategy. Stop search if an indication is found.
4. Decision	Compare indication to standards for crack. Use re-bleed process to differentiate cracks from other features. Confirm with white light and magnifying loupe.
5. Response	If cleaning is below standard, then return to cleaning. If indication confirmed, then mark extent on component. Complete paperwork procedures and remove component from booth.

Humans can operate at several different levels in each function depending upon the requirements. Thus, in Search, the operator functions as a low-level detector of indications, but also as a high-level cognitive component when choosing and modifying a search pattern. It is this ability which makes humans uniquely useful as self-reprogramming devices, but equally it leads to more error possibilities. As a framework for examining inspection functions at different levels the skills/rules/knowledge classification of Rasmussen

(1983) will be used. Within this system, decisions are made at the lowest possible level, with progression to higher levels only being invoked when no decision is possible at the lower level.

For most of the functions, operation at all levels is possible. Presenting an item for inspection is an almost purely mechanical function, so that only skill-based behavior is appropriate. The response function is also typically skill-based, unless complex diagnosis of the defect is required beyond mere detection and reporting.

Table 2. Generic Function, Outcome, and Error Analysis of Test Inspection.

Function	Outcome	Logical Errors
Initiate	Inspection system functional, correctly calibrated and capable.	1.1 Incorrect equipment 1.2 Non-working equipment 1.3 Incorrect calibration 1.4 Incorrect or inadequate system knowledge
Access	Item (or process) presented to inspection system	2.1 Wrong item presented 2.2 Item mis-presented 2.3 Item damaged by presentation
Search	Individuals of all possible non-conformities detected, located	3.1 Indication missed 3.2 False indication detected 3.3 Indication mis-located 3.4. Indication forgotten before decision
Decision	All individuals located by Search, correctly measured and classified, correct outcome decision reacted	4.1 Indication incorrectly measured/confirmed 4.2 Indication incorrectly classified 4.3 Wrong outcome decision 4.4 Indication not processed
Response	Action specified by outcome decision taken correctly	5.1 Non-conforming action taken on conforming item 5.2 Conforming action taken on non-conforming item

3.1 Critical Functions: search and decision: The functions of search and decision are the most error-prone in general, although for much of NDI, setup can cause its own unique errors. Search and decision have been the subjects of considerable mathematical modeling in the human factors community, with direct relevance to airframe and engine inspection.

In FPI, visual inspection and X-ray inspection, the inspector must move his/her eyes around the item to be inspected to ensure that any defect will eventually appear within an area around the line of sight in which it is possible to have detection. This area, called the visual lobe, varies in size depending upon target and background characteristics, illumination and the individual inspector's peripheral visual acuity. As successive fixations of the visual lobe on different points occur at about three per second, it is possible to determine how many fixations are required for complete coverage of the area to be searched.

Eye movement studies of inspectors show that they do not follow a simple pattern in searching an object. Some tasks have very random appearing search patterns (e.g., circuit boards), whereas others show some systematic search components in addition to this random pattern (e.g., knitwear). However, all who have studied eye movements agree that performance, measured by the probability of detecting an imperfection in a given time, is predictable assuming a random search model. The equation relating probability (p_1) of detection of an imperfection in a time (t) to that time is

$$\mathrm{p}_t = 1 - \exp\left(-\frac{t}{\bar{t}}\right)$$

where $\bar{t}$ is the mean search time. Further, it can be shown that this mean search time can be expressed as

$$\bar{t} = \frac{t_o A}{apn}$$

where

t_o = average time for one fixation

A = area of object searched

a = area of the visual lobe

p = probability that an imperfection will be detected if it is fixated. (This depends on how the lobe (a) is defined. It is often defined such that p = ½. This is an area with a 50% chance of detecting an imperfection.

n = number of imperfections on the object.

From these equations we can deduce that there is speed/accuracy tradeoff (SATO) in visual search, so that if insufficient time is spent in search, defects may be missed. We can also determine what factors affect search performance, and modify them accordingly. Thus the area to be searched (A) is a direct driver of mean search time. Anything we can do to reduce this area, e.g. by instructions about which parts of an object not to search, will help performance. Visual lobe area needs to be maximized to reduce mean search time, or alternatively to increase detection for a given search time. Visual lobe size can be increased by enhancing target background contrast (e.g. using the correct developer in FPI) and by decreasing background clutter (e.g. by more careful cleaning before FPI). It can also be increased by choosing operators with higher peripheral visual acuity (Eriksen,1990) and by training operators specifically in visual search or lobe size improvement (Drury, Prabhu and Gramopadhye,1990). Research has shown that there is little to be gained by reducing the time for each fixation, t_o , as it is not a valid selection criterion, and cannot easily be trained.

The equation given for search performance assumed random search, which is always less efficient than systematic search. Human search strategy has proven to be quite difficult to train, but recently Wang, Lin and Drury (1997) showed that people can be trained to perform more systematic visual search. Also, Gramopadhye, Prabhu and Sharit (1997) showed that particular forms of feedback can make search more systematic.

Decision-making is the second key function in inspection. An inspection decision can have four outcomes, as shown in Table 3. These outcomes have associated probabilities, for example the probability of detection is the fraction of all nonconforming items which are rejected by the inspector shown as p_2 in Table 3.

Table 3. Attributes Inspection Outcomes and Probabilities.

	True State of Item	
Decision of Inspector	**Conforming**	**Nonconforming**
Accept	Correct accept, p_1	Miss, (1 - p_2)
Reject	False alarm, (1 - p_1)	Hit, p_2

Just as the four outcomes of a decision-making inspection can have probabilities associated with them, they can have costs and rewards also: costs for errors and rewards for correct decisions. Table 4 shows a general cost and reward structure, usually called a "payoff matrix," in which rewards are positive and costs negative. A rational economic maximizer would multiply the probabilities of Table 3 by the corresponding payoffs in Table 4 and sum them over the four outcomes to obtain the expected payoff. He or she would then adjust those factors under his or her control. Basically, SDT states that p_1 and p_2 vary in two ways. First, if the inspector and task are kept constant, then as p_1 increases, p_2 decreases, with the balance between p_1 and p_2 together by changing the discriminability for the inspector between acceptable and rejectable objects. p_1 and p_2 can be changed by the inspector. The most often tested set of assumptions comes from a body of knowledge known as the theory of signal detection, or SDT (McNichol, 1972). This theory has been used for numerous studies of inspection, for example, sheet glass, electrical components, and ceramic gas igniters, and has been found to be a useful way of measuring and predicting performance. It can be used in a rather general nonparametric form (preferable) but is often seen in a more restrictive parametric form in earlier papers (Drury and Addison, 1963). McNichol (1972) is a good source for details of both forms.

Table 4. Payoff Matrix for Attributes Inspection.

	True State of Item	
Decision of Inspector	**Conforming**	**Nonconforming**
Accept	A	-b
Reject	-c	D

The objective in improving decision making is to reduce decision errors. There can arise directly from forgetting imperfections or standards in complex inspection tasks or indirectly from making an incorrect judgement about an imperfection's severity with respect to a standard. Ideally, the search process should be designed so as to improve the conspicuity of rejectable imperfections (nonconformities) only, but often the measures taken to improve conspicuity apply equally to nonrejectable imperfections. Reducing decision errors usually reduces to improving the discriminability between imperfection and a standard.

Decision performance can be improved by providing job aids and training which increase the size of the apparent difference between the imperfections and the standard (i.e. increasing discriminability). One example is the provision of limit standards well integrated into the inspector's view of the item inspected. Limit standards change the decision-making task from one of absolute judgement to the more accurate one of comparative judgement. Harris and Chaney (1969) showed that limit standards for solder joints gave a 100% performance improvement in inspector consistency for near-borderline cases.

One area of human decision-making that has received much attention is the vigilance phenomenon. It has been known for half a century that as time on task increases, then the probability of detecting perceptually difficult events decreases. This has been called the vigilance decrement and is a robust phenomenon to demonstrate in the laboratory. Detection performance decreases rapidly over the first 20-30 minutes of a vigilance task, and remains at a lower level as time or task increases. Note that there is not a period of good performance followed by a sudden drop: performance gradually worsens until it reaches a steady low level. Vigilance decrements are worse for rare events, for difficult detection tasks, when no feedback of performance is given, and where the person is in social isolation. All of these factors are present to some extent in FPI, so that prolonged vigilance is potentially important here.

A difficulty arises when this body of knowledge is applied to inspection tasks in practice. There is no guarantee that vigilance tasks are good models of inspection tasks, so that the validity of drawing conclusions about vigilance decrements in inspection must be empirically tested. Unfortunately, the evidence for inspection decrements is largely negative. A few studies, e.g. for chicken carcass inspection (Chapman and Sinclair, 1975) report positive results but most, e.g. eddy current NDI (Spencer and Schurman, 1995; Murgatroyd, Worrall and Waites, 1994) find no vigilance decrement.

It should be noted that inspection is not merely the decision function. The use of models such as signal detection theory to apply to the whole inspection process is misleading in that it ignores the search function. For example, if the search is poor, then many defects will not be located. At the overall level of the inspection task, this means that PoD decreases, but this decrease has nothing to do with setting the wrong decision criteria. Even such devices as ROC curves should only be applied to the decision function of inspection, not to the overall process unless search failure can be ruled out on logical grounds.

4. NDI/Human Factors Links: As noted earlier, human factors has been considered for some time in NDI reliability. This often takes the form of measures of inter-inspector variability (e.g. Herr and Marsh, 1978), or discussion of personnel training and certification (Herr and Marsh, 1978). There have been more systematic applications, such as Lock and Strutt's (1990) classic study from a human reliability perspective, or the initial work on the FAA/Office of Aviation Medicine (AAM) project reported by Drury, Prabhu and Gramopadhye (1990). A logical task breakdown of NDI was used by Webster (1988) to apply human factors data such as vigilance research to NDI reliability. He was able to derive errors at each stage of the process of ultrasonic inspection and thus propose some control strategies.

A more recent example from visual inspection is the Sandia National Laboratories (SNL/AANC) experiment on defect detection on their B-737 test bed (Spencer, Drury and Schurman, 1996). The study used twelve experienced inspectors from major airlines, who were given the task of visually inspecting ten different areas. Nine areas were on AANC's Boeing 737 test bed and one was on the set of simulated fuselage panels containing cracks which had been used for the earlier eddy-current study.

In a final example an analysis was made of inspection errors into search and decision errors (Table 5), using a technique first applied to turbine engine bearing inspection in a manufacturing plant. This analysis enables us to attribute errors to either a search failure (inspector never saw the indication) or decision failure (inspector saw the indication but came to the wrong decision). With such an analysis, a choice of interventions can be made between measures to improve search or (usually different) measures to improve decision. Such an analysis was applied to the eleven inspectors for whom usable tapes were available from the cracked fuselage panels inspection task.

Table 5. Observed NDI errors from classified by their function and cause (Murgatroyd et al, 1994).

Function	Error Type	Etiology/Causes	Miss	False Alarm
3. Search	3.1 Motor failure in probe movement	Not clamping straight edge Mis-clamping straight edge Speed/accuracy tradeoff	X X X	X
	3.2 Fail to search sub-area	Stopped, then restarted at wrong point	X	
	3.3 Fail to observe display	Distracted by outside event Distracted by own secondary task	X X	
	3.4 Fail to perceive signal	Low-amplitude signal	X	
4. Decision	4.1 Fail to re-check area	Does not go back far enough in cluster, missing first defect		
	4.2 Fail to interpret signal correctly	.Marks nonsignals with ? Notes signals but interprets it as noise Mis-classifies signal	 X	X X X
5. Response	5.2 Mark wrong rivet	Marks between 2 fasteners	X	

The results of this analysis are shown in Table 6. Note the relatively consistent, although poor, search performance of the inspectors on these relatively small cracks. In contrast, note the wide variability in decision performance shown in the final two columns. Some inspectors (e.g. B) made many misses and few false alarms. Others (e.g. F) made few or no misses but many or even all false alarms. Two inspectors made perfect decisions (E and G). These results suggest that the search skills of <u>all</u> inspectors need improvement, whereas specific individual inspectors need specific training to improve the two decision measures.

Table 6. Search and decision failure probabilities on simulated fuselage panel inspection (derived from Spencer, Drury and Schurman, 1996).

Inspector	Probability of Search Failure	Probability of Decision Failure (miss)	Probability of Decision Failure (false alarm)
A	0.31	0.27	0.14
B	0.51	0.66	0.11
C	0.47	0.31	0.26
D	0.44	0.07	0.42
E	0.52	0.00	0.00
F	0.40	0.00	1.00
G	0.47	0.00	0.00
H	0.66	0.03	0.84
I	0.64	0.23	0.80
J	0.64	0.07	0.17
K	0.64	0.17	0.22

With linkages between NDI reliability and human factors such as these given above, it is now possible to derive a more detailed methodology for this project.

5. Practical Issues in Inspection Human Factors: As can be seen from the review of human factors in inspection, a number of interventions is derivable from models and field data. Human factors recognizes that any system comprises several components that must work together harmoniously to ensure system performance and human well being. There have been several proposed taxonomies of system components, including ICAO's SHELL model, but here we will use the TOMES model for simplicity: Task/ Operator/ Machine/ Environment/ Social. For detailed reference on each see, for example Drury (1992).

5.1 Task Interventions. The task comprises all of the steps necessary to perform the inspection reliability. Task factors affecting performance include:

- Time available for task completion. Because search is resource-limited, overall probability of detection is very sensitive to time limitations. In particular, external pacing of inspection tasks increases errors.

- Nature of defect. Some detects are inherently more difficult to find than others. In addition, defect size is a major driver of probability of detection. This makes early detection of progressive defects such as cracks and corrosion difficult.

- Mix of defects. If the inspector must search simultaneously for several defects, performance on detecting any particular defect decreases.

- Probability of a defect. As noted under decision models, inspectors have a higher probability of detection where a defect is more likely. Conversely, rare defects are very difficult to detect, providing an ultimate limit to human inspection performance.

5.2 Operator Interventions. The operator here is usually the inspector, although others involved for example in set-up or part cleaning may also be operators.

- Selection and Placement. Historically there has been a continuing interest in providing tools to select a "good" inspector. However, such efforts have been largely unsucessful when "good" is defined in terms of detection probability. A primary reason has been that performance of inspectors is task-dependent, with no guarantee that an inspector who performs well on inspection task A will also perform well on task B.

- Training. Human factors engineers have had considerable success in using the generic inspection functions (Tables 1, 2) as the basis for improved training. Both in manufacturing industry (Kleiner and Drury, 1993) and in aviation maintenance (Gramopadhye, Drury and Sharit, 1997) such training must cover search strategy as well as decision criteria if it is to be effective.

5.3 Machine Interventions. Hardware and software aspects of the task inspection.

- Inspection object handling. If the component inspected is difficult to reach or has poor visual access, inspection performance will suffer to some extent. Access is limited by aircraft and engine design factors, but steps can be taken for improvement. Examples include customized access stands for airframe inspection, easily maneuverable hangers for engine components and improved mirrors/loupes.
- Software aspects of inspection cover the design of documentation such as workcards, manuals and service bulletins. Poor wording and layout of these documents, or their computer equivalents, can have a major effect on error rates (Patel, Prabhu, and Drury, 1992).

5.4 Environment Interventions: All inspection takes place in a physical environment (this section) and a social environment (following section).

- Visual environment. Obviously, enough light must be available for inspection, but performance typically depends more on the quality of the visual environment than the intensity of illumination. Lighting must be developed to maximize the probability of defect detection.
- Other environmental factors. Human performance decreases in adverse noise and thermal environments. For inspectors, such adverse conditions are common, both in line inspection and within the maintenance hangar.

5.5 Social Interventions. Inspection is part of a socio-technical system of aircraft maintenance, so that relationships between the inspector and others will influence inspection performance.

- Management interactions. If inspectors' decis-ions are contradicted by management, then the inspectors are likely to change their decision criteria for reporting defects (see Section 3.1). Most inspectors are fiercely independent, and their departmental managers respect this. But external pressures for hurried work will have obvious effects on inspection reliability.
- Peer interactions. Inspectors hand off work whenever a shift changes or an interruption occurs. The handover procedures have been implicated in incident and accident reports so that good practices need to be followed whenever ownership of a job changes.
- Working hours. Inspection demands continuous vigilance, which is a cognitively demanding activity. People do not perform well during long hours of work. Nor do they perform well when sleep patterns are disrupted. Yet much inspection is carried out on night shifts, and large amounts of overtime are common during initial inspection. Neither practice helps inspection reliability.

6. Conclusions: Airframe and engine inspection is a complex activity dependent upon its human and hardware components alike for its reliability. Human factors engineers have developed useful models of the generic tasks in inspection. Such models can be used both to guide field investigation of inspection tasks and to predict those factors having the greatest impact on inspection reliability. Using this approach it is possible to derive good practices to improve inspection performance. For one unique inspection task, Fluorescent Penetrant Inspection, a set of good practices has been derived and is available at www.hfskyway.com.

7. Acknowledgement: This work was performed under contract from the Office of Aviation Medicine (Ms. Jean Watson), Federal Aviation Administration.

8. References

Chapman, D. E., and Sinclair, M. A. (1975). Ergonomics in inspection tasks in the food industry. In C. G. Drury and J. G. Fox (eds.), *Human Reliability in Quality Control*. London: Taylor and Francis, 231-252.

Drury, C. G. (1999). Human Factors Good Practices in Fluorescent Penetrant Inspection, *Human Factors in Aviation Maintenance - Phase Nine, Progress Report, DOT/FAA/AM-99/xx,* National Technical Information Service, Springfield, VA.

Drury, C. G. (1992). Inspection Performance. In G. Salvendy *(ed.), Handbook of Industrial Engineering*. New York: John Wiley & Sons**,** 2282-2314.

Drury, C. G. and Addison, J. L. (1973). An Industrial Study of the Effects of Feedback and Fault Density on Inspection Performance. *Ergonomics*, **l6,** l59-l69.

Drury, C. G. and Fox, J. G. (1975). *Human Reliability in Quality Control*, London: Taylor & Francis, Ltd.

Drury, C. G. and Prabhu, P. V. (1994). Human factors in test and inspection. In G. Salvendy and W. Karwowski (eds.), *Design of Work and Development of Personnel in Advanced Manufacturing*. New York: John Wiley and Sons, Inc., 355-402.

Drury, C. G. and Sinclair, M. A. (1983). Human and machine performance in an inspection task. *Human Factors,* **25**(4,391-400).

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*, Human Factors and Ergonomics Society, 1181-1185.

Eriksen, C. W. (1990). Attentional search of the visual field. In D. Brogan (ed), *Visual Search,* London: Taylor & Francis**,** 3-19.

Gramopadhye, A. K., Drury, C. G. and Sharit, J. (1997). Feedback strategies for visual search in airframe structural inspection. *Int. Journal of Industrial Ergonomics*, **19(5)**, 333-344.

Harris, D. H. and Chaney, F. B. (1969). *Human Factors in Quality Assurance*. New York, John Wiley and Sons.

Herr, J. C. and Marsh, G. L. (1978). NDT Reliabiity and Human Factors. *Materials Evaluation*. Evanston, IL, American Society for Nondestructive Testing. **36,** 41-46.

Hovey, P. W. and Berens, A. P. (1988) Statistical evaluation of NDE reliability in the aerospace industry. *Review of progress in Quantitative Nondestructive Evaluation, Vol 7B*, Thompson D. O. and Chimenti, D. E. (eds.), Plenum Press, 1761-1768

Kleiner, B.M., and Drury, C.G. (1993). Design and evaluation of an inspection training program. *Applied Ergonomics*, **24(2),** 75-82.

Lock, M. W. B. and Strutt, J. E. (1990). Inspection reliability for transport aircraft structures: a three part study, CAA Paper 90003. London: Civil Aviation Authority.

Lock, M. W. B. and J. E. Strutt, and (1985). *Reliability of In-Service Inspection of Transport Aircraft Structures. CAA Paper 85013*. London, Civil Aviation Authority.

McEwan, W. (1992). NDT Personnel certification and validation. In W. E. Gardner (ed.), *Improving the Effectiveness and Reliability of Non-Destructive Testing*. Oxford, England: Pergamon Press Ltd., 173-192.

McNichol, D. (1972). *A Primer of Signal Detection Theory*, Allen and Unwin.

Murgatroyd, R. A., Worrall, G. M. and Waites, C. (1994). A Study of the Human Factors Influencing the Reliability of Aircraft Inspection, AEA/TSD/0173. Risley, AEA Technology.

NDE Capabilities Data Book (1997) Nondestructive Testing Information Analysis Center.

Patel, S., Prabhu, P., & Drury, C.G. (1992). Design of work control cards. *Proceedings of the Seventh Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection*, Washington, DC: Office of Aviation Medicine, pp. 163-172.

Panhuise, V. E. (1989). Quantitative Nondestructive Evaluation - Introduction. *Metals Handbook*, **17,** 663-665.

Rasmussen, J. (1983). Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions: Systems, Man and Cybernetics*, **SMC-13**(3), 257-267.

Rummel, W. D. (1998). Probability of Detection as a quantitative measure of nondestructive testing end-to-end process capabilities. *Materials Evaluation*, **56**, 29-35.

Rummel, W. D, Hardy, G. L. and Cooper, T. (1989). Applications of NDE reliability to systems. *Metals Handbook*, Ninth Edition, 17, 674-688.

Rummel, W. D., Rathke, R. A., Todd, Jr., P. H. and Mullen, S. J. (1975). The detection of tightly closed flaws by nondestructive testing methods, NASA Lyndon B. Johnson Space Center.

Spencer, F. and Schurman, D. (1995). Reliability Assessment at Airline Inspection Facilities. Volume III: Results of an Eddy Current Inspection Reliability Experiment, DOT/FAA/CT-92/12. Atlantic City: FAA Technical Center.

Spencer, F. W., Drury, C. G. and Schurman, D. (1996). *Visual Inspection Research Project Report on Benchmark Inspections.* Washington, D.C.: Federal Aviation Administration Technical Center.

Swain, A. D. (1990). Human reliability analysis: need, status, trends and limitations. *Reliability Engineering and System Safety*. **29:** 301-313.

Wang, M.-J. J., Lin, S.-C. and Drury, C. G. (1997). Training for strategy in visual search. *International Journal of Industrial Engineering*, 101-108.

Webster, C. (1988). Some Individual Psychological Factors Leading to Error in Ultrasonic Testing. *Reliability in Non-Destructive Testing NDT-88, Proceedings of the 27th Annual British Conference on Non-Destructive Testing*, Portsmouth, UK: Pergamon Press Ltd.

Yang, J. N. and Donath, R. C. (1983). *Improving NDE capability through multiple inspection with application to gas turbine engine disks*. Wright-Patterson Air Force Base, OH 45433, Materials Laboratory, Air Force Wright Aeronautical Laboratories.