Chapter 9
Support of the FAA/AANC Visual Inspection Research Program (VIRP)

Colin G. Drury, Ph.D.
State University of New York at Buffalo

9.0 Objective

This project's objective is to provide human factors inspection expertise to support the Visual Inspection Research Program (VIRP). Note: The material in this chapter is the result of a collaborative effort among many organizations and is not solely the work of C. G. Drury, SUNY at Buffalo, or of Galaxy Scientific Corporation.

9.1 Background and Need

Over the past two decades there have been several studies of human reliability in aircraft structural inspection (Rummel, Hardy, & Cooper, 1989; Spencer & Schurman, 1994; and Murgatroyd, Worrall, & Waites, 1994). All of these studies to date have examined the reliability of Non-Destructive Inspection (NDI) techniques, such as eddy-current or ultrasonic technologies. However, over 80% of civil aircraft inspection does not use NDI and is classified as Visual Inspection (Goranson & Rogers, 1983). Both the FAA *(National Aging Aircraft Research Program Plan*, 1993, p. 26, p. 35) and the ATA have recognized the need for equivalent studies of the reliability of visual inspection as a research priority.

Flight safety is dependent upon airframe integrity; for the civil airline fleet, this includes the detection and repair of structural defects as they appear. Data on airframe structural forces, material characteristics, and models of crack growth are used in the Maintenance Steering Group-3 (MSG-3) process to determine safe inspection schedules. This assumes that there are multiple inspection opportunities between the time a crack becomes detectable and the time it compromises safety. This process is, thus, very sensitive to assumptions about crack detectability. For example, overestimation of inspection reliability would lead to longer inspection intervals, compromising safety. Conversely, underestimation of inspection reliability would lead to shorter intervals, increasing costs because of unnecessary inspection.

While there is a need to obtain accurate measures of in-service visual inspection reliability, there is also a parallel need to understand the process of aircraft visual inspection to improve it. There is a large body of literature on visual inspection in the manufacturing industry (e.g., Drury, 1992), and an increasing number of papers applying this to aircraft inspection (e.g., Drury, 1995). However, there are still no on-aircraft studies which quantify the effects of the many variables affecting human factors in visual inspection. Thus, a second major goal of the VIRP is to provide quantitative evaluations of the effectiveness of visual inspection enhancements.

9.2 Definitions

Quantifying visual inspection is inherently more complex than quantifying NDI. Visual inspection uses many senses and is expected to detect many indications beyond cracks. It may be applied to many different structures and surface treatments.

Bobo and Puckett (1994), in the FAA's latest Advisory Circular on Visual Inspection for Aircraft, use the following definition:

Visual Inspection is the process of using the eye, alone or in conjunction with various aids, as the sensing mechanism from which judgments may be made about the condition of a unit to be inspected.

Visual inspection involves using the "eye, alone or with various aids," and also shaking, listening, feeling, and sometimes smelling, the aircraft and its components. Additionally, the process of any inspection can be analyzed as a combination of various functions, the two most important functions are search and decision-making (e.g., Latorella & Drury, 1992).

In visual inspection, a search process uses most of the human body's senses to detect and locate an indication. There is then a secondary process of combining relevant knowledge, sensory input, and pertinent logic to determine if the indication represents a flaw. The inspector must then make a decision whether or not this flaw is sufficiently sensitive to pose a risk to the continued safe operation of the aircraft or aircraft part.

The Visual Inspection Research Program uses the following definition of "Visual Inspection":

Visual inspection is the process of examination and evaluation of systems and components by use of human sensory systems, aided only by mechanical enhancements to sensory input, such as magnifiers, dental picks, stethoscopes, and the like. The visual input to the inspection process may be accompanied by such behaviors as listening, feeling, smelling, shaking, twisting, etc.

Table 9.1 Classification of Indication & Defect Type

CODE		DESCRIPTOR
20		Wear and Tear
30	21	Loose
40	22	Pulled
	23	Bent
	24	Dent
	25	Scratch
	26	Frayed
	27	Leaking
	28	Lighting Hole
	31	Corrosion
	32	Pillowing
	33	Exfoliation
	34	Intergranular
	41	Material Missing
	42	Broken
	43	Crack
	44	Disbond
		Delamination
		Part Missing

In addition to defining the process of visual inspection, definitions of both the types of indications, i.e., potential defects detectable with visual inspection and the structure on which this inspection is practiced, need to be addressed.

The types of indication possible in aircraft structures were derived from findings at The Aging Aircraft Non-Destructive Inspection Center (AANC) and on other documents relating to inspection. A two-level classification scheme was developed; each major heading was given a two-digit number ending in zero. Below this level, individual indication types shared the same first digit with the appropriate major heading. Table 9.1 shows the current version of this scheme, which can be expanded or modified as needed.

To fully characterize an indication on an aircraft, it is necessary to know the type of indication (Table 9.1) and the structure on which it is found. As results of the baseline inspection of the fuselage area of the AANC's Boeing-737 test bed became available, the findings were classified into the two-level scheme shown in Table 9.2 . This table only includes structural items needed in the current research; there are obviously many more structural elements on an aircraft. As with Table 9.1, this classification scheme gives sufficient detail for the test bed used in VIRP, but should be expanded and modified as necessary to better characterize visual inspection tasks.

From the definitions given in this section, the VIRP was able to design representative experimental evaluations.

9.3 Design of the VIRP Experiments

The research team responsible for designing, conducting, and analyzing the VIRP experiments includes personnel from Sandia National Laboratories/AANC, SAIC, AEA (U.K.) as well as State University of New York (SUNY) at Buffalo. To design the experiments, we held working sessions which included airline inspection representatives (through the ATA) and FAA Technical Center representatives. This group met formally on two occasions during 1994 at AANC facility in Albuquerque; the research team performed its detailed design work outside these meetings.

Table 9.2 Classification of Structure for Fuselage Inspection (Only includes items with indications in Baseline Data)

CODE		DESCRIPTOR
10		Skin
20	11	Doubler
30	12	Extension Skin
40	13	Interior Skin
50	14	Bulkhead
	15	Panel
	21	Fasteners
	22	Rivet
	23	Screw
	31	Bolt
	32	Support Structure
	33	Frame
	34	Stringer
	35	Track
	36	Bracket
	37	Web
	41	Mount

42 Clip
51 Other Structure
52 Rod
Strap
Other Material
Seal
Paint

Reliability of NDI for crack detection is typically reported as one or more Probability of Detection (POD) curves, plotted against crack length. As the design progressed, it became obvious to the research team that visual inspection was a multifaceted activity; unlike NDI of cracks, it could not be characterized by a series of performance curves plotted against a single characteristic. While an equivalent curve can be generated for visual inspection for the single defect type of crack, as Table 9.1 shows, it would only give a partial description of inspection performance. Thus the goals of VIRP were defined as follows:

A. *To establish probabilities of detection for a range of different types of visual inspection (cracks, corrosion, wear and tear, and mechanical) for a "typical" aircraft visual inspection.*

B. *To provide quantified "best practice" guidance on improving visual inspection reliability.*

A research program was developed based on these goals. This process has been described fully in the research team's 1994 White Paper on VIRP and is only summarized here.

The VIRP experiments are designed to achieve Goals A and B (above) in a series of experiments. Because of the large number of factors potentially affecting performance, a single experiment cannot economically provide a measure of overall performance and simultaneously quantify the effects of important parameters. Thus, the program was developed as a Benchmark Experiment (Goal A), followed by a series of Follow-On Studies giving parametric measures of various factors of interest (Goal B).

The detailed protocols for the Benchmark Experiment were partly based upon AANC's 1992-94 study of human reliability in eddy current inspection (Spencer, et al., 1994). Because the main vehicle for testing was AANC's high-cycle Boeing-737, that aircraft had to be subjected to a thorough inspection to determine potential indications/defects. This was performed in a Baseline study during 1994, using qualified commercial inspection personnel to perform a D-check package on the fuselage structure. This study's findings were placed into a database that could be accessed either by the job card (workcard) on which the defect was found or by the defect type. This database was used to develop a new set of job cards specific to VIRP, each containing known defects. These job cards were often designed as subsets of the original job cards so as to include specific areas and specific defects of most interest.

To determine the factors to be included in the experimental program's design, the working group (ATA, FAA, and research team) listed factors known or suspected to affect inspection performance under four headings (see Czaja, Drury, & Shealy, 1981):

- **Task**: The actions the inspector performs, for example: which defects are inspected for, the level of inspection, the time constraints, etc.
- **Operator**: Individual characteristics of the inspector, such as visual ability, training, motivation, familiarity with the task.
- **Machine:** Details of the structure inspected and of the tools used, from mirrors and flashlights to layout of the job card.
- **Environment**: The surroundings of the inspection task. This obviously includes visual, thermal, and auditory environments, but can also include restrictiveness of access and even managerial climate.

Based on these considerations, the working group decided that the Benchmark experiment would be concerned primarily with using the factors to ensure that results would be representative of industry practice. The Follow-On experiments would then examine specific factors one or two at a time. In this way, any data obtained in the Follow-On experiments, e.g., new flashlight designs or better training, could be compared directly against the Benchmark study to measure the effectiveness of any changes in inspection "best practice."

9.3.1 Benchmark Study

During the benchmark study, a group of inspectors, who have not seen the test aircraft previously, will be asked to make a visual inspection of specific areas defined by the VIRP jobcards. The benchmark will be set up as a "typical" scenario by controlling key variables. Each inspector will inspect a number of areas of the aircraft in order to assess that inter-inspector reliability. Videotapes of inspectors performing inspection tasks will be made. Following the actual aircraft inspection, each inspector will be interviewed using a structured interview schedule to elicit his or her expert judgments about the factors influencing successful performance. Analysis of the results will include consideration of the types of errors inspectors may make. The outputs of the benchmark study will be as follows:

Quantitative Results

1. probabilities of detection for different flaw/defect types and sizes
2. inter-inspector reliability
3. estimate of the effects of inspector characteristics included in the design (see below)

Use of videotape as a recording medium will allow a classification of whether an unreported defect was due to an inspector not reacting to the defect (search failure), or reacting, but deciding not to report it (decision failure). After this experiment, it will be possible to measure the reliabilities of the search process and of the decision process so that detailed guidance can be given on suitable improvement interventions.

Both factors to be varied in this experiment concern difficulty of the task. Job cards were developed to provide inspection tasks with either high or low physical access difficulty and with high or low visual complexity. Twelve experienced airline inspectors, recruited through the ATA members, will inspect each area of the B-737 test bed over a two-and-a-half-day period (Figure 9.1). They will also inspect a sample of the crack test panels developed for the NDI eddy-current reliability experiment (Figure 9.2) to determine how reliable inspectors are on a highly-controlled, but realistic, task of the aircraft.

Factors to be fixed were chosen so that they would be at the "best practice" level. Thus, only experienced inspectors will be used. Each will use a good standard tool kit (mirror, flashlight, etc.), and the jobcards will be well-designed (Patel, Drury & Lofgren, 1994). The hangar environment is low-noise with minimum distractions, and the support stands are sturdy and of the correct height.

In addition to the primary data of whether or not each inspector detected each defect, secondary data will be available from a video debriefing procedure. This procedure prompts inspectors to describe what they were doing, and why they were doing it, during various inspection procedures. The procedure we will use is called a Retrospective Verbal Protocol (e.g., Ohnemus & Biers, 1993). It provides valuable insight into the cognitive mechanisms of inspection (e.g., Kleiner, Drury, Sharit, & Czaja, 1989). To improve the precision of the experiment and to obtain a greater understanding of individual factors in aircraft visual inspection, a small battery of tests will be given to each subject. These tests, which provide co-variates for later analysis, include visual performance, mechanical comprehension, and field dependence (e.g., Thackray, 1992; Drury, & Wang, 1986).

As of March 1995, a pilot subject has been tested, and the lessons learned were incorporated into the Benchmark Study. Ten test subjects have now been run.

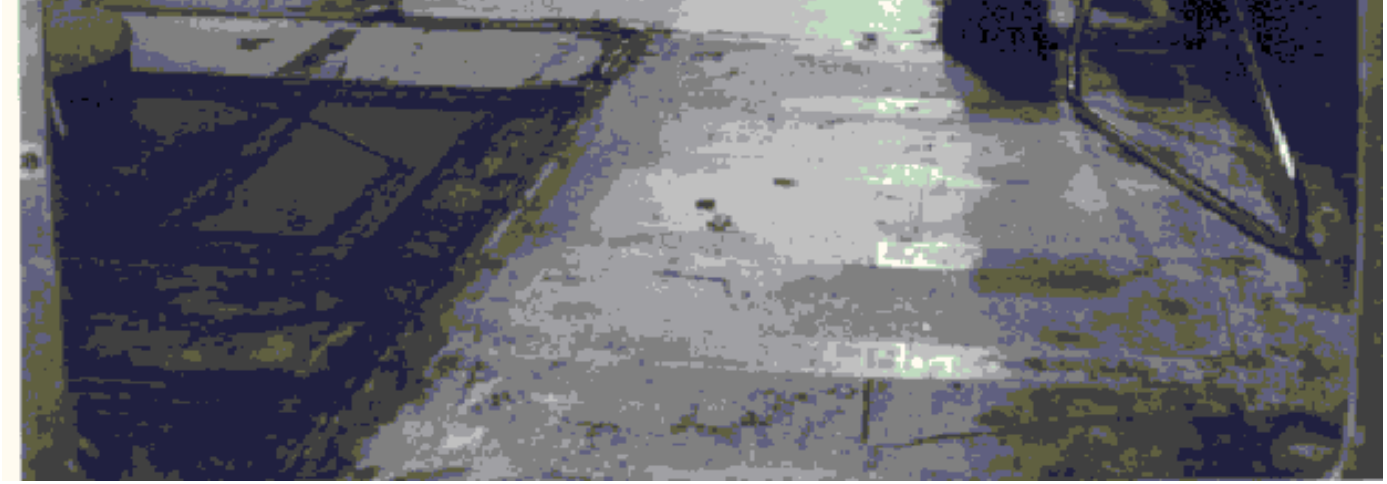

Figure 9.1 Subject Inspecting B-737 Structure

Figure 9.2 Inspector Inspecting Test Panels for Cracks

9.3.2 Follow-On Studies

While a large variety of studies are possible following the benchmark study, only those of most direct benefit to the user community, e.g., to FAA and ATA, will be performed as part of the VIRP. The developed protocols and the characterized B-737 test bed could be used as the basis for specific commercial studies in a manner similar to AANC's continuing work in NDI. No follow-on studies will be finalized until the results of the benchmark study are available; indeed, the design of the follow-on studies is likely to be an ongoing activity of the group as industry and FAA needs are better defined.

In the White Paper produced before the Benchmark Study began, we identified four potential follow-up studies:

1. Effects of fatigue and rest pauses on the detection of flaws

   Objective: To assist in providing guidance on the effective use of rest pauses or other work changes to enhance inspection and to combat the effects of fatigue.

Background: Studies of human reliability in other domains have shown that, with fatigue/time on shift, the performance of experts tends to deteriorate; in extreme instances performance reverts to that of relatively untrained personnel. Studies have also clearly related the ability to detect signals to levels of attentiveness. The negative effects of both of these factors may be controlled with rest pauses. Data from this study could be compared with that from the benchmark study.

2. Perceptual factors

   Objective: To form a basis for guidance on suitable lighting levels, color enhancements, etc., needed to design an appropriate physical environment for visual inspection tasks.

   Background: Visual detection will be influenced by pertinent factors in the physical environment such as contrast, color enhancement, light levels, etc. Job aids such as flashlights, mirrors, etc., will interact with such factors. Aspects such as the color of the inspection surface may affect ease of detection.

3. Search criteria

   Objective: To study the effects of search criteria on the probability of detecting flaws and to assist in the development of guidance on suitable search criteria.

   Background: The ability to detect signals has been shown to be dependent on the search criteria provided, e.g., general versus detailed inspection. Factors such as the number of type of flaws to be searched for may influence the probability of detection of both these and other types of flaws.

4. Decision criteria

   Objective: To study the effects of decision criteria on the probability of detection of flaws and to provide guidance on suitable decision criteria.

   Background: The criteria provided to or assumed by inspectors will influence both the hit/miss and false alarm rates. Criteria may also be affected by the actual or perceived consequences of calling or failing to call a flaw.

9.4 Conclusions

The VIRP is designed to respond directly to industry needs, as expressed through the ATA, and to FAA concerns. Over the first year a test bed has been characterized, protocols developed, and job cards produced so that subsequent studies will benefit in terms of reduced design time and effort. As the Benchmark study is completed and analyzed (Spring, 1995), benefits in data handling and analysis for subsequent studies will also be available. The whole VIRP effort has been unique in the way it has combined knowledge of human inspection behavior, experience of aircraft inspection, and statistical design of experiments. Future experiments will extend the VIRP effort to investigate the effects of inspector fatigue, the visual environment, and for the criteria used by the inspector.

9.5 References

Bobo, S. N. and Puckett, C. H. (1994). *Visual Inspection for Aircraft*, Draft Advisory Circular AC 43, FAA Aging Aircraft Program, Federal Aviation Administration, Atlantic City International Airport, New Jersey.

Czaja, S., Drury, C. G. and Shealy, J. (1981). *Human Factors in Manufacturing*. Unpublished training guide. Buffalo, NY: Applied Ergonomics Group, Inc.

Drury, C. G. (1992). Inspection performance. In G. Salvendy (Ed)., *Handbook of Industrial Engineering, Second Edition, 88*, New York: John Wiley and Sons, 2282-2314.

Drury, C. G. (in press). Human Factors in Aviation Maintenance. In Garland, D. J., Wise, J. A. and Hopkin, V. D. (Eds.), *Aviation Human Factors*, Chapter 25, New York: L. Erlbaum Associates Inc.

Drury, C. G. and Wang, M-J. (l986). Are Research Results in Inspection Tasks Specific? *Proceedings of the Human Factors Society 30th Annual Meeting l986*, 476-480.

Goranson, U. F. and Rogers, J. T. (1983). Elements of Damage Tolerance Verification, *12th Symposium of International Commercial Aeronautical Fatigue*, Toulouse, France.

Kleiner, B. M., Drury, C. G., Sharit, J. and Czaja, S.J. "Evaluating the Effects of Automation on the Human Operator", *Proceedings of the Human Factors Society 33rd Annual Meeting, 1989*, 733-736.

Latorella, K. A. and Drury, C. G. (1992). A framework for human reliability in aircraft inspection. *Proceedings of the Seventh Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection*. Washington, DC: Office of Aviation Medicine, 71-82.

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, U. K.: AEA Technology.

National Aging Aircraft Research Program Plan (1993). Atlanta City, NY: Federal Aviation Administration Technical Center.

Ohnemus, K. R. and Biers, D. W. (1993). "Retrospective versus Concurrent Thinking Out Loud in Usability Testing," 37th Annual Meeting of the Human Factors & Ergonomics Society, Seattle, Washington, October, 1993.

Patel, S., Drury, C. G. and Lofgren, J. (1994). Design of workcards for aircraft inspection, *Applied Ergonomics 1994*, 25(5), 283-293

Rummel, W. D., Hardy, G. L. and Cooper, T. D. (1989). Applications of NDE reliability to systems. *Metals Handbook*, 9th Edition, Volume 17, 674-688.

Spencer, F. and Schurman, D. (1994). *Reliability Assessment at Airline Inspection Facilities, Volume III: Results of an Eddy Current Inspection Reliability Experiment*. DOT/FAA/CT-92/12,III, Atlanta City, NJ: FAA Technical Center.

Spencer, F. and Schurman, D. (1994). Human factors effects in the FAA eddy current inspection reliability experiment. In *Meeting Proceedings of the 8th FAA/OAM Meeting on Human Factors in Aviation Maintenance and Inspection, "Trends and Advances in Aviation Maintenance Operations"*, 16-17 November, Alexandria, VA, 63-74.

Thackray, R. (1992). *Human Factors Evaluation of the Work Environment of Operators Engaged in the Inspection and Repair of Aging Aircraft.* Report No. DOT/FAA/AM-92/3. Washington, D.C.: Federal Aviation Administration.