CHAPTER 71

Human Factors and Automation in Test and Inspection

COLIN G. DRURY
State University of New York at Buffalo

1. DECISION FUNCTIONS IN A GLOBAL BUSINESS ENVIRONMENT

Throughout the world, the business environment is changing, primarily because of globalization. This chapter examines the human factors aspects of one function in industry, test, and inspection to summarize the state of knowledge and practice. Test and inspection are the sense organs of industry; many decisions are based upon their data. They may even be the decision functions themselves when inspection is closely coupled to the control of a business process. Changes in the business environment directly impact the functioning of industry, including in the areas of test and inspection, so these impacts must be considered throughout the chapter. While much of this chapter concentrates on test and inspection in a manufacturing context, there are many other applications, such as in medicine, maintenance, security, and design review, some of which will be considered in Section 7.

Globalization is the combination of market and political forces that has reshaped business and politics since the end of the Cold War in the late 1980s. Some of these forces, such as technological change, are obvious to those who study work, but other forces are changing the nature of jobs, perhaps even more profoundly. Globalization of customers, finance, and production of goods and services has been driven by forces of deregulation, inexpensive transportation, and rapid diffusion of distributed computing (Friedman 1999). Industry is becoming spread across more regions of the world and is shifting away from manufacturing and agriculture towards communications and service. Global capital markets force "creative destruction," the often brutal flow of capital away from enterprises with low shareholder value to enterprises where the capital will generate the greatest return. Investment moves rapidly, forcing industries to respond equally quickly to changing customer demands. We have moved from the managerial capitalism of the first part of this century to investor capitalism with more demanding shareholders (e.g., large pension funds) and more information available instantly (Whitman 1999).

While these changes may seem remote from the lives of human factors professionals, they in fact have very direct effects as they drive industry into new modes of operation beyond technological changes. Perhaps the most important change from the viewpoint of test and inspection is that most workers will be more closely involved with customers because consumers are increasingly demanding a combination of high quality, customization, and low price (Whitman 1999). Exposing the workforce to customers will change the skill requirements of jobs to include communications skills, as has always been the case in service industries such as travel and banking. It also means that the effective batch size is driven towards unity, the individual customer. Quality control in such environments poses a number of challenges because the concepts of continuous control or large-batch sampling no longer apply. Customer involvement can increase job satisfaction but can also increase performance pressure.

Secondly, many operations will be outsourced, leading to reduced levels of job security and more temporary jobs (Whitman 1999). In fact, layoffs in U.S. industry peaked in 1992–1995 at a time of maximum job growth. The fastest-growing category was in temporary jobs, which rose six-fold between 1972 and 1995 (Whitman 1999). The total number of temporary jobs is still small but is a major concern of workers (Kanter 1995, chap. 6). Global competition has forced many companies to downsize their workforces to remain competitive and increase shareholder value. Budros (1999) examines the reasons for downsizing as separate from reorganization. He sees the downsizing trend as being caused by technological innovation and by the existence of highly paid long-term employees, concluding that downsizing is not always effective. From a work perspective, downsizing can be expected to increase workloads for those remaining and to remove some of the company expertise. Work hours overall may be increasing for Americans, according to Schor (1991). She analyzes national data on long-term employment hours of work, vacation time, and work in the home, concluding that total hours of work have increased by about one month per year over the past 40 years or so. While her data and analyses have been questioned, we seem to find very few people who are not working harder than they used to.

Global changes are also driving job demands in ways beyond employment security. Increasingly, work at the world-class levels demanded by global competition generates greater worker skill requirements and a greater rate of worker knowledge obsolescence. Kanter (1999) shows that even in manufacturing, physical assets represented 63% of company capitalization in 1982 but only 38% in 1991. The remainder of the assets are largely composed of company knowledge and competence. Indeed, Siemieniuch and Sinclair (1999) show that even if useful industrial knowledge has a half-life as long as 10 years, only 6% of the knowledge at the start of a working life will be useful at the end 40 years later. Before we retire, we will be producing unknowable offerings (goods and services) with unborn people and uninvented techniques. In turn, this creates a demand for life-long training. Whitman (1999) notes that companies with a heavy emphasis on training show a 19% greater productivity gain over a three-year interval than other companies.

There are other industrial changes taking place at the same time that are not specifically part of the globalization of work. Information technology is becoming a part of ALL jobs as computing power becomes less expensive and more distributed. We not only use information technology to replace workers, but to change the nature of their jobs. While only a few years ago the National Research Council was investigating the gap between IT investment and productivity improvement (NRC 1997), it now appears that the gap has closed and that computing power is having a significant effect on both productivity growth and the nature of work.

As far as the sensing and decision functions of test and inspection are concerned, the factors considered above have large impacts. The pressures for competitiveness have increased the demands on the test and inspection systems for simultaneous improvement of both effectiveness (quality) and efficiency (low cost). In turn, some of these demands have been met by new forms of production. In the past decades, production changes have been driven by influential, although not always effective, movements in industry such as the quality movement and business process re-engineering. Some of the major impacts are described below.

1.1. Increasing Decision Options through IT

Information technology increases the options open to designers while creating a climate in which automation at any price is seen as a virtue per se. Major breakthroughs in vision systems, visual-scene analysis, and decision systems have allowed systems designers to consider replacing human labor in test and inspection with automation. In addition, the increased speeds of production now being experienced and the legal pressure for a return to 100% inspection have meant that in some industries automated systems are the only feasible answer. Throughout this chapter, we will consider the roles of humans and automation within test and inspection as parallel options for allocation of function so that the full impact of information technology can be brought to bear in the most effective manner rather than be a design goal in itself.

1.2. Quality Control/Assurance/Management

A major influence on industry since the 1980s has been the quality movement. The quality revolution has grown over two decades, beginning in quality technology and proceeding through the quality circles movement. It has stabilized over the past 10 years or so under the general titles of total quality control (Hancock et al. 1992) or total quality management (Evans and Lindsay 1993). Quality itself is defined as *meeting or exceeding customer expectations* (fitness for use) or *conformance to specifications* (manufacturing quality). TQM is seen as having both technical and managerial components, in that quality requires both technical knowledge and organizational knowledge.

The most basic quality requirement is freedom from design error, or "off-line quality," as Taguchi and Wu (1979) characterize it. Agricultural products, manufactured goods, and delivered services must meet the needs of their customer over a wide range of customer environments. Products must be functional and must be free from user/product mismatches because any mismatch is by definition a design error. The implications are that designers know their customers' needs and have the techniques to turn these needs into product design. For test and inspection the implications are largely in the design stage, where human factors test and evaluation is an important component of customer service (O'Brien and Charlton 1997).

Secondly, globalized production cannot work unless all elements of the company's global operations fit together without error. Thus, new production systems based on pervasive computer-mediated design and manufacturing can assemble parts from up and down the supply chain without first-time errors of fit. Modern civil and military aircraft production give prime examples of this freedom from physical error. In a customer-oriented system, any over-cost or delayed delivery is an error with the same effects on company performance as a defective product. The implication for test and inspection is that we must be able to measure these customer needs and convert them into measurable precursors of error states.

1.3. Decisions on Quality Made at Source

From TQM and sociotechnical systems design (Taylor and Felten 1993) comes the concept of controlling key variances at their source. One major effect of the quality movement has been to renounce post-production inspection to a large extent and replace it by in-process inspection. With the introduction of more tightly coupled production systems, error cannot be tolerated without widespread consequences, so that control of quality at source is a key component of advanced manufacturing systems such as flexible manufacturing systems (FMS) cells. The corresponding push towards just-in-time (JIT) manufacturing has also made early control of errors a priority.

Control at source means not producing defective products in the first place rather than trying to sort defective products from good products later in the process. With the ultralow defect rates now being demanded and achieved, even highly effective sorting leaves the error rate too high. If we require defective rates in parts per million ($\sim 10^{-6}$) or at a six-sigma standard, then even a hit rate on inspection of 99% will not give the necessary result unless the production defective rate is $\sim 10^{-4}$. Even this level may be difficult with human inspectors because they tend to move their criterion for reporting a defect to more and more stringent values as the overall quality improves. Thus, we look instead for necessary precursors to error rather than error. If we know the mean of a process has shifted by a given amount, then we can deduce that defects will be more common, *even if we find no defects*. This is the principle of in-process quality control, often using control charts (e.g., Vardeman and Jobe 1998).

1.4. Distribution of Test and Inspection Effort

Given the effects of globalization on all enterprises, it should not be surprising that quality is demanded with minimum expenditure of resources. Test and inspection resources are inherently difficult to justify because they can be easily dismissed as not contributing value added to the goods or services produced. They do contribute by providing process control, but not necessarily directly. A product is physically unchanged after test or inspection, except perhaps for the addition of a sticker with the inspector's name on it.

In a broader perspective, production systems design had considered for many years the optimal distribution of inspection effort in control of complex processes. For example, Nurani and Akella (1996) studied the best strategies for sampling semiconductor wafers. A later paper (Shindo et al. 1999) showed that the classification of defects by family allowed the accurate in-line yield prediction required in modern production systems. Such work implies that in designing inspection and test systems, we need to look beyond the individual inspection work point and examine the whole range of defects possible, where they are generated, and how they can best be detected using the technology and models available. This theme will be returned to in Section 7.

2. TEST AND INSPECTION REQUIREMENTS

Test and inspection are decision functions in manufacturing and service industry. They provide decisions concerning the fitness for use of either an item of product or a production process. As such, their prime responsibility is in decision quality, and these functions of test and inspection are often placed in a department with "quality" in its title. Decisions made by the test and inspection subsystem should be:

- **Precise:** Enough depth of information should be incorporated into the decision process so that its conclusions are unbiased by underspecification or rounding errors.
- **Valid:** Decisions should be the same as would be reached after the product or process has been allowed to proceed to actual use.
- **Reliable:** Valid decisions must be possible repeatedly, with different items of product or different states of the process having consistently correct decisions. An inspection or test system should not need frequent recalibration for reliable operation.
- **Robust:** The inspection system must be capable of detecting and classifying a range of defect types large enough to cover customer concerns. This aspect has also been termed "flexibility" (Drury 1992a)
- **Rapid:** Information on a decision is needed rapidly enough for the system to react before many, or even any, defective items have been produced.

Both "test" and "inspection" are from Latin roots, with the former defined by Websters as "to view closely in critical appraisal: look over," and the latter implying "a critical examination, observation, or evaluation" or "a procedure, reaction, or reagent used to identify or characterize a substance or constituent." Precision, depth, and validity are key elements in these definitions, while reliability appears of necessity in a manufacturing context where a sequence of decisions is required. To differentiate between test and inspection, test typically requires a determination of functional suitability for use, while inspection is more usually confined to indications of fitness for purpose short of actual use testing. For example, each new aircraft produced will be flight tested before delivery to determine whether it fulfils its functional requirements. Prior to this flight test, considerable inspection of components and subassemblies will have taken place to ensure that they are free from manufacturing or assembly defects.

2.1. Inspection Measures and Scales

Note that the rather loose term *accuracy* of inspection and test is related to precision, validity, and reliability. To measure the performance of an inspection system, we need to consider how well its decisions match the decisions that should have been made given complete knowledge of the system and the items inspected. First, note that decisions can be of three types, following quality control terminology:

1. Decisions about the fitness of the design of a product. This is off-line quality control, as discussed in Section 1.2. above.
2. Decisions about the fitness of a single item (or group of items) of product. This is the traditional customer protection aspect of any test and inspection system. It is the area traditionally classified as (statistical) quality control, although it can include physical sorting of product items.
3. Decisions about the fitness of a process to continue manufacturing. This is the *jidoka* concept (Monden 1992), aimed at preventing a production system from ever producing a single defective product. A decision not to continue the process does not always imply that individual items produced are unfit (see Section 1.3). Decisions about process fitness are traditionally termed (statistical) process control.

TABLE 1 Definition of Outcome Probabilities

Decision of Test and Inspection System	True State of Conforming	Item (or Process) Nonconforming	Total
Conforming	$p_1 (1 - p')$	$(1 - p_2) p'$	$p_1 + p' (1 - p_1 - p_2)$
Nonconforming	$(1 - p_1)(1 - p')$	$p_2 p'$	$(1 - p_1) - p'(1 - p_1 - p_2)$
Total	$1 - p'$	p'	1.0

Each of these types of decision can only have two final outcomes: fitness and unfitness (of item or process). There are many names for these outcomes:

Conforming/nonconforming
In control/out of control
Good/faulty
Effective/defective

A Note on Measurement Scales. Each decision is a binary one, although subdivisions are possible. For example, an individual item may be conforming or nonconforming (scrap) or nonconforming (rework). Similarly, a process may be fit to continue, or fit to continue with increasing monitoring, or unfit to continue. This brings up the question of level of measurement. Final outcome (above) is always measured on a nominal scale with a discrete number of categories, although some implied ordering may exist. Nominal scales are used directly in test and inspection for the recording of discrete defects on an item—for example, scratches on roller bearings (Drury and Sinclair 1983). Ordered, or ordinal, scales are rarely used as such in test and inspection, although they have occasionally been proposed (e.g., Kelly, 1955). Interval scales, where a continuous measurement is taken, or even ratio scales (which are interval scales with a true zero), are often a part of the measurement process leading to a decision. Thus, the outside diameter of a roller bearing may be measured with high precision and plotted upon a statistical process control chart, although the application of standard rules to this measurement will give a decision outcome about whether the process should continue.

The outcomes of test and inspection can thus be classified by the level of measurement upon which they are based and the type of decision required. These outcomes in turn define the typical statistical methods used, such as prototype testing for decision 1 above. For in-process quality control (decision 3 above), nominal scale decisions lead to attributes control charts, while interval or ratio scales lead to variables control charts (Vardeman and Jobe 1998). The second type of decision above would lead to either attributes of variables sampling plans, although the whole concept of sampling a batch to determine its quality has largely been abandoned in favor of in-process quality control.

Because the final decisions are binary ones, we can define performance in terms of the outcomes for product items (or processes) that were in fact either fit or not fit (conforming or non-conforming). If these probabilities are defined as:

p_1 = probability of deciding that a conforming item or process is conforming
p_2 = probability of deciding that a nonconforming item or process is nonconforming

then the probabilities of the various outcomes are as shown in Table 1, where p' is the true fraction of items (or processes) nonconforming.

In the special case of inspection of items for attributes, these cells have common names, as shown in Table 2.

TABLE 2 Outcomes of Inspection of Items by Attributes

Decision of Test and Inspection System	True State of Conforming	Item (or Process) Nonconforming
Accept	Correct Accept	Miss
Reject	False Alarm	Hit

However, as noted above, in cellular and JIT manufacturing, final outcome should be predicted rather than measured. Thus, statistical process control becomes the technique most used for deciding upon process fitness, with attributes inspection of individual product items playing a progressively smaller role. To move from output control to process control, however, requires detailed, predictive models of all processes. Until these are achieved, the dream of eliminating all final functional test and inspection will perhaps remain a rarely attained goal. In particular, large process changes arising through continuous improvement will mean that predictive models still lag behind the need to deliver product. Thus, the practical strategy to achieve fitness with high reliability would appear to be cyclic changes from item to process control as each process cycles through maturity and periodic replacement. The rapidity with which predictive control is achieved can be expected to increase on successive cycles as parts of predictive models can be expected to be usable across innovation cycles.

3. LOGICAL STRUCTURE OF TEST AND INSPECTION

If we are to proceed towards predictive control of any process, detailed models of the process are needed. This applies equally to the test and inspection system itself. In this section, we develop a classification system for test and inspection systems and generic structure for test and inspection activities. These structures will help guide our more detailed examination of human and automated systems in later sections while showing explicitly the similarities between human and automation. Such similarities are not emphasized, and indeed rarely addressed, in the test and inspection literature.

3.1. Human and Automated Test and Inspection

Logically, no inspection or test system can be either fully automated or fully manual. These extremes are approached, for example with the automated functional tests for integrated circuit chips or the visual examination in proofreading. But there is always some human operator involved in the system, even if only for setting up the system and maintaining it. Equally, an apparently unaided human inspector still uses technology in the workplace, from desks and chairs to flashlights and hand tools.

We can thus define the degree of automation of inspection from one extreme of fully manual, through various levels of hybrid inspection, to the other extreme of fully automated. In this chapter, we will use *manual inspection* as a shorthand for inspection primarily using the human operator's minimally aided senses and minimally aided decision mechanisms. At the other extreme, *automated inspection* will be used to denote an inspection system where the human operator is not involved in on-line sensing or decision functions. Most real systems will involve hybrid inspection, as some functions involve humans and some automation.

The design of systems, such as test and inspection, that can logically incorporate alternative designs (manual or automated) of different system functions has a long history in human factors under the label of allocation of function. This is a primary concern to system designers who must choose which parts of the inspection and test system to automate, although in practice few papers on automated inspection mention function allocation explicitly. The typical automation paper (e.g., Komatsu et al. 1999) list the shortcomings of humans as inspectors and uses this as the justification for developing an automated system to replace the humans in the system. Often the automated system is characterized as being able to detect smaller defects, or automated-system performance is assessed by comparing how well the system classifies a test set of items previously classified by human inspectors. Allocation of function has never had much influence as a formal discipline outside human factors circles (e.g., McCarthy et al. 2000), but it still provides a logical framework for considering a range of automation alternatives. In fact, the allocation of function at the design stage is becoming recognized as perhaps too rigid a design tool, with flexible allocation during operations seen as a more natural and useful approach. For inspection systems this means designing alternate means of fulfilling each objective and having these parallel means available during operations. This requires that we still consider automation alternatives function by function.

3.2. Mission and Function in Test and Inspection

In the classical systems design process (e.g., Singleton 1972), a top-down design starts by defining the system mission, then splits the mission into logical functions. Each function is a unit of system behavior that can be assigned to either human or machine without consideration (at least at first) of other function allocations. Functions can be specified by their goals (or outcomes) in a way analogous to the mission specification of the overall system. In a test or inspection system, we have implicitly defined the mission in Section 2 by defining characteristics of the final system (precision, validity, etc.). Now we must be more explicit about such a definition, both to guide our later evaluation of test and inspection systems (Section 7.1) and to help define the system's logical functions.

The implicit mission in Section 2 concerns determination of functional suitability, so this can become the basis for a more formal definition:

The test and inspection system determines the suitability of a product or process to fulfil its intended function, within given parameters of accuracy, cost, and timeliness.

This definition incorporates both the idea of evaluation of suitability and the notions of predefined system performance. It also forces us to concentrate on the ultimate suitability of the product from the point of view of the customer, so that we do not perform test and inspection activities unrelated to customer needs. Finally, the definition covers both products and processes, so that it can apply equally to output control and process control.

Before we can define the mission for any particular test or inspection system we must be able to specify customer needs. While a detailed framework for designing inspection systems is given in Section 7, we must consider now how to define such needs. One way is to apply a failure modes and effects analysis (FMEA) to the product and design a test and evaluation system to cover each of the potential failure modes. But this technique does not make the customer an explicit part of the design process, whereas we have seen earlier (Section 1) that direct customer input is increasingly needed in more customized products. A preferable technique is to begin with customer function and quality requirements as the basis for a list of product attributes that form the basis of test and inspection. In attributes inspection (Section 2.1), this list is often a ‘‘defect list’’ or ‘‘fault list’’ defining the discrete defects that the inspection system must ensure the customer never experiences.

An explicit approach has been used in a series of studies of paint-inspection systems for automotive painted body panels culminating in Lloyd et al. (2000). Instead of starting with the list of defects traditionally used by process engineers, they sampled many body panels with and without defects and had customers directly rate their dissatisfaction with each defect. This technique produced a list of the most important defects from the customer’s point of view and associated criteria for defects to be below the customer’s threshold of dissatisfaction. They could then design improved inspection systems, for example, using better lighting for human inspection, that reduced warranty data significantly and had a payback time of only six months.

Function lists for inspection have been defined generically by Drury (1978), Sinclair (1984), and Wang and Drury (1989), and for the specific case of inspecting aircraft structures by Drury et al. (1990). Table 3 defines five functions of inspection, with the goal (or outcome) and representative errors for each. Note that the functions are allocation independent in the systems-design sense, in that the functions can be assigned to either human or machine (mechanical, optical, electrical, computer) components. The outcomes and errors remain the same, although the detailed error-causation mechanisms will obviously differ. Thus, error 3.1, ‘‘Indication missed,’’ in the Search function could be caused by the human searcher not detecting a visual indication through inattention or by an automated visual inspection system (AVIS) failing to reach the threshold required for a response to a pixel configuration. The result is identical, although the error mechanism is different.

Because each function can (in principle) be allocated to human or machine, parallel models need to be presented for both types of component. This is done in the next two sections, which provide detail and models for the human and automated components available for building integrated, usually hybrid, systems.

TABLE 3 Generic Function, Outcome, and Error Analysis of Test and Inspection

Function	Outcome	Errors
Setup	Inspection system functional, correctly calibrated, and capable	1.1. Incorrect equipment 1.2. Nonworking equipment 1.3. Incorrect calibration 1.4. Incorrect or inadequate system knowledge
Present	Item (or process) presented to inspection system	2.1. Wrong item presented 2.2. Item mispresented 2.3. Item damaged by presentation
Search	Indications of all possible nonconformities detected, located	3.1. Indication missed 3.2. False indication detected 3.3. Indication mislocated 3.4. Indication forgotten before decision
Decision	All indications located by Search correctly measured and classified and correct outcome decision reached	4.1. Indication incorrectly measured 4.2. Indication incorrectly classified 4.3. Wrong outcome decision 4.4. Indication not processed
Respond	Action specified by outcome taken correctly	5.1. Nonconforming action taken on conforming item 5.2. Conforming action taken on nonconforming item

4. THE HUMAN ROLE IN TEST AND INSPECTION

The framework presented in Table 3 suggests a linear sequence of functions, but in practice there are some branches and reentries in the sequence. For example, an indication of one type may be located, a decision made, and action taken without consideration of any other indications. This occurs when a discrete fault with major consequences is located, such as a missing chip on a circuit board. For simplicity, this treatment will consider only sequential steps.

However, the concept of a one-dimensional sequence is inappropriate for human inspection functions for another reason. 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 being invoked only when no decision is possible at the lower level.

In inspection, two of the functions (2. Present and 5. Respond) have no logical higher-level functions. Note, however, that if test and inspection is regarded as part of a total process-control system, function 5 (Respond) involves considerable higher-level cognitive functioning for diagnosis and prediction of the optimum changes to be made (e.g., Moray et al. 1986; Bainbridge 1990; Umbers 1981). Each function will be considered in turn, but first the terms used in classifying human errors need to be reviewed. System-specific errors have been classified by Hollnagel (1989) as error phenotypes. They are instances of errors generated by an error-generating model or mechanism, which provides the prototypical errors, or error genotypes. Thus, the errors in Table 3 are all phenotypes, but for humans each is an instance of a genotype. Each level of the Rasmussen S/R/K classification has its associated genotypes. Reason (1990) identifies these as skill-based slips, rule-based mistakes, and knowledge-based mistakes. Following Reason (1990), mistakes are choosing the wrong action (or decision), whereas slips are failures to implement the chosen action correctly. Thus, mistakes represent wrong intentions, while slips represent the failure of correct intentions.

With these classifications in mind, the functions will be considered.

4.1. Human Inspection Function by Function

4.1.1. Setup

In setting up a test and inspection system, the measurement devices, decision aids, and recording mechanisms must be procured, checked for functionality, and calibrated. For optical inspection, for example, the correct lighting and magnification levels must be implemented. Calibration consists of checking that the measurement devices respond in the correct manner to known inputs. At the lowest level, these are a sequence of psychomotor activities, equivalent to the items on a pilot's preflight checklist. Job aids consist of checklists, while the tasks themselves are typically closing switches and observing outputs or, more usually now, interacting with computer-based equipment. All of the factors important in good control panel design (Chapter 39) or in human–computer interaction (Chapter 44) become important if these steps are to have a high reliability. Possible slips are those common to any sequential activity, listed by Hollnagel (1989) as:

Wrong place	repetition reversal omission
Wrong time	omission delay premature action
Wrong type	replacement
Not in current plan	insertion intrusion

Note that the errors possible can be mapped onto the set in Table 3 in many different ways, that is, the outcome phenotypes do not necessarily specify the error genotypes.

At the rule-based level, there are typically changes in the setup to accommodate different customers, different products, or different process conditions. Thus, in Drury and Kleiner (1990), a job aid was produced to help inspectors of roller bearings reason more effectively from customer specifications to choice of inspection standards. In aircraft-structure inspection, the calibration of an eddy-current meter must be changed by known rules to accommodate different thicknesses of fuselage skin or detect cracks of different depths. Errors at this level consist of misapplication of rules, particularly reasoning within multiple conditions. Again, principles for designing job aids to assist in rule application are well documented (Johnson 1990).

Knowledge-based reasoning should rarely be required of a setup inspector under normal conditions. It will be invoked when new products are introduced; after process changes due to continuous improvement; occasionally to substitute an alternate inspection device for one that has failed; to troubleshoot and diagnose errors in the inspection device; or even as an innovation on the inspector's part to improve the inspection function. Examples are specifying an increased magnification for detection of smaller severities of visual defects in circuit chip inspection, or increasing the threshold of an eddy current meter to avoid frequent false alarms in lap joint inspection of an aircraft fuselage. Errors can arise from incomplete models of process, product, and/or inspection device or from faulty application of these models. Examples of such knowledge-based reasoning, and of its error genotypes, are more commonly found in the process-control literature (e.g., Moray et al. 1986). System improvement consists of knowledge-based training of the set-up inspector as well as job aids such as schematics and computer programs (e.g., Johnson 1990) to aid understanding of the various components.

4.1.2. Present

Each selected item of product, or test point in the process, must be presented to, and interfaced with, the inspector. Although typically a machine function, it can be given to the human, for example, picking items from a mass-production process and placing each in turn into a gaging device. In more extended inspection tasks such as inspecting aircraft for structural defects (Wenner et al. 1997), the inspector must move to the inspected area rather than the area being moved to the inspector. In such cases, inspectors must move their whole bodies or their limbs to ensure that each area of the item is available to the (visual) sense. The tasks are again psychomotor, typically picking up, orienting, placing, attaching sensors, and disposing. The reliability of such processes is high, with errors due to either misperception of orientation, or to slips in musculoskeletal execution. Models of such activities are widely available (e.g., Knight and Salvendy 1992; Holden 1981). Standard design techniques for preventing misassembly (e.g., the *poka-yoke* concept of Japanese JIT manufacture, Monden 1992) can be used to reduce error rates. Training in manual skills (e.g., Salvendy and Seymour 1973) should not be ignored to improve reliability.

4.1.3. Search

Search is an active process in any inspection context. The item or process inspected must be searched in stages. At times these stages can involve moving an area of limited extent across the item, as in using a flashlight to inspect the inside of an aircraft fuel tank or a microscope's limited field of view to cover the whole area of a microchip. Within a field of view, a target is visible to the inspector only within a limited area, called the visual lobe. This visual lobe must be actively moved in successive fixations across the field of view with saccadic eye movements. Within a single visual lobe, information can only be extracted at a limited rate (Eriksen 1990), so that an attention area can be said to move successively within the visual lobe.

Note that search is a serial sequential process, terminating upon successful detection of a discrete indication (such as a surface scratch) called a target in visual search. Because of its sequential nature, it is a resource-limited rather than data-limited process (Wickens 1991). As such, the probability of detection increases with time spent searching in a manner dependent upon the search plan.

Visual search is so well practiced as to be automatic at the visual lobe and field of view levels, and experienced inspectors' movements of the field of view across the item can also require little cognitive effort. Thus, to a large extent, the search process occurs at the skill level. Aspects of this skill, such as the attention conditions (Eriksen 1990; Engel 1971) and the optical factors affecting the visual lobe (e.g., Overington 1973) have received considerable study so that reasonable models of the probability of detection are possible (e.g., Greening 1975). Errors at the skill level are failure to detect/locate a target and detection/location of a nontarget. However, there are still strategic issues in visual search as evidenced by error phenotypes. Targets that are fixated do go unreported: parts of the item are not fixated at all, while others receive multiple fixations: search terminates too quickly, or continues unproductively.

At the rule-based level, it is the search plan that is altered. Thus, the choice of the next fixation point for the visual lobe is made. This is a decision partly based upon a fixed top-down search plan, but also partly based upon bottom-up information gained during the search. For example, Kundel and LaFollette (1972) show that searchers of medical X rays return a number of times to fixate places where a possible indication was seen. Search is usually modeled as either a fixed systematic process or as a purely random process (e.g., Morawski et al. 1980), although these are simplifications for mathematical convenience. As noted earlier, the probability of detection increases with search time. Typical cumulative search time distributions are shown in Figure 1 for the extreme cases of random search, where each fixation is placed randomly on the field of view, and repeated systematic search,

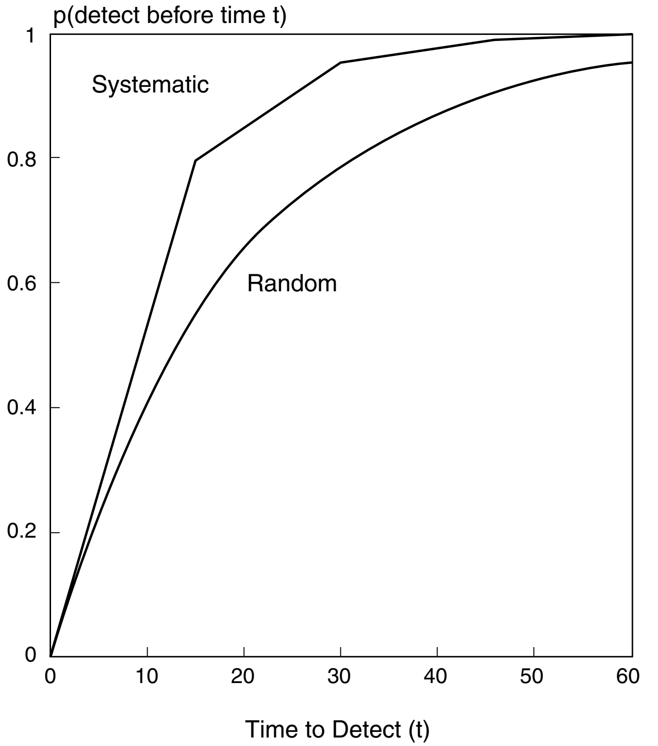

Figure 1 Speed–Accuracy Trade-off Curves for Random and Systematic Visual Search.

where fixations are only placed in unsearched parts of the field of view within each complete scan. Errors at the rule-based level are failing to choose a plan that will lead to the target being fixated.

Knowledge-based functioning in search consists of optimizing choice of parameters for the search. For example, prior information concerning the likelihood of a target being in a particular location will lead to that location being searched first. Choice of how long to spend searching before deciding that no target exists is another optimization aspect of search. There is evidence (Baveja et al. 1996; Chi and Drury 1995) that humans do choose a close-to-optimal stopping policy for search. Errors at this level are stopping too soon, neglecting an area entirely, or the lesser error of failing to detect the target in an optimally short time.

To improve search, the conspicuity of the target in the background must be increased (e.g., by overlays, Chaney and Teel 1967), or the search time increased (e.g., Schoonhard et al. 1973), or the search plan improved by training (e.g., Kundel and LaFollette 1972; Gramopadhye et al. 1997a). If the task requires simultaneous search for more than one type of indication, performance degrades rapidly (e.g., Gallwey and Drury 1986). Recent research (Wenner 1999) on large-scale search tasks, such as inspection of aircraft or searching offices for files, has shown that people need good information support if they are to devise effective and efficient search strategies.

4.1.4. Decision

The output from the search function is either zero, one, or many indications that must be evaluated by the decision process. Logically, if no indications have been found, the item or process must be classified as fit for use. Note that a false alarm is not possible unless at least one indication reaches the decision stage; thus, false alarms are directly diagnostic of decision error rather than search error. Given that decisions will be nontrivial only with an indication, the nature of the decision or indication must be examined.

Decision conditions must always be provided in terms of rules to the inspector, although whether these rules are well formulated depends upon the organization. At times they can be passed on by on-the-job experience from more senior inspectors, while in other instances complex, written, decision procedures exist. Typically, the rules are of one of three types:

Rule 1: IF the magnitude M_i of an indication of type (i) exceeds a severity S_i, THEN item is not fit.

Rule 2: IF the magnitude M_i of an indication of type (i) under circumstances (j) exceeds a severity $S_{i,j}$, THEN item is not fit.

Rule 3: IF the number of indications of type (i) with magnitude M_i exceeding severity S_i exceeds N_i, THEN item is not fit.

The existence of such defining rules would naturally suggest rule-based behavior, but other factors enter into the decision. For extreme indications, such as total absence of an indication or very high severity of indication, the decision will be trivial and essentially skill based. Thus, a missing component in an assembly will "automatically" trigger a rejection response, as will a major crack in a magnesium casting. In the former case, the implication is that

$S_i = 0$
for i = missing component

while the latter case has

$M_i >> S_i$
for i = crack

Most decisions, however, are nontrivial. In inspection, the decision component has been identified with the theory of signal detection (TSD) (e.g., Drury and Addision 1973), although not without necessary warnings about validity (e.g., Megaw 1992). Here, the perceived magnitude of an evidence variable is compared with a perceived criterion, and a decision to reject as unfit made if the evidence variable exceeds the criterion (McNichol 1972). If we modify the notation above with primes indicating perceived variables, we have:

Perceived magnitude for type (i) = M'_i
Perceived criterion (standard) for type (i) = S'_i

TSD states that the decision is based upon $(M'_i - S'_i)$ so that

IF $(M'_i - S'_i) < 0$ THEN accept item
IF $(M'_i - S'_i) > 0$ THEN reject item

The distribution of $(M'_i - S'_i)$ depends upon both the true signal, and the noise in the decision, thus:

$$\begin{aligned}\text{Signal} &= \text{mean }(M'_i - S'_i)\\ &= \text{mean }(M'_i) - \text{mean }(S'_i)\\ \text{Noise} &= \text{var }(M'_i - S'_i)\\ &= \text{var }(M'_i) + \text{var }(S'_i)\end{aligned}$$

To maximize decision reliability, that is, the fraction of total decisions made correctly, then the signal must be as large as possible and the noise as small as possible. This gives a maximum signal-to-noise ratio. In inspection terms, this means magnifying the difference between indication magnitude and standard, for example, using optical devices or computer-enhanced images (e.g., Komorowski et al. 1991) to increase the signal. To reduce the noise, the variability in both the indication and the standard should be minimized. Standard variability can be reduced by having comparison standards available at the workplace (Harris and Chaney 1969). Perceived magnitude variability can be reduced by training (e.g., Gramopadhe et al. 1997b; Drury and Kleiner 1990).

For simple rules, such as rule 1, used in the above exposition, performance is skill based and the only possible errors are:

1. Failing to invoke the decision process
2. Deciding that a conforming item is nonconforming (false alarm)
3. Decision that a nonconforming item is conforming (miss)

All are slip errors, but the first is an example of Hollnagel's sequence errors. Misses and false alarms are classic examples of correct intentions failing.

Rule-based decision making is well supported by training and job aids for complex rules (rules 2 and 3) or for lengthy lists of rules of type 1. Errors can be due to invoking the wrong rule, misapplying the correct rule, or failing to invoke the decision stage. Examples of the first of these errors were given by Drury and Sinclair (1983) in aircraft roller bearing inspection, where there was confusion about defect names among the inspectors, leading to choice of rule appropriate to the wrong type of indication.

In a manufacturing setting, it is rare that the false alarm and miss errors will have equal weights. Thus, in inspection of fuel flow valves for a space shuttle, the consequences of a miss are potentially catastrophic, whereas a false alarm leads only to the delay of component replacement. The weighting of these consequences is an essential component of TSD and represents a form of knowledge-based behavior when consciously applied. An inspector can define an optimum strategy as one that maximizes the expected value of outcomes rather than maximizing the fraction of correct decisions. Expected value depends upon:

1. The costs and values associated with the four decision outcomes, correct accepts, false alarms, misses, hits (V1, V2, V3, V4)
2. The prior probabilities of an item being conforming $(1 - p')$ or non-conforming (p').
3. The criterion (S_i') chosen by the inspector for reaching an accept or reject decision

It is the third of these factors that is under the inspector's control, based upon the perceived values of the other two factors. It is possible to calculate an optimum placement of criteria so as to maximize the expected value across the decision outcomes (McNichol 1972). Experimental evidence from both laboratory studies (Chi and Drury 1995) and field studies (Drury and Addison 1973) shows that inspectors do indeed modify their decision criterion in the direction indicated by the optimum criterion. Changes in chosen criterion are usually less than optimal predictions, an effect dubbed the sluggish beta phenomenon (Wickens 1991).

Clearly, knowledge-based functioning in decision depends upon accurate knowledge of costs, probabilities, and the process of optimization. Misperception of costs and values and misperception of prior defect probabilities can lead to incorrect decisions on criterion placement. Decision support for cost and value perceptions is simple in principle, although it is almost never formalized in an inspection task. Support for estimation of the true defective rate for each type of defect is much more difficult because it is based upon the output of the inspector, or of another inspector who is likely to be equally error prone. It is possible to provide such data to the inspector using another inspector who is allowed more time and/or resources for the decision. When these data are provided as feedforward of likely defect types and probabilities (Sheehan and Drury 1971) or as performance feedback (Drury and Addison 1973), dramatic reductions in error are found.

An alternative way of approaching the decision function in inspection is through the notion of fuzzy sets. Indications of defects such as wear, corrosion, scratches, stains, and quality of cloth have imprecisely defined criteria, and judgments cannot often be represented by a single precise number (Karwowski et al. 1990; Watson et al. 1979). The framework of fuzzy sets (Zadeh 1965) provides us a way of dealing with the category of problems where an absence of sharply defined criteria of class membership is the source of imprecision. Thus, inexact knowledge can be represented, and so can situations where membership in sets cannot be defined purely on an yes/no basis.

Wang et al. (1986) postulate that if the relative contribution of the different cognitive skills towards the performance of a human–machine system can be either known or inferred, then the corresponding subjective assessment of these skills can help in the designing cognitive aids for the human. They developed a fuzzy set approach to formulate a multicriteria decision-making problem to determine whether individuals can prioritize cognitive skills considered important for inspection performance.

A fuzzy set approach was more appropriate than a probabilistic approach because the question was not whether a cognitive factor belongs to a set representing cognitive skills important for inspection, but how strongly it belongs to this set. Thus, the decision alternatives (cognitive factors) for the individuals were imprecise.

The problem was formulated as follows: $X = (X_1, X_2, X_3, X_4)$ constitutes the set of decision-making alternatives corresponding to four cognitive factors. These factors are X_1 = memory, X_2 = attention, X_3 = perception, and X_4 = judgment. The objective was prioritizing the weights associated with these factors and determining the correspondence between these weights and their representation in a multiple-regression model as predictors of inspection performance. The weights are given by

$$FD\ (X_i) = \mathrm{MIN}\ [F_{s,1}\ (X_i), \ldots, F_{s,12}\ (X_i)] \text{ for each } X_i \text{ in } X|S_j$$

where S_j = the fuzzy set of cognitive skills important for inspection performance corresponding to subject j.

$F_{sj}\ (X_i)$ = the membership function associating with each of the four cognitive factors, a number in interval (0, 1) that indites the grade of membership of factor i in S_j

D = the fuzzy subset that results from selecting, for each X_i, the smallest membership value from any of the fuzzy subsets S_1 through S_{12}, under the assumption that the judgments of all subjects are equally reliable.

The membership function was derived using the method proposed by Saaty (1974, 1977), which determined the grade of membership through the process of magnitude estimation derived from pairwise comparisons. The normalized membership for each factor in D was found to be:

$$D = \begin{bmatrix} 0.130, & 0.354, & 0.183, & 0.333 \\ \text{memory} & \text{attention} & \text{perception} & \text{judgment} \end{bmatrix}$$

The priority of importance is seen to be attention, judgment, perception, and memory.

The above study demonstrates an application of fuzzy sets in modeling inspection performance, namely the human's perception of relative importance of various cognitive processes; this could serve as a selection device for inspection tasks. It has to be noted, however, that fuzzy set approaches are applicable in situations where humans must cope with the inexact or imprecise knowledge of the process or system being observed or controlled. It is a descriptive theory that attempts to model the way humans actually cope with a complex problem. Thus, if the space of actions is precise (e.g., inspection of variables such as length, diameter, etc.), then conventional approaches should be followed. Some of the issues that need to be addressed are methodologies for determining membership functions, clarifying the relationship between probability theory and fuzzy set theory, and interpreting of fuzzy utilities (guidelines to interpret the fuzzy advice).

4.1.5. Respond

Actions taken in response to a decision involve both the item itself and a data-capture system. Thus, at the simplest level, defective items can be removed from a production system, or a process can be stopped because it is no longer in control. At the same time, the inspector may need to capture the data in a form usable by the manufacturing system. Again, in a simple form, counters for different defect types are often observed at inspection workstations, with the counter readings recorded at specific times. Another example of a manual system is for aircraft structural inspection (Drury et al. 1990), where an inspector provides details of each defect on a nonroutine repair card.

With modern manufacturing system, evidence of a process or item not being fit to function is expected to be an increasingly rare event, so that action needs to be immediate and data dissemination widespread. Errors in response are, like errors of presentation, likely to be slips rather than mistakes. Actions taken on the basis of the test and inspection response may be complex and require cognitive processing, but the response itself is confined to the skill level. This is not to diminish its importance, as it is the final function of a sequence of processes that can have involved considerable thought and effort. Even if an inspector uses information intelligently to optimize search and decision, the outcome error is just as bad if an item is wrongly tagged or if part of the defect report is omitted. To improve the reliability of the response on the item, principles of workplace design should be followed. Space should be left for rejected items to be stored, and care should be taken to ensure that the response required when an item is rejected does not become so onerous as to discourage a rejection response. Data capture should be efficient and can be simply automated as an alternative to relying on error-prone handwritten forms. One airline company, for example, uses bar codes of all possible faults so that nonroutine repairs (NRRs) can be documented with little written or even keyboard input from the inspector. Portable computer-based systems for aircraft inspection (e.g. Patel et al. 2000) include functionality to allow the inspector to complete NRR forms directly in the computer, ensuring correct identification of the airframe number and the inspector.

4.2. Overarching Considerations: Job Design

Because we are unlikely to reach total automation of the whole manufacturing organization, there will be humans involved somewhere in test and inspection for the foreseeable future, even if only in a supervisory role (Sheridan 1987). Hence, organizational design will continue to impact upon the test and inspection function, requiring at least some familiarity with organizational variables even in highly automated systems.

Many prescriptive models of organizational design exist (e.g., Hackman 1990), but relatively few studies applied specifically to test and inspection. Early work by Jamieson (1966), Thomas and Seaborne (1961), and McKenzie (1958) established that humans change their inspection behavior, and hence performance, in predictable ways when social and organizational variables are changed. Inspection, McKenzie notes, is always of people. The inspector is always judging the work of others, or even his or her own work. Thus, pressures on the inspector are to be expected. More recent work (e.g., Taylor 1991) has examined the sociotechnical systems context of inspection in aviation maintenance and shows that similar pressures still exist. As industry moves from an inspection-based philosophy towards control at the point of manufacture, inspection activities are increasingly a part of all manufacturing jobs. Many years ago (Stok 1965), the visual presentation of information using X-bar and R charts was shown to be highly effective in reducing the production of defects by operators who measured their own production quality.

While the change from item to process control has been taking place, management in manufacturing has also been interested in organizational design for quality, also known as total quality management (TQM) (Evans and Lindsay 1993), as noted in Section 1.2 above. The interaction between such quality initiatives and human factors has been considered elsewhere in some detail (Drury 1996).

Job demands for the operator as inspector and controller logically include an enriched range of tasks. Job enrichment has been shown to be highly effective in improving inspection performance. Thornton and Matthews (1982) found missed defects reduced from 35% to 11%, while false alarms also fell slightly when an enriched inspection job was implemented, with the improvement in performance continuing well beyond the initial change. In a more broad-reaching job-enrichment program, Maher et al. (1970) found a halving of both human errors and time for inspection. To reach the level of knowledge and skill required in those new tasks, operators will need expanded training in process control and inspection procedures (Drury and Kleiner 1990).

The group-oriented aspects of organizational changes have been aimed at empowering a small, self-contained group of operators in the same way that job design empowers the individual (Hackman 1990). Good results have been reported in group inspection work within a maintenance environment (Rogers 1991; Diehl 1990). However, when groups such as manufacturing cells are implemented, the training needs again increase. Drury (1991) notes that operators in such environments require communication and interaction skills as well as group decision-making training to function effectively.

5. AUTOMATION IN TEST AND INSPECTION

Automated inspection and test systems are now finding a regular place in many industrial applications, alongside or replacing human inspection. Their main advantage is that they can perform a relatively limited number of inspection services rapidly and reliably, suiting them to high-volume repetitive test and inspection tasks. In this section, we provide a classification of automated systems based on their functions and give some detailed examples of automated systems in the literature. A typical system (Steinmetz and Delwiche 1993) is first introduced to illustrate many of the points raised.

The Steinmetz and Delwiche (1993) example is a Franco-American design of a system for automatic grading of cut roses. It was chosen because it uses a product of considerable geometric and chromatic complexity and variability but that is instantly recognizible and is likely to be well known to readers from different industries. In fact, a human grading was used to develop the test batches required for performance evaluation of the system. A second consideration was that the paper contains some numerical evaluation of system performance.

The paper begins, as is typical, with making the economic case for automation, that is, that the industry has high throughput and the grading and packaging costs (currently human activities) account for about half of total production costs. Direct cost savings should be possible through automation of the grading process. The cut rose has a set of quality characteristics that form the basis for the current grading system:

1. Stem length
2. Stem diameter
3. Stem straightness
4. Bud maturity (degree of opening)
5. Bud color

First, a holding and lighting system was designed, supporting the rose below the bud on a simple hook and surrounding it with diffuse fluorescent light of appropriate spectral characteristics. Note the similarity to the Setup and Present functions of Section 4. A standard color video camera was used to give a 480×512 pixel image (a low figure by today's standards), giving about 1.4 mm/pixel.

Next, a set of algorithms was devised for the computer to "understand" the image captured by the camera. Because the rose was in a constant position and orientation, defined by the support hook, the stem and bud could be unambiguously differentiated if the coordinates of the hook were known. Stem length was derived from the distance from the hook to the first cross-section of the image containing a nonwhite pixel, starting from the side farthest from the hook. At this point, other stem features could be extracted. The leaves were differentiated from the stem itself by searching column by column along the length of the stem for wide groups of nonwhite pixels. At each end of the stem, an authorized region (AR) was defined as a rectangle, while where leaves were encountered, triangular ARs were defined. These regions were large enough to include thorns, which had also to be differentiated from the stem proper. Logical detection of stem and leaf boundaries was performed using an edge-detection algorithm. Feature recognition made use of the geometric fact that a stem must be a set of rectangular segments of approximately equal widths located in the space defined by the two ends of the stem already found and oriented approximately parallel to the line between the stem end points. Note the similarity to the early stages of the human function of Search, where overall salient features of the item are used to guide the detailed search process in a top-down manner. Stem straightness was defined by the maximum angle between any of the rectangular stem segments, or by the maximum distance between any segment and the line joining the stem end points. Bud maturity measurement started by searching above the hook for an area of nonwhite pixels. This area was recognized by its edges, transformed to a convex shape and the centroid calculated. Maturity was defined as the maximum distance of any bud edge to either side of the centroid. Bud color was determined by averaging separately the red, green and blue chromaticity coordinates of the convex region of the bud.

From this measurement phase, a set of nine measures was available as a vector for each rose. In fact, the measurements were repeated for two orthogonal views of each rose to ensure that straightness was not an artifact of the direction of view. To test the system, two sets of roses were first graded by experienced graders, and the grades were refined in the laboratory using the raw camera images to improve grading consistency. The two sets were a training set and a test set, for use with the later classification evaluation. For the two direct measurements (length and diameter of the stem), standard errors were used to show that the automated system was accurate enough for use. For each of the other characteristics, the training set was used to train a multilevel neural net (or alternatively a Bayes classifier) to reach the correct classification of each rose. After training, the neural net then classified the test set to measure performance. Compare this technique with the Decision function for human inspectors, where training was again found to be useful. For straightness, the neural net classified 21 of 27 straight roses correctly and 28 of 33 crooked roses correctly, for an overall error rate of 18%. Bud maturity was classified in the same way, but onto a three-point scale rather than into two classes. The misclassification rate was between 15% and 21%, depending on rose color sample. Finally, comparison to a simple threshold of the average red chromaticity was used to classify red and white roses with 0% errors. The final human function of Respond was not used in this prototype system.

Recommendations were finally developed for improvement of the system, including a low-pass filter to prevent thorns interfering with diameter measures, removal of lower leaves to make length detection more reliable, and extension of the system to the more difficult task of detecting defects on the leaves.

From this description, the logic of automated systems is clear, as are many parallels with human inspection. The functions can be separated and implemented, performance measured, and recommendations made for future improvement. The functions of automated inspection will now be considered in turn because they involve quite different mechanisms than the equivalent human functions.

5.1. Automated Inspection Function by Function

Although automated inspection must logically fulfill the five functions presented in Section 4, these functions do not represent the most natural breakdown for automated inspection. Some functions are combined in a single piece of hardware, while some automated functions (e.g., sensing) have a major classificatory role. One example of combined roles is that the Present and Search functions can overlap in a CCD readout. Another is that Search and Decision can be part of one computer program rather than separated as they were in the rose-inspection example. A more natural breakdown for automated inspection would be as follows:

1. Setup
2. Materials handling
3. Sensing
4. Signal processing

Figure 2 shows how these two classification schemes map onto each other. The Materials-handling function of automation covers both the Access and Respond functions of our more general scheme, while, as noted above, sensing and signal processing may be difficult to differentiate physically. All are, however, separable logically.

5.1.1. Setup

The setup function in automated inspection corresponds to configuring the system for the task. Several environmental factors affect the setup function. In a machine vision system, lighting factors such as illumination levels, type of illumination, reflectivity, and contrast affect performance. Most vision systems are quite sensitive to these variables. In many cases, the automated system has failed because

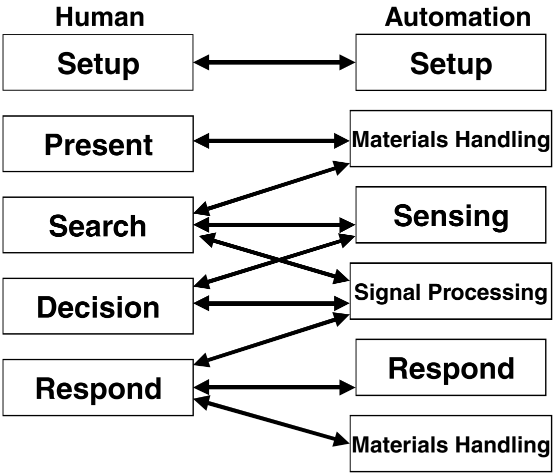

Figure 2 Mapping of Classification Schemes for Manual and Automated Inspection.

the stability of the environment and the effect that environmental factors may have on the measurements have not been correctly understood. In fabric inspection, cloth brightness can change due to environmental conditions that can lead to incorrect detection of dirt on cloth. Thus, we need to have either vision algorithms that can detect defects independent of illumination changes or a tightly constrained environment (e.g., Steinmetz and Delwiche 1993). Lighting intensity is a critical factor and must be sufficient to swamp interference from ambient sources to improve environmental reliability. Similarly, the contrast of the object against its background must be greater than the local lighting variation around the feature that is of interest.

5.1.2. Materials Handling

The materials-handling function for manufacturing test and inspection consists of devices to select items for test, mount them into the test fixture, move them within the fixture (if needed) and transport the item to the next stage. Thus, in the inspection of aircraft engine blades (Rosemau et al. 1999), a robot is used to position the blade in different orientations so as to cover all of the cooling holes with a sequence of video images. Often the fixture for inspection will be that used for transport throughout the system, for example in a flexible manufacturing system. Parts are held in the fixtures by gravity, quick-release clamps, or magnetic couplings, depending upon stability and human compatibility requirements. Pick-and-place robots can be used if an item needs moving from its fixture to the test/inspection device. In nonmanufacturing applications, the sensor must be moved physically around the item inspected, as in the robotic inspection of aircraft skins by Siegel et al. (1998) or the inspection of water pipes by Moraleda et al. (1999).

Very small items produced and tested at very high rates (e.g., fasteners, microcircuit chips) can be handled by vibratory bowl feeders, slides, and small conveyors to ensure that they are correctly presented to the inspection device. These systems often have automated disposal mechanisms built in to gate any defective items along a different track from the good items, an example of a materials-handling component equivalent to the human Response function.

The materials-handling system can be automated or manual. In a system with manual material handling, an operator manually loads the part and then initiates the inspection cycle. In some cases, automated materials handling is suitable and eliminates lifting and lowering of heavy components. The type of materials handling system (and the degree of its automation) depends on the inspection system and its place in the production system (i.e., preprocess, online, or postprocess inspection).

5.1.3. Sensing

In both inspection and process control, we can sense a wide range of properties of the product or item produced, or of the process itself. Thus, in test and inspection, a classification of these sensors aids in choosing the most suitable one from the many available. Here we use a standard listing from a recent computer automation text (Boucher 1996). All use the transducer principle, converting energy of one form (e.g., light, water flow) into a common electrical form with well-defined properties. The electrical signal can then be processed in many ways, independently of the type of transducer used.

A necessary part of sensing is signal generation. Typically some form of energy must be input into the sensor to elicit a response. Thus, a temperature sensor works because the electrical resistance changes, but detecting such a change requires that a voltage be applied across the resistance element and an output voltage measured. Signal generation involves the application of a suitable constant voltage or the illumination of an object to be sensed by appropriate lighting etc. We will consider this explicitly only when we deal with parts of nondestructive inspection in the next section.

5.1.3.1. Discrete-Event Sensors Note that these can be combined into arrays, for example to detect complex logical events, or even items of different sizes when some elements are activated and some not.

- *Mechanical limit switches:* These sense when an object displaces an actuator, either on a lever or by pushing/pulling. Their logic can be normally open or normally closed. Limit switches are used to sense when a machine (e.g., robot) has reached a particular position or when an item arrives on a conveyor.
- *Proximity switches:* Two types are inductive and capacitive, both of which measure when an object enters a defined sensing field. Inductive switches sense only conducting materials such as metals, while capacitive switches indicate the proximity of any object. They are particularly useful because they do not require contact with the item sensed and hence cannot damage it.
- *Photoelectric sensors:* An emitter unit gives out a beam of light or infrared radiation, which is sensed by a receiver unit. The emitter and receiver can be used in an opposed sensing mode, where an electrical event is registered if the beam is blocked. They can also be used in the reflective mode to sense objects of high reflectivity, and even used to sense only objects at a given range of distances using a convergent-beam system.
- *Fluid-flow switches:* An analog of proximity switches for fluid detection, fluid-flow sensors use a sprung or magnetic valve to detect when flow takes place in a pipe.

5.1.3.2. Continuous Sensors Continuous transducers convert energy directly from one form to electrical voltages or currents. A measurement circuit, such as a bridge circuit, is used to make the conversion from the direct sensor signal. The voltage or current is then digitized using an analog-to-digital converter, which provides input to the signal processor. Each type of transducer has a fixed relationship between the input energy and the output digital signal, a form known as the calibration curve.

- *Linear position transducers:* The simplest form is a linear potentiometer, where the position of the slider is proportional to the output voltage. Linear variable differential transformers (LVDIs) move a metal core between primary and secondary coils to produce a voltage proportional to core position.
- *Rotary (angluar) transducers:* Rotary potentiometers sense angle in a way analogous to linear potentiometers. Resolvers use rotary transformers similar to linear variable differential transformers. Optical encoders go directly from angular position to a digital signal by passing light beams through a disc attached to the rotating shaft. The disc has sectors printed in a pattern of black and white so that each angle corresponds to a unique pattern of transmitted and blocked light beams.
- *Float transducers:* As in an automobile gas tank, a float transducer uses a rotary potentiometer to produce a voltage proportional to the angle of an arm with a float attached.
- *Ultrasonic sensor:* The reflection time of a high-frequency sound pulse can be measured to give an accurate estimate of distance from an object or from a liquid level. In the latter case, an ultrasonic sensor is a noncontact alternative to a float transducer.
- *Velocity encoders:* Both velocity (speed) of solid objects and liquid flow rates can be measured. One way is to use an optical encoder and measure time between position changes to give the differential of position, i.e., velocity. Tachometers use an AC or DC generator to give a voltage proportional to angular velocity. Flow rate transducers can use the differential pressure in a pipe to measure a flow rate if pipe diameters are known. Alternatively, a small turbine blade can be inserted into the flow and its rotation velocity used to measure flow rate.
- *Force/pressure transducers:* The simplest force measures use load cells based on an LVDT, which displaces by a small amount as a force is applied. Strain gauges can measure the (small) deformation of thin film attached to a surface by measuring the resistance of an attached conducting element.
- *Temperature transducers:* The most common temperature sensor is a thermocouple made from two different metal wires. As the temperature changes, so does the relative resistance of the wires, giving a signal proportional to the temperature. Thermistors are semiconductors with the property of changing resistance with temperature.

5.1.3.3. Scene Sensors: While the above sensing devices convert discrete or continuous information into electrical signals, they are all essentially one-dimensional. Their output can be characterized by the changing of a single quantity over time, such as position or temperature. However, many automated inspection and test systems need to utilize two-dimensional or even three-dimensional data, for example in inspection of sheet materials or circuit boards. Such systems demand two-dimensional sensing.

The typical sensor for these applications is a charge-coupled device (CCD), which consists of an array of picture elements (pixels) that are light sensitive. Electrons emitted from each pixel are integrated for a field time (typically 1/60 sec) and a voltage proportional to this value is sent out. This voltage is thus proportional to the light intensity at the pixel, although the relationship is not linear. CCD pixels have a number of limitations, such as saturation and bloom, that can degrade image quality.

- **One-dimensional arrays:** Two-dimensional information can be obtained by sweeping a one-dimensional row of sensors across the two-dimensional surface of interest. With a one-dimensional array, such as a row of charge-coupled devices, the necessary condition is that the two-dimensional surface remain constant while the scan is in progress. Thus, any fixed object, such as a circuit board, can be scanned in this way to produce a two-dimensional image for later processing. The essential condition is that the sensor be moved relative to the material to provide an undistorted two-dimensional image. Alternatively, the inspected item itself may be moving at a known velocity, so that data collected from a one-dimensional array will give a complete picture of the material. In such a way, continuous sheet production, such as float glass or sheet steel, can be scanned continuously.
- **Two-dimensional arrays:** If the array is composed of rows and columns of sensors, as in a television picture sensor such as a two-dimensional CCD, then two-dimensional information is available simultaneously for all parts of the inspected material. A lens may be used to form a sharp image on the sensor array. For three-dimensional information, height maps may be constructed with appropriate two-dimensional sensors. The common methods for 3D image acquisition include laser radar, confocal microscopes, and high-speed triangulation-based 3D systems. It is currently believed that triangulation based systems offer the greatest potential, although this is somewhat application dependent (Svetkoff et al. 1989; Gieles et al. 1989).

5.1.4. Signal Processing

While the output from a discrete-event sensor is essentially digital, digitization is first required for all of the sensors of continuous one-dimensional or two-dimensional information,. Digitization is typically performed by an analog-to-digital converter (A to D converter). Such a device quantitizes the continuous information by comparing the signal with known levels or thresholds. If a signal exceeds threshold (i) but not ($i + 1$), then it is digitized at the level corresponding to threshold (i). The reverse process of digital-to-analog (D to A) conversion will not be considered here because it pertains to system effectors rather than receptors. D to A conversion is required when moving from the test and inspection task to a process control task.

For scene sensing, signal processing is more complex and will thus be covered in a separate section.

5.2. Image Processing in Automated Visual Inspection Systems (AVIS)

Because so much of test and inspection is now concerned with vision systems, whether human or automated, this section will use automated vision as an extended example through which to discuss issues of automation in general. A typical example has already been presented, that of rose grading, and will be used to illustrate concepts as they arise.

Lumai (1994) reviews the applications of vision systems and image processing in manufacturing, not all of which are concerned with test and inspection. For example, automated vision systems are applied to control of guided vehicles, control of robot positioning, and determination of orientation of parts for assembly. They may use the same sensors and signal-processing algorithms as an automated visual inspection system (AVIS), but they will not be considered further here.

Figure 3 shows the basic functions of image processing, adapted from Lumai (1994). Note as in Figure 2 that these partially correspond to the generic functions given earlier, with Classification very similar to Decision, Respond the same in both lists, and the other functions partially overlapping with Search. Ventura and Chen (1994) use a similar but more specific breakdown of functions in their system for classifying component shapes based on a two-dimensional image and the original CAD specification.

Signal generation here is the lighting system, as for example in the careful design of even illumination in the rose-grading study. The rest of the image capture system, after the lighting, is composed of a lens and a CCD. The lens is chosen to give the correct ratio between angular coverage

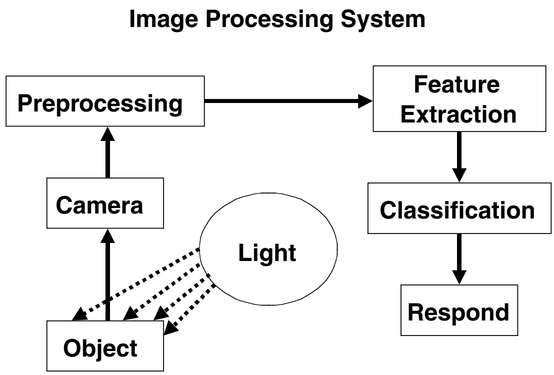

Figure 3 Functions in Image Processing.

and resolution, defined as the solid angle subtended by individual pixels. Note that lenses have their own aberrations, but modern lenses are amazingly free from distortions and also inexpensive. CCD devices are also increasing rapidly in the total number of pixels and the grayscale dynamic range. As of 2000, inexpensive CCDs contain well over 2 million pixels and have at least an 8-bit grayscale range. Note that for color imaging, very important in such inspection applications as food processing, three pixels with red, green and blue filters are used to capture the RGB image, giving $3 \times 8 = 24$ bits per data point. The signal is typically stored in a frame grabber so as to be available as an ensemble to the computer system for further processing. The signal at this stage consists of a grayscale level or RGB value for each pixel for each sampling interval.

Signal preprocessing now has the task of preparing the signal for feature extraction. In the rose-grading example, very little preprocessing was required because the lighting ensured that the object was captured as a dark stem and leaves against a uniform white background. In most industrial applications, however, dirt, visual noise, and stray reflections can degrade both the object inspected and the background. A typical visual noise-reduction technique is smoothing, performed by a low-pass filtering technique. This operation consists of convoluting a small template, say 3×3 pixels, with the original image. The value of the pixel at the center of the template is replaced by the value calculated by summing the product of the 3×3 array of pixel grayscale values and the template values around the center pixel. Thus (Lumia 1994), if the template is composed of nine identical values of 1/9, then the resulting smoothed value is the average of the 3×3 array of pixels. This means that any deviant (noise) value has its effect reduced by a factor proportional to the square root of the template area. The penalty for this smoothing is that signal information may be reduced along with the noise. If the signal were, to use an extreme example, a single deviant pixel, its discriminability would be reduced as if it were noise. The preprocessing may also consist of changing the overall brightness and contrast of the image, as is done, for example, in image-processing software for photographic purposes. This allows the subsequent feature extraction to utilize the full dynamic range of the image information despite changes in incident illumination.

Feature extraction uses a variety of techniques for converting a pixel value array into a set of discrete features useful for classification. One set of operations involves convolution to detect specific features. For example the template (Sobel operator):

$$\begin{matrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{matrix}$$

when convoluted with a pixel array will yield a high value if the image contains a horizontal edge. If the operator is rotated about its leading diagonal, it will detect vertical edges. If edges are neither horizontal nor vertical, the two Sobel operators can be combined by finding the effective gradient in each direction. Typically, the edge is detected by comparing the value of the convolution at each point to a threshold. This threshold can be a fixed value or chosen algorithmically (e.g., Hou et al. 1993).

In addition to edge detection, there are also operators for thinning edges and for converting a set of pixels forming an edge to a straight line or a circular arc. Additional algorithms can be used to detect circles (e.g., holes drilled in components) or irregular features such as the leaves or thorns of roses. Some, as in Steinmetz and Delwiche (1993), use the concept of an allowed area or search for the maximum extent of a contiguous area such as a stem or bud. In fact, as with human perception, the more closely a feature can be specified, the easier it is to detect.

Classification: The set of features from feature extraction must be classified to reach conclusions about the overall acceptability or otherwise of the item inspected. Features such as end points, edges, and blobs need to be combined and classified in order to classify the object as a whole. While for a single measured parameter (e.g., force or temperature) a single threshold will provide this classification, for complex two- and three-dimensional objects the process is less obvious. For example, with the rose grading, one algorithm finds the size of the area representing the bud, but eventually a three-level scale is needed to judge bud development. This requires thresholds to be set in the simplest case, or more usually an algorithm must be developed to mimic human judgement of these factors. Fuzzy set theory can provide such an integration (see Section 4), but a preferred solution is to use neural nets as classifying algorithms. One such neural net architecture was used in the rose-grading example.

Neural network models have the ability to explore several competing hypotheses simultaneously. This is done using computational elements connected by links having variable weights that form massive parallel nets. Neural networks can be classified based on the nature of input data, that is, whether they process binary or continuous valued inputs. The pixels of a digitized image can be expressed either in binary (black or white pixels only) or grayscale format. Another classification is based on the type of learning. Learning can be either supervised or unsupervised. In supervised learning, the network is told what the correct answer should be during training. It can then determine whether or not its output was correct and use its learning law to adjust its weights. In unsupervised learning, on the other hand, the network has no knowledge of the correct answer (Caudill 1988a, b). For most inspection applications, neural networks with continuous valued inputs that are trained under supervision seem appropriate.

Current application of neural networks has included speech recognition, handwritten character recognition, target recognition, industrial parts inspection, and signature recognition (Buffa 1985; Cormier 1991). Cormier (1991) developed a three-layer, back-propagation neural network for circuit board component inspection. The system verifies presence of correct surface-mount components on board. Graphical images of various components were used. The scale of the images was chosen to accomplish an effective resolution of 1 mm per pixel, which is the most common resolution for surface-mount inspection systems. Noise was added to the images to simulate variations due to inconsistent lighting, surface irregularities, imaging equipment, and so on. The network was found to have false alarm rate, miss rate, and nonclassification rates nearing zero at a speed equivalent to covering a 100 mm $\times$ 100 mm area per second.

In a study that will be considered in detail in the section on function allocation, a neural net was developed as one of the automated alternatives to human visual inspection of circuit boards (Hou et al. 1993). It used the output from edge detection and thinning algorithms, followed by a Hough transformation as the inputs to a three-layer back-propagation network, with either a fixed threshold (for fully automated inspection) or a multi-level threshold where a human inspector could help the system make the final decision. The network was given a number of boards as the training set and then given another sample as the test set to measure its performance.

Neural network research and its application to the decision-making component in inspection offer interesting possibilities towards more realistic automated inspection systems. However, a number of issues have to be addressed before neural networks can make a significant impact in the area of test and inspection. The training time for neural networks can be extremely long, especially if there is an overlap of categories in the decision space. The speed of the network also has to meet line-speed requirements. With respect to neural networks, the speed is largely a function of the size of the middle layer. Adding middle layer units can significantly increase computation time while only increasing accuracy up to a point. Accuracy of neural networks in terms of both false alarms and misses has to be established and is always application specific. As with human training, the amount of fine tuning that is needed to get the desired performance may be quite large.

The final output from the overall system is a unit of product classified into one of a small number of categories corresponding to different system responses. Thus, a circuit board could be classified as ‘‘Conforming’’ or ‘‘Nonconforming’’ (Hou et al. 1993) or into three levels of acceptability, such as ‘‘Tight,’’ ‘‘Open 1,’’ or ‘‘Open 2,’’ for rose grades.

5.2.1. Examples

Any recent conference on automated vision from an automation or applied optics viewpoint will provide copious current examples of AVIS applications in a variety of industries. Table 4 lists a number of recent applications of automated vision to show the variety of products and industries covered by the pace of this form of automation. Unfortunately, few technical papers contain the depth of performance evaluation required to assess system reliability or its components. The Steinmetz and

TABLE 4 Examples of Automated Test and Inspection Systems

Industry and Reference	Task	Parameters/ Defects Detected	Technology
Civil Aviation Siegel et al. 1998	In-service structural inspection of aircraft	Cracks, corrosion	3D stereo vision and eddy current
Semiconductor manufacture Komatsu et al. 1999	Detection of defects in lithographic process after development	Contamination, scratches, reduced conductor thickness	Image processing
Aircraft engine manufacture Rosemau et al. 1999	Classification of cooling holes in jet engine fan blades	Cooling hole wrong diameter, blocked cooling holes	Robot movement of fan blade for different views, image processing to classify defects
Food products Legeard et al. 1999	Evaluation of quality of pork hams	Detection and classification of muscle color and fat thickness	Color vision system using hue, saturation, and intensity measures, using thresholding
Ceramics Kälviäinen et al. 1998	Classification of color matches in tile production	Classification of brown tiles by color and feature	Color machine vision system measures RGB pixel values, classify color features using neural nets
Chemical Industry Liu et al., 1998	Detection of welding defects in pipelines	Weld defects of all types	X-radiography followed by image processing and pattern recognition
Automobile Manufacturing Hung and Park 1996	Detection of dents in large automobile steel panels	Dents	3D computer vision measures slope, curvature and depth.

Delwiche (1993) paper was chosen as a worked example partly because such performance data was available and published.

6. NONPRODUCTION TEST AND INSPECTION

Production and manufacturing are not the only realms of inspection and test. Service industries need to inspect goods and services before they are released to the customer. A number of these nonmanufacturing applications are shown below:

- *Regulatory inspection:* To ensure that regulated industries meet meet or exceed regulatory norms. Examples are review of restaurants against local service codes, fire safety inspection of buildings, and safety inspection of workplaces.
- *Maintenance:* To detect failure arising during the service life of a product. This failure detection function can be seen in inspection of road and rail bridges for structural determination or of civil airlines for stress cracks or corrosion (see below).
- *Security:* To detect items deliberately concealed. These may be firearms or bombs carried onto aircraft, drugs smuggled across borders, or camouflaged targets in aerial photographs. They can also be suspicious happenings on a real-time video monitor at a security station. Law enforcement has many examples of searching crime sites for evidence.

- *Design review:* To detect discrepancies or problems with new designs. Examples are the checking of building drawings for building code violations, of chemical plant blueprints for possible safety problems, or new restaurant designs for health code violations.
- *Functionality testing:* To detect lack of functionality in a completed system. This functional inspection can often include problem diagnosis, as with checks of avionics equipment in aircraft. Often functional inspection is particularly dangerous and costly, as in test flying aircraft or checking out procedures for a chemical process.

6.1. Maintenance Inspection

The example chosen here is that of inspection of civil aircraft as part of the system for assuring the public that airworthiness is maintained throughout the service life of airframes, avionics, and aircraft structures. It is part of a maintenance process and is typical of many transportation applications, such as for maritime transport, heavy goods vehicles, or even the space shuttle.

Airworthiness of civil aircraft depends upon a process by which a team composed of aircraft manufacturers, regulators, and one or more airlines predicts possible system failures. This process, Maintenance Steering Group 3 (MSG-3), considers possible failure pathways (e.g., in structures, controls, avionics) and for each pathway determines a recovery strategy. For structural failure, this may be replacement after a fixed service life, regular inspection to ensure detection, or an indication to crew of the malfunction. The concern here is with the reliability of the primary failure recovery system for aircraft structural inspection: regular inspection to ensure detection.

Failure modes of aircraft structures can be cracks, corrosion, fastener/bonding failure, or deformation beyond the plastic limit. Inspection systems are designed to detect all of these in a timely manner, that is, before the failure has a catastrophic effect on structural integrity. For example, crack growth rates can be predicted probabilitistically from material properties and applied stresses, so that the MSG-3 process can schedule inspections before a potential crack becomes dangerous. However, the detection system has certain limits on the size of crack that can be detected, so MSG-3 typically schedules several inspections between the time the crack becomes detectable and the time it becomes dangerous. If too many inspections are scheduled, the costs are driven up in a highly competitive industry and the risk of collateral damage is increased due to the handling activities involved in the inspection process itself. Conversely, if too few inspections are scheduled, the probabilistic rate of the crack growth-prediction process may combine with the probabilistic nature of the detection process to cause dangerous cracks to remain undetected. Spectacular failures of this inspection process have occurred both for aircraft structures (Aloha incident, Hawaii, 1988) and engine components (Pensacola incident, Florida, 1997).

The MSG-3 process thus requires quantitative data on inspection reliability to function correctly. In addition, no rule-based prediction system can foresee all possible malfunctions, so that once an aircraft is in service, regular detailed inspections are made of the whole structure to discover any unexpected cracks. When such "new" cracks are found, the information is typically shared among manufacturers, operators, and regulators in the form of supplementary inspections. Similar considerations apply to other failure modes, such as corrosion.

This whole reliability assurance process thus rests upon an inspection system that checks both points where malfunctions are expected and points where they are not expected, for a variety of malfunctions. For good reasons, human inspectors are part of this inspection system, and thus human inspection reliability is an essential element in ensuring structural integrity and hence airworthiness.

The inspection task combines two goals: detection of expected malfunctions and detection of unexpected malfunctions. Neither detection is particularly easy or particularly rapid, so inspection can be a difficult and time-consuming task. In some ways inspection can be classified as an ill-structured task (Wenner 1999) because there is no simple step-by-step procedure that will ensure success and usually no knowledge of task success available during the task. Finding (n) malfunctions in a structure still leaves an unknown number (hopefully zero) potentially undetected.

In addition, inspection is typically scheduled at the beginning of an aircraft's maintenance visit so that malfunctions can be detected early and their repair scheduled to overlap in time with other maintenance activities. As airlines streamline their parts inventory to reduce holding costs, the lead time for replacement components can increase, again pressuring the inspection system to ensure early detection. Aircraft typically arrive following scheduled service, that is, after the last flight of the day. Following opening up and cleaning processes, maximum inspection resources are committed to the initial inspection. In practice, this means inspectors working overtime, even double shifts, starting with a night shift, under some implied pressure for early detection. Human inspection reliability may not be optimal under these conditions.

The inspection task itself is classified in aviation as either visual inspection or nondestructive inspection. Regulatory bodies have issued formal descriptions of both of these tasks (e.g., Bobo 1989, for the FAA), and both have somewhat different characteristics in aviation

Nondestructive inspection (NDI) includes a set of techniques to enhance the ability to detect small and/or hidden malfunctions. One set of NDI techniques is those that enhance what is essentially still a visual inspection task, such as X ray, fluorescent particle, magnetic particle, or D Sight. They show cracks that are very small (fluorescent particle) or hidden within other structures (X ray). Apart from the steps necessary to ensure a good image, they have many of the human interface characteristics of visual inspection. The other set of NDI techniques is focused on specific malfunctions in specific locations, such as eddy current and ultrasound. For this reason, they are useful only for detection of malfunctions already predicted to exist. In practice, such NDI techniques are much more proceduralized than visual inspection or NDI techniques, which contain a human visual inspection component.

Visual inspection is much more common, making up 80% of all inspection (Goranson and Rogers 1983). It consists of using the inspector's eyes, often aided by magnifying lenses and supplementary lighting, as the detection device. Inspectors must visually scan the whole structure of interest, typically using portable mirrors to examine areas not directly visible. Whether the task is categorized as visual inspection or NDI, its aim is to detect flaws (indications) before they become hazardous. Next we consider the bodies of knowledge potentially applicable to aircraft inspection reliability (e.g., Drury and Spencer 1997).

Over the past two decades there have been several studies of human reliability in aircraft structural inspection (Rummel et al. 1989; Spencer and Schurman 1995; Murgatroyd et al. 1994). All of these to date have examined the reliability of nondestructive inspection (NDI) techniques, such as eddy current and 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., Spencer and Schurman 1995). This curve expresses reliability of the detection process (PoD) as a function of a variable of structural interest, such as 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 (e.g., cracks) and much of the variance in PoD can be described by just crack length, the PoD is a realistic reliability measure. It also provides the planning process with exactly the data required because remaining structural integrity is largely a function of crack length.

Both the FAA (1993, pp. 26, 35) and the Air Transport Association (ATA) have recognized the need for equivalent studies of the reliability of visual inspection as a research priority.

Aircraft inspection has already benefited from models of human inspection such as those in Section 2. Thus, the ECRIRE program (Spencer and Schurman 1995) examined one NDI technique, eddy current inspection, incorporating human factors variables. They were able to test one-person vs. two-person teams (no consistent effects) and gross body posture (a small decrease in detection performance when the inspector had to work at about knee height). A FAA program on human factors in aviation maintenance and inspection (e.g., Drury et al. 1997) has had some success in improving documentation design, lighting, and communications. This program expanded the search-plus-decision model following industrial inspection findings to include the five generic inspection functions presented earlier.

Such a task description invites task analysis, which would lead naturally to human reliability analysis (HRA). Indeed, perhaps the earliest work in this field applied HRA techniques to construct fault trees for aircraft structural inspection (Lock and Strutt 1985). The HRA tradition lists task steps, such as expanded versions of the generic functions above, lists possible errors for each step, then compiles performance shaping factors for each error. Such an approach was tried early in the FAA's human factors initiative (Drury et al. 1990) but was ultimately seen as difficult to use because of the sheer number of possible errors and PSFs. It is occasionally revised, such as in the current FRANCIE project (Haney 1999), using a much expanded framework that incorporates inspection as one of a number of possible maintenance tasks. Other attempts have been made to apply some of the richer human error models (e.g., Reason 1990; Hollnagel 1997; Rouse 1985) to inspection activities (Latorella and Drury 1992; Prabhu and Drury 1992; Latorella and Prabhu 2000) to inspection tasks. These have given a broader understanding of the possible errors but have not helped better define the PoD curve needed to ensure continuing airworthiness of the civil air fleet.

One outcome of this work has been a detailed analysis of a single process to develop recommendations of human factors good practices for use by industry managers and inspectors (Drury 1999). The project was performed for the regulatory body, the FAA, as part of its response to recommendations arising from investigation of the engine hub failure at Pensacola in 1997. That failure was due in part to failure of the fluorescent penetrant inspection (FPI) system for titanium hubs. Hence a human factors analysis of engine FPI systems was performed.

The human inspection models of visual search and human decision making were shown to be particularly applicable leading to a task analytic framework using hierarchical task analysis (HTA). In this way, human factors knowledge could be applied systematically to observed FPI processes. Visits were made to several engine repair facilities owned by major air carriers and engine manufacturers. The site visit team consisted of a human factors engineer specializing in inspection reliability and two senior Federal Aviation Administration (FAA) personnel specializing in NDI.

At each visit the HTA model of FPI was further developed and both good and poor human factors practices were noted. The HTA had seven major tasks, most of which would be classified under Setup in our five-function model. The tasks are shown in a HTA format in Figure 4. Note that Load/transport part in FPI would be equivalent to the Present function, and Read part for defects to Search, Decision, and Respond. For each of these seven tasks, two levels of subtasks were developed and potential errors listed. A fault tree for the primary failure, defect not reported, was developed and used to help generate a set of human factors good practices. Table 5 shows an example of the most detailed level for function 3, Apply penetrant, with the listing of human factors issues and process variances.

Using this process, a set of 90 human factors good practices was generated, with each good practice keyed to one of the seven major tasks listed above. Each was also keyed to the potential errors that that good practice can prevent. In this way, users are given the reasons for the recom-

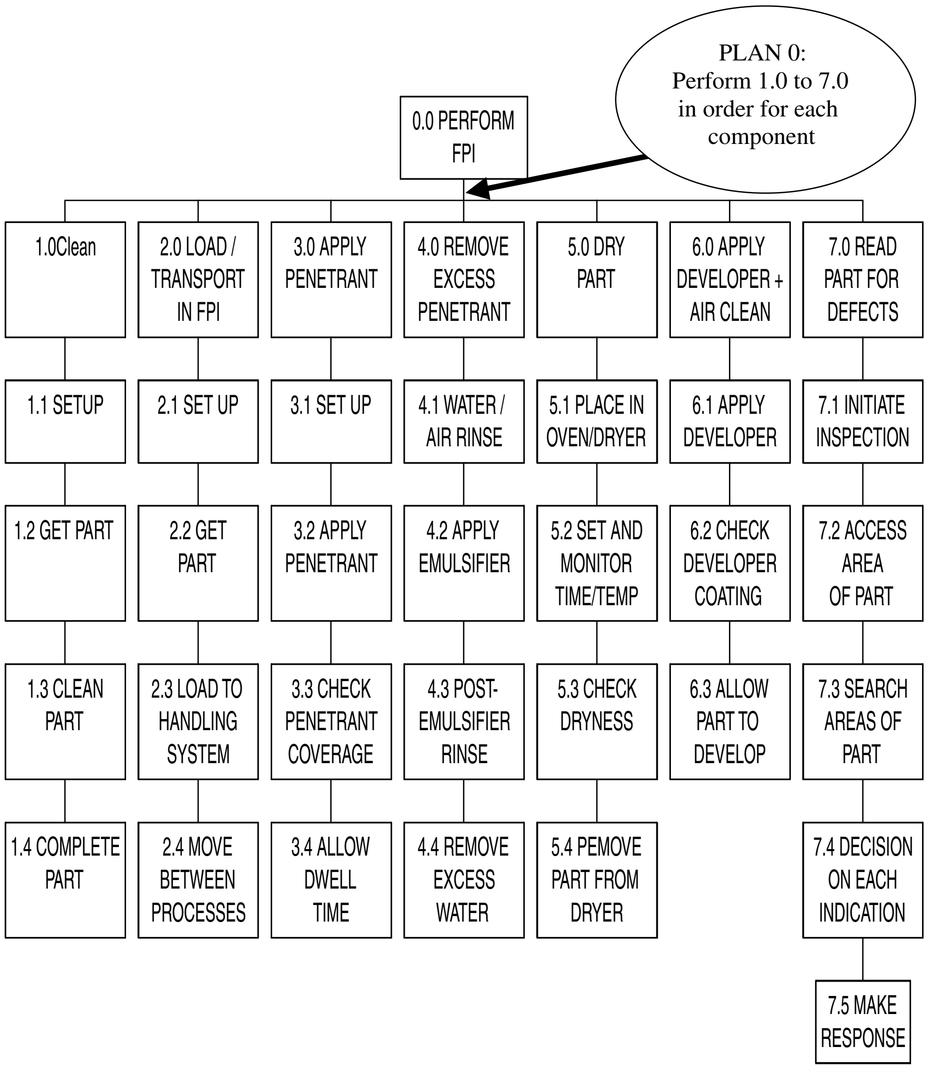

Figure 4 Hierarchical Task Analysis for the Top Level of Fluorescent Penetrant Inspection.

TABLE 5 Detailed Level of HTA for Function 3: Apply Penetrant

Task Step	Task Description	Task Analysis
3.1. Set-up	3.1.1. Monitor penetrant type, consistency (for electrostatic spray) or concentration, chemistry, temperature, level (for tank)	Are measurements conveniently available? Are measurement instruments well human-engineered? Do recording systems require quantitative reading or pass/fail?
3.2. Apply		
3.2.1. Electrostatic spray	3.2.1.1. Choose correct spray gun, water-washable or postemulsifiable penetrants available.	Are spray guns clearly differentiable? Can feeds be cross-connected?
	3.2.1.2. Apply penetrant to all surfaces.	Can sprayer reach all surfaces?
3.2.2. Tank	3.2.2.1. Choose correct tank, water washable or postemulsifiable penetrants available.	Are tanks clearly labeled? Is handling system well-designed to use for part placement?
	3.2.2.2. Place in tank for correct time, agitating/turning as needed.	Does operator know when to agitate/turn? Does carrier interfere with application?
	3.2.2.3. Remove from tank to allow to drain for specified time.	Is drain area available?
3.2.3. Spot	3.2.3.1. Choose correct penetrant, water-washable or postemulsifiable penetrants available.	Are spot containers clearly differentiable? Does operator know which areas to apply penetrant to?
	3.2.3.2. Apply to specified areas with brush or spray can.	Can operator reach all areas with spray can/brush? Is handling systems well human-engineered at all transfer stages?
3.3. Check coverage	3.3.1. Visually check that penetrant covers all surfaces, including holes.	Can operator see penetrant coverage? Is UV light/white light ratio appropriate?
	3.3.2. Return to 3.2 if not complete coverage.	Can operator see all of part? Can handling system back up to reapplication?
3.4. Dwell time	3.4.1. Determine dwell time for part.	Does operator know correct dwell time? How is it displayed?
	3.4.2. Allow penetrant to remain on part for specified time.	Are production pressures interfering with dwell time? Is timer conveniently available, or error-proof computer control?
	Errors/Variances for 3.0 Apply Penetrant Process measurements not taken Process measurements wrong Wrong penetrant applied Wrong time in penetrant Insufficient penetrant coverage Penetrant applied to wrong spots No check on penetrant coverage Dwell time limits not met	

mendations so that they can develop a knowledge base in addition to the rule-based good practices. In addition, there were five general control mechanisms where applications of human factors principles should lead to improved FPI reliability: Operator selection, training and turnover, Hardware design, Software and job aids, Interpersonal systems design, and Environmental control. Each was considered to show how human factors could be applied at a higher level than the 90 specific recommendations. Finally, the analysis established a set of research and development needs to provide better support for improved FPI reliability: Improved solvent and developer, Better magnifying loupe, Better process test panel validity, Job aids for search strategy, and Realistic expectation control.

It was concluded that the FPI system for critical rotating engine components can reach significantly higher reliability through application of specific and general human factors good practices. For the current purposes, the analysis provides a worked example of how human factors knowledge can be applied to find practical changes in an existing inspection system.

7. LOGICAL FUNCTION ALLOCATION IN TEST AND INSPECTION

Given that test and inspection systems must be designed and redesigned on an accelerated cycle, systematic design procedures would be advantageous to modern manufacturing. With the array of devices and human factors interventions available to the designer, a procedure that highlights functional and organizational differences between alternative solutions would assist in the evaluation or convergence phase of design (Jones 1981, p. 68). In Section 3, the functions of inspection were defined and the principles of allocation of function between human and machine proposed (Section 3.1).

Function allocation is a basic technique of human factors engineering, although it has periodically been called into question. Most recently, McCarthy et al. (2000) edited a special issue of a journal devoted to a modern consideration of function allocation in manufacturing. Although test and inspection are not mentioned, the issues raised clearly apply to these manufacturing functions. They argue that function allocation at the design stage is against the principles of localized design and is often rational and atomistic, both currently disfavored concepts in the field. They see function allocation as being at times arrogant instead of growing more naturally out of each particular operating environment. However, in any design there is always allocation of function, even if it is only implicit or a default to automation. Human operators still have too many jobs composed of leftover functions in a system designed for full automation but failing to achieve that in practice.

In test and evaluation, much of the automation literature is frankly antihuman operator. For example, Lumai (1994) states that "the subjective nature of human inspectors prevents achieving a 'six-sigma' manufacturing capability" going on to give the argument that if inspectors only achieve a 90% success rate, this will not meet the six-sigma requirement. In fact, if the defective rate is low enough, even a 90% hit rate can achieve arbitrarily low final defect rates. But 100% inspection by humans is rarely the issue in industry. Most research papers, as noted earlier, start by justifying automation based on human error and/or human variability. The implication is that automation will prevent errors and remove unwanted variability. Much of modern manufacturing is based on continual variability reduction (e.g., total quality management), so this aim is reasonable. There is no justification for variability for its own sake. But variability of *response* may well be required in a practical system to allow the system to cope with unexpected variance in inputs or process conditions. Human inspectors provide the required ability to respond to external change, for example by finding defects not initially defined for the process. In fact, one characteristic of automated systems is that they function best when inspecting for a limited set of defect types, often a single type. Humans share this characteristic but can expand their effective set of detect types much more easily than automated devices.

Three examples of direct tests of function allocation have been reported. The first, Drury and Sinclair (1983), compared two alternative designs for inspection of precision roller bearings for discrete surface imperfections (scratches, toolmarks, nicks, and dents). The current manual method using trained inspectors was compared to a computer-based automated visual inspection system (AVIS). In fact, the systems differed on the functions of Setup, Search, and Decision. The latter two functions could not be separated for the AVIS, indicating one difficulty of performing allocation at the function level. Results, in terms of misses and false alarms, were poor for both systems, with the automated system slightly worse than the human inspectors. In fact, the company embarked upon a very effective retraining program for its inspectors (Kleiner 1986), while redesigning the AVIS for better discrimination between conforming and nonconforming indications.

The second example, Drury and Goonetilleke (1992), used an inspection device for populated printed circuit boards (called a CVC) that could be reconfigured to allocate functions progressively between human and device. Direct comparisons of performance (misses, false alarms, performance speed) were possible between six levels of automation from level 1 (fully manual) to level 6 (automated Setup, Present, and Search), with only the final decision left to the inspector. Response via a control panel was kept constant throughout. Both laboratory and in-plant studies were undertaken, but only the former are presented here. Figure 5 shows performance for a number of measures of defect-detection performance that gave significant effects. Note that performance generally falls with increasing automation until an automated search procedure that detects all indications is reached. Performance differences between the fully manual and fully automated systems were very small.

The final example is from Hou et al. (1993), who examined five function-allocation alternatives for the task of inspecting circuit board height maps for a set of three defects: missing component, misaligned component, and wrong-sized component. The five function allocations were based on the fact that humans are expected to perform worse than computer-based automation on the search function but may be better for decision making. Thus, the extreme allocations of all tasks to the human and all tasks to the machine were complemented by an allocation of machine search/human decision and another of machine search/shared decision. For automated inspection both a template matching system and a neural net system were used. Each of the five systems was optimized, for example by finding the optimum human search time (Morawski et al. 1992) for human search or by choosing parameters for the automated systems that maximized performance. All five systems were tested using four human inspectors for a range of three circuit board sizes and three different image contrast levels to simulate environmental degradation. Results were measured by the set of performance measures [p(miss), p(false alarm), time for inspection], with the two probabilities being combined using signat-detection theory into the single nonparametric measure of discriminability, $p(A)$. There were significant differences between the five systems on $p(A)$, with the two fully automated systems giving the worst discriminability. When inspection time was factored into the evaluation, the graph in Figure 6 could be drawn. This shows the speed–accuracy trade-off (SATO) across the five systems. Optimum performance is in the upper left corner of the figure with maximum discriminability in minimum time. The contours represent potential cost trade-offs for speed and accuracy, shown as straight lines for simplicity. From this figure, it can be seen that the ultimate choice of system depends upon the exact SATO cost weighting, but some systems dominate others no matter what the weightings. For example, template automation dominates neural net automation. Also, the human decision system dominates both the shared decision and the totally human systems. The conclusion from this study was that some form of human/automation hybrid system is preferable to either total allocation to human or total allocation to automation. It was also found that the hybrid systems were more consistent across different levels of environmental degradation.

Note that the set of criteria for deciding on function allocation was different for the three studies. The first used only the error probabilities, the second expanded this set to include performance measures such as misclassification rate, while the third also included performance speed. The issue of flexibility was addressed in the third study by comparing the competing systems across different contrast levels and board sizes. In none of these examples was there any measure of the cost of

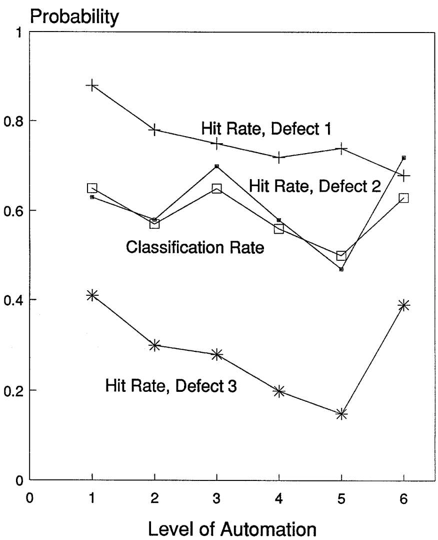

Figure 5 Performance with Increasing Automation.

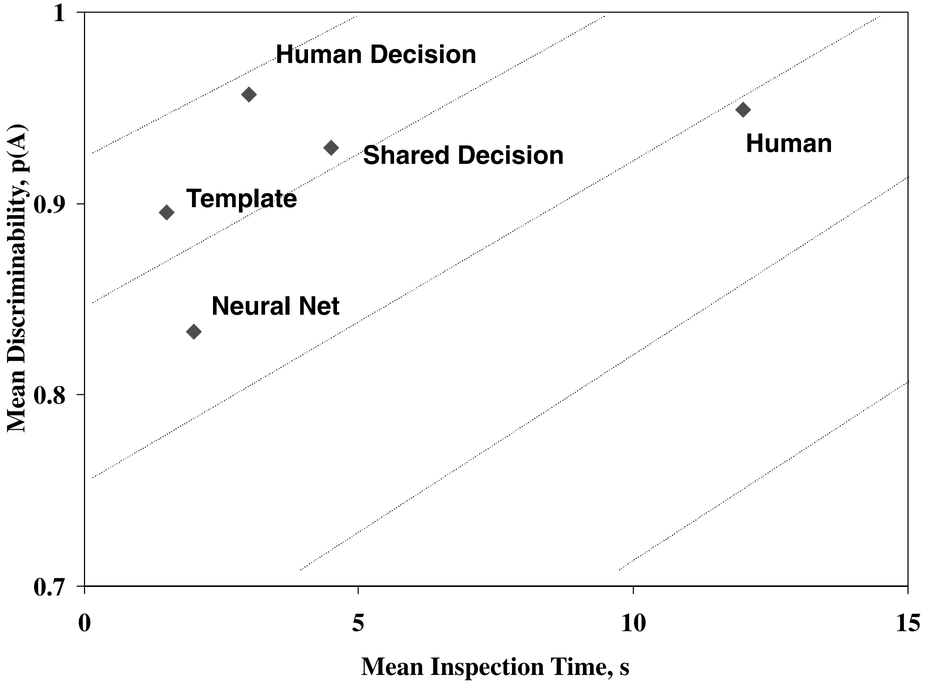

Figure 6 Accuracy and Speed Measures for the Five Alternative Function Allocations of Hou et al. (1993). The Dotted Lines Represent Typical Isocost Contours.

performance to the human inspector, although another part of the second study did use measures of workload and stress. The final choice of allocation is dependent on the set of criteria used to compare the systems.

7.1. A Methodology for Test and Inspection Systems Design

The implication that there are five functions in Table 3 that can each be allocated to either component is perhaps too simplistic. Not only are there practical interactions between functions, but there are multiple potential solutions to the design problem represented by each function. The best function allocation may well change over time, meaning that a dynamic rather than a static allocation procedure is needed. One approach to the design problem is the design for inspectability procedure proposed by Drury (1992b), although this was concerned primarily with the design of a product to enhance its inspectability. Here the procedure is further developed to examine different design alternatives.

Note that the outcome errors for test and inspection were defined as misses, false alarms, and (to a lesser extent) delays. Any system must attempt to minimize some function of these errors, although it has been shown that they can have trade-offs. For example, in the Search function, allowing greater time per item (increased delay) reduces misses. Also, in Decision, variation of the criterion for reporting (S_i) has the effect of causing misses and false alarms to covary inversely. Thus, compromises among all three measures will usually be required. Because the functions are sequential, failure to detect a defect (miss) is usually a much more common error than a false alarm (e.g. Megaw 1992).

If the probability of search success is p_s and the probability of decision success is p_d (here for simplicity both are assumed constant for good and faulty items):

$$p\ (\text{miss}) = p_s\ (1 - p_d) + (1 - p_s)$$

$$p\ (\text{false alarm}) = p_s\ (1 - p_d)$$

Thus,

$$p \text{ (false alarm)} < p \text{ (miss) for}$$
$$0 < p_s < 1$$
$$0 < p_d < 1$$

In addition to being numerically the most important error, misses are usually the most expensive in consequences because they imply failure of the system in use. False alarms, while still costly, can often be reevaluated by slower, off-line testing and usually corrected.

Thus, while the complete set of performance measures [p(miss), p(false alarm), time for inspection] is required to evaluate competing systems and systems components, for design purposes the detection of nonconforming items must be the priority. This is particularly so in modern manufacturing environments, where nonconforming items are increasingly rare and increasingly important to detect early. Thus, rather than optimizing the set of all three measures, in practice it is often imperative to optimize (minimize) p(miss) subject to fixed limits or constraints on p(false alarm) and time for inspection.

The design for inspectability procedure noted above starts from the assumption that a list of defects to be detected can be developed at the product design stage, perhaps from a failure modes and effect analysis (FMEA) of the product's components. These defects (D_1, D_2, . . . , D_i, . . . , D_n) represents the challenges to the test and inspection system because all must be detected, and detected at the appropriate stage of manufacture. Only three components of the inspection system can be changed, or changes can affect:

Product: the item produced and its subcomponents

Process: each test or inspection process used on the item and its subcomponents

Person: the human or humans interfacing with the product or process at each test or inspection

A design procedure must find an optimum mix of product, process, and person design features for each function to ensure that the probability of detection is maximized, subject to false alarm and time constraints. Note that the function-allocation procedure has been transformed into a function design feature, with design suggestions for each function (or even group of functions) being evaluated for how it impacts product, process, and person design and how it affects the probabilities of detection of each defect. Table 6 summarizes this design procedure, with each row representing a design alternative for one function and containing design impacts on both system components (product, process, people) and detection probabilities for the set of all defects. Entries in the System Design Impacts columns are references to notes concerning, for example, impact upon training or product weight. Entries in System Performance Impacts are detection probabilities for each defect or, where these are not available, in indication of the direction of change from a reference test and inspection system, perhaps the current system. Symbols of +, −, or 0 would be suitable in such a procedure, with actual magnitudes included if available.

TABLE 6 Design Procedure for Test and Inspection

Function	Design Alternative	System Design Impacts			D_1,	D_2, . . .	D_i, . . .	D_n,
		Product	Process	Person				
Setup	SU Alternative 1							
	SU Alternative 2							
	—							
Present	*P* Alternative 1							
	P Alternative 2							
	—							
Search	*S* Alternative 1							
	S Alternative 2							
	—							
Decision	*D* Alternative 1							
	D Alternative 2							
	—							
Response	*R* Alternative 1							
	R Alternative 2							
	—							

In many automated inspection systems reported in the literature, there is a concentration on a single performance measure, which makes it difficult to handle allocation decisions. For example, a vision system may be quoted as being able to detect surface defects ‘‘in the 20 μm range’’ without saying what the false alarm rate would be in those conditions. As Drury and Sinclair (1983) have shown, it is possible to change the ratio of misses to false alarms over the complete range by varying the sensitivity or threshold of an automated system. Most papers on human inspection quote at least the miss and false alarm rates and often the time per item. For automated systems, this is not typically true, making direct comparisons currently difficult.

8. CONCLUSIONS ON TEST AND INSPECTION

While much ground has necessarily been covered in this review, certain rather general and simple conclusions are possible. First, there is indeed a role for test and inspection in modern business. Global pressures are likely to increase the quality requirements over time while simultaneously requiring efficient performance with minimal resources. Give that test and inspection will still be needed, their role within the organization is moving from checking output for customer protection to collecting process data for point of manufacture control. This contextual change should in fact benefit human operators by including an explicit test and inspection function within most jobs. In the context of automation, we now have a plethora of sensors and signal processors to choose from, with capability increasing while cost decreases over time. Our choice of human and automation roles is likely to require hybrid systems, at least for image data where pattern recognition is required. Having said this, there appears to be a general conclusion that automation is preferable for the search function, at least where defects can be specified in advance. For routine measurement, such as of forces or temperatures, human roles are much reduced. Fortunately, for such measurements the ultimate goal is process control, where humans have a definite role in interpretation of patterns of results and choice of remedial action.

REFERENCES

Bainbridge, L. (1990), ‘‘Verbal protocal analysis,’’ in *Evaluation of Human Work,* J. R. Wilson and E. N. Corlett, Eds., Taylor & Francis, London, pp. 161–179.

Baveja, A., Drury, C. G., Karwan, M., and Malone, D. M. (1996), ‘‘Derivation and Test of an Optimum Overlapping-Lobe Model of Visual Search,’’ *IEEE Tranactions on Systems, Man, and Cybernetics,* Vol, 28, pp. 161–168.

Bobo, S. (1989), ‘‘Communication and Transfer of Nondestructive Inspection Information,’’ in *Human Factors Issues in Aircraft Maintenance and Inspection—Information Exchange and Communications, Second Federal Aviation Administration Meeting,* Washington, DC.

Boucher, T. O. (1996), *Computer Automation in Manufacturing: An Introduction,* Chapman & Hall, London, pp. 92–132.

Budros, A. (1999). ‘‘A Conceptual Framework for Analyzing Why Organizations Downsize, *Organizational Science,* Vol. 10, No. 1, pp. 69–82

Buffa, M. (1985), ‘‘Process Control for Automatic Component Assembly of Surface Mounted Devices Utilizing a Machine Vision Controller,’’ *SME Vision ‘85 Conference Proceedings,* Society of Manufacturing Engineers, Dearborn, MI, pp. 5.180–5.200.

Caudill, M. (1988a), ‘‘Neural Networks Primer—Part I,’’ *Al Expert,* Vol. 31, No. 4, pp. 53–59.

Caudill, M. (1988b), ‘‘Neural Networks Primer—Part II,’’ *Al Expert,* Vol. 31, No. 5, pp, 53–59.

Chaney, F. B., and Teel, K. S. (1967), Improving Inspector Performance through Training and Visual Aids,’’ *Journal of Applied Psychology,* Vol. 51, pp. 311–315.

Chi, C.-F., and Drury, C. G. (1995), ‘‘A Test of Economic Models of Stopping Policy in Visual Search,’’ *IIE Transactions,* Vol. 27, pp, 392–393.

Cormier, D. R. (1991), ‘‘Neural Network Based Surface Mount Component Verification,’’ Master’s Thesis, State University of New York at Buffalo, Buffalo, NY.

Diehl, A. (1990), ‘‘The Effectiveness of Aeronautical Decision-Making Training,’’ in *Proceedings of the Human Factors Society 34th Annual Meeting,* pp. 1367–1371.

Drury, C. G. (l978), ‘‘Integration of Human Factors Models into Statistical Quality Control,’’ *Human Factors,*, Vol. 20, No. 5, pp. 56l–572.

Drury, C. G. (1991), ‘‘Ergonomics Practice in Manufacturing,’’ *Ergonomics,*, Vol. 34, pp. 825–839.

Drury, C. G. (1992a), ‘‘Inspection Performance,’’ in *Handbook of Industrial Engineering,* 2nd Ed., G. Salvendy, Ed., New York, John Wiley & Sons, pp. 2282–2314.

Drury, C. G. (1992b), ''Design for Inspectability,'' in *Design for Manufacturability: A Systems Approach to Concurrent Engineering and Ergonomics,* M. H. Helander and M. Nagamachi, Eds., Taylor & Francis, London, pp. 204–216.

Drury, C. G. (1996), ''Ergonomics Society Lecture 1996: Ergonomics and the Quality Movement,'' *Ergonomics,* Vol. 40, No. 3, pp. 249–264.

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

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

Drury, C. G., and Goonetilleke, R. S. (1992), ''Stress, Performance and Automation: A Field Study,'' in *Proceedings of the 28th Annual Conference of the Ergonomics Society of Australia,* pp. 48–55.

Drury, C. G., and Kleiner, B. M. (1990), ''Training in Industrial Environments,'' in *Proceedings of the Human Factors Association of Canada Meeting* (Ottawa), pp. 99–108.

Drury, C. G., and Sinclair, M. A. (1983), ''Human and Machine Performance in an Inspection Task,'' *Human Factors,* Vol. 25, No. 4, pp. 391–400.

Drury, C. G., and Spencer, F. W. (1997), ''Human Factors and the Reliability of Airframe Visual Inspection,'' in *Proceedings of the 1997 SAE Airframe/Engine Maintenance and Repair Conference* (AEMR '97) (August).

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,* pp. 1181–1185.

Drury, C. G., Patel, S. C., and Prabhu, P. V. (2000), ''Relative Advantage of Portable Computer-Based Workcards for Aircraft Inspection,'' *International Journal of Industrial Ergonomics,* Vol. 26, No. 2, pp. 163–176.

Engel, F. L. (1971), ''Visual Conspicuity, Directed Attention, and Retinal Locus,'' *Visual Research,* Vol. 11, pp. 563–576.

Eriksen, C. W. (1990), ''Attentional Search of the Visual Field,'' in *Visual Search,* D. Brogan, Ed., Taylor & Francis, pp. 3–19.

Evans, J. R., and Lindsay, W. (1993), *The Management and Control of Quality,* West, Minneapolis-St. Paul.

Federal Aviation Administration (FAA) (1993), *National Aging Aircraft Research Program Plan,* FAA Technical Center, Atlantic City, NJ.

Friedman, T. L. (1999), *The Lexus and the Olive Tree: Understanding Globalization,* Farrar, Straus & Giroux, New York.

Gallwey, T. J., and Drury, C. G. (1986), ''Task Complexity in Visual Inspection, *Human Factors,* Vol. 28, No. 5, pp. 595–606.

Gieles, I., and Venema, W. J. (1989), ''Inspection of SMD's with 3-D Laser Scanning,'' *SME Vision '89 Conference Proceedings* (Chicago), pp. 5.59–5.71.

Gramopadhye, A. K., Drury, C. G., and Prabhu, P. (1997a), ''Training for Aircraft Visual Inspection,'' *Human Factors and Ergonomics in Manufacturing,* Vol. 3, pp. 171–196.

Gramopadhye, A. K., Drury, C. G., and Sharit, J. (1997b), ''Feedback Strategies for Visual Search in Airframe Structural Inspection,'' *International Journal of Industrial Ergonomics,* Vol. 19, No. 5, pp. 333–344.

Greening, C. P. (1975), ''Mathematical modeling of Air-to-Ground Target Acquisition,'' *Human Factors,* Vol. 18, pp. 111–147.

Hackman, J. R. (1990), *Groups That Work,* Jossey-Bass, San Francisco.

Hancock, W. M., Sathe, P., and Edosomwan, J. A. (1992), ''Quality Assurance,'' in *Handbook of Industrial Engineering,* 2nd Ed., G. Savendy, Ed., John Wiley & Sons, New York, pp. 2221–2234.

Haney, L. (1999), ''Framework Assessing Notorious Contributing Influences for Error (FRANCIE)'' in *Proceedings of International Workshop on Human Factors in Space* (Tokyo).

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

Holding, D. H. (1981), *Human Skills,* John Wiley & Sons, New York.

Hollnagel, E. (1989), ''The Phenotype of Erroneous Actions: Implications for HCI Design,'' in *Human-Computer Interaction and Complex Systems,* G. R. S. Weir and J. L. Alty, Eds., Academic Press, London.

Hollnagel, E., Ed. (1997), *CREAM—Cognitive Reliability and Error Analysis Method,* Elsevier Science, New York.

Hou, T.-S., Lin, L., and Drury, C. G. (1993), "An Empirical Study of Hybrid Inspection Systems and Allocation of Inspection Function," *International Journal of Human Factors in Manufacturing,* Vol. 3, pp. 351–367.

Hovey, P. W., and Berens, A. P. (1988), "Statistical Evaluation of NDE Reliability in the Aerospace Industry," in *Review of Progress in Quantitative Nondestructive Evaulation,* Vol. 7B, D. D. Thompson and D. E. Chimenti, Eds., Plenum Press, New York, pp. 1761–1768.

Hung, Y. Y., and Park, B. G. (1996), "Automated Inspection and Quantification of Dents in Automotive Bodies Using 3-D Computer Vision Technique," SME Technical Paper IQ, Society of Manufacturing Engineers, Dearborn, MI.

Jamieson, G. H. (1966), "Inspection in the Telecommunications Industry: A Field Study of Age and Other Performance Variables," *Ergonomics,* Vol. 9, pp. 297–303.

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,* pp. 1171–1175.

Jones, J. C. (1981), *Design Methods: Seeds of Human Factors,* John Wiley & Sons, New York.

Kälviäinen, H., Kukkonen, S., Parkinnen, J., Hyvärinen, T. (1998), "Quality Control in Tile Production," *Proceedings of SPIE,* Vol. 3522, pp. 355–360.

Kanter, E. M. (1995), *World Class: Thriving Locally in the Global Economy,* Simon & Schuster, New York.

Kanter, E. M. (1999), "Challenges in the Global Economy," *The Washington Quarterly,* Vol. 22. No. 2, pp. 39–58.

Karwowski, W., Kosiba, E., Benabdallah, S., and Salvendy, G. (1990), "A Framework for Development of Fuzzy GOMS Model for Human Computer Interaction," *International Journal of Human–Computer Interaction,* Vol. 2, pp. 287–307.

Kelly, M. L. (1955), "A Study of Industrial Inspection by the Method of Paired Comparisons," *Psychological Monographs,* Vol. 69(9) (No. 394), p. 16.

Knight, J. L., and Salvendy, G. (1992), "Psychomotor Work Capabilities," in *Handbook of Industrial Engineering,* 2nd Ed., G. Salvendy, Ed., John Wiley & Sons, New York, pp. 978–1004.

Komatsu, K., Omori, T., Kitamura, T., Nakajima, Y., Aiyer, A., and Suwa, K. (1999), "Automatic Macro Inspection System," in *Proceedings of SPIE,* Vol. 3677, pp. 764–771

Komorowski, J. P., Simpson, D. L., and Gould, R. W. (1991), "Enhanced Visual Technique for Rapid Inspection of Aircraft Structures (D. Sight)," *Materials Evaluation,* Vol. 49, pp. 1486–1490.

Kundel, H. S., and Lafollette, B. S. (1972), "Visual Search Patterns and Experience with Radiological Images," *Radiology,* Vol. 103, pp. 523–528.

Latorella, K., and Drury, C. G. (1992), "A Framework for Human Reliability in Aircraft Inspection," in *Proceedings of the 7th FAA Meeting on Human Factors Issues in Aircraft Maintenance and Inspection,* Federal Aviation Administration, Washington, DC.

Latorella, K., and Prabhu, P. (1998), "Review of Human Error in Aviation Maintenance and Inspection," *International Journal of Industrial Engineering,* Vol. 26, No. 2, pp. 133–161.

Legeard, D., Marty-Mahe, P., Camillerapp, J., Marchal, P., and Leredde, C. (1999), "Real-Time Quality Evaluation of Pork Hams by Color Machine Vision," in *Proceedings of SPIE,* Vol. 3652, pp. 138–149.

Liu, Y., Li, X., Ren, D., Ye, S., Wang, B., and Sun, J. (1998), "Computer Vision Application for Weld Defect Detection and Evaluation," in *Proceedings of SPIE,* Vol. 3558, pp. 354–357.

Lloyd, C. J., Boyce, P., Ferzacca, N., Eklund, N., and He, Y. (2000), "Paint Inspection Lighting: Optimization of Lamp Width and Spacing," *Journal of the Illuminating Engineering Society,* Vol. 28, pp. 99–102

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

Lumai, R. (1994), "Image Processing in Manufacturing," in *Handbook of Design, Manufacturing and Automation,* R. C. Dorf and A. Kusiak, Eds., John Wiley & Sons, New York.

Maher, J., Overbach, W., Palmer, G., and Piersol, D. (1970), "Enriched Jobs Improve Inspection Performance," *Work Study and Management Services,* October, pp. 821–824.

McCarthy, J. C., Fallon, E., and Bannon, L. (2000), "Dialogues on Function Allocation," *International Journal of Human Computer Studies,* Vol. 52, No. 2, pp. 191–202.

McKenzie, R. M. (1958), "On the Accuracy of Inspectors," *Ergonomics,* Vol. 1, No. 3, pp. 258–272.

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

Megaw, E. D. (1992), *Contemporary Ergonomics,* Taylor & Francis, London.

Monden, Y. (1992), "Just-in-Time Production System," in *Handbook of Industrial Engineering,* 2nd Ed., G. Salvendy, Ed., John Wiley & Sons, New York, pp. 2116–2130.

Moraleda, J., Ollero, A., and Orte, M. (1999), "Robotic System for Internal Inspection of Water Pipes," *IEEE Robotics and Automation Magazine,* Vol. 6, pp. 30–41.

Morawski, T., Drury, C. G., and Karwan, M. H. (l980), "Predicting Search Performance for Multiple Targets," *Human Factors,* Vol. 22, No. 6, pp. 707–7l8.

Morawski, T. B., Drury, C. G., and Karwan, M. H. (1992), "The Optimum Speed of Visual Inspection Using a Random Search Strategy," *IIE Transactions,* Vol. 24, pp. 122–133.

Moray, N., Lootseen, P., and Pajak, J. (1986), "The Acquisition of Process Control Skills," *IEEE Transactions,* Vol. SMC-16, pp. 497–505.

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, AEA Technology, Culham, UK.

National Research Council (1997), *Enhancing Organizational Performance,* National Academy Press, Washington, DC.

Nurani, R. K., and Akella, R. (1996), "In-line Defect Sampling Methodology in Yield Management: An Integrated Framework," *IEEE Transactions on Semiconductor Manufacturing,* Vol. 9, No. 4, pp. 506–518.

O'Brien, T. G., and Charlton, S. G. (1997), *Handbook of Human Factors Testing and Evaluation,* Erlbaum, Mahwah, NJ.

Overington, I. (1973), *Vision and Acquisition,* Pentech Press, London.

Prabhu, P., and Drury, C. G. (1992), "A Framework for the Design of the Aircraft Inspection Information Environment," in *Proceedings of the 7th FAA Meeting on Human Factors Issues in Aircraft Maintenance and Inspection,* Federal Aviation Administration, Washington, DC.

Rasmussen, J. (1983), "Skills, Rules and Knowledge: Signals, Signs and Symbols, and Other Distinctions in Human Performance Models," *IEEE Transactions on Systems Man and Cybernetics,* Vol. SME-13, No. 3, pp. 257–266.

Reason, J. (1990), *Human Error,* Cambridge University Press, Cambridge.

Rifkin, J. (1995), *The End of Work,* Putnam, New York.

Rogers, A. (1991), "Organizational Factors in the Enhancement of Military Aviation Maintenance," in *Proceedings of the Fourth International Symposium on Aircraft Maintenance and Inspection,* Federal Aviation Administration, Washington, DC, pp. 43–63.

Rosemau, R. D., Nawaz, S., Niu, A., and Wee, W. G. (1999), "Aircraft Engine Blade Colling Holes Detection and Classification from Infrared Images," *Proceedings of SPIE,* Vol. 3586, pp. 85–93.

Rouse, W. B. (1985), "Optimal Allocation of System Development Resources to Reduce and/or Tolerate Human Error," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. SMC-15, No. 5, pp. 620–630.

Rummel, W. D., Hardy, G. L., and Cooper, T. D. (1989), "Applications of NDE Reliability to Systems," *Metals Handbook,* Vol. 17, pp. 674–688.

Saaty, T. L. (1974), "Measuring the Fuzziness of Sets," *Journal of Cybernetics,* Vol. 4, pp. 53–61.

Saaty, T. L. (1977), "A Scaling Method for Priorities in Hierarchical Structures," *Journal of Mathematical Psychology,* Vol. 15, pp. 234–280.

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

Schor, J. B. (1991), *The Overworked American,* Basic Books, New York.

Schoonhard, J. W., Gould, J. D., and Miller, L. A. (1973), "Studies of Visual Inspection," *Ergonomics,* Vol. 16, pp. 365–379.

Sheehan, J. J., and Drury, C. G. (l97l), "The Analysis of Industrial Inspection," *Applied Ergonomics,* Vol. 2, No. 2, pp. 74–78.

Sheridan, J. J. (1987), "Supervisory Control," in *Handbook of Human Factors,* G. Salvendy, Ed., John Wiley & Sons, New York, pp. 1243–1268.

Shindo, W., Wang, E. H., Akella, R., Strojwas, A. J., Tomlinson, W., and Bartholomew, R. (1999), "Effective Excursion Detection by Defect Type Grouping in In-line Inspection and Classification," *IEEE Transactions on Semiconductor Manufacturing,* Vol. 12, No. 1, pp. 3–10.

Siegel, M. Gunatilake, P., and Podnar, G. (1998), "Robotic Assistants for Aircraft Inspectors," *Industrial Robot,* Vol. 25, pp. 389–400.

Siemieniuch, C. E., and Sinclair, M. A. (1999), "Knowledge Lifecycle Management in Manufacturing Organizations," in *Contemporary Ergonomics,* M. A. Hanson, E. J. Lovesey and S. A. Robertson, Eds., Taylor & Francis, London, pp. 322–332.

Sinclair, M. A. (1984), "Ergonomics of Quality Control." Workshop document, *International Conference on Occupational Ergonomics,* Toronto.

Singleton, W. T. (1972), *Man–Machine Systems,* Penguin, Harmondsworth.

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,* FAA Technical Center, Atlantic City, NJ.

Steinmetz, V., and Delwiche, M. (1993), "Machine Vision Techniques for Rose Grading," *Proceedings of SPIE Conference on Vision, Sensors, and Control for Automated Manufacturing Systems,* SPIE, Bellingham, WA, pp. 32–43

Stok, T. (1965), *The Worker and Quality Control,* University of Michigan Press, Ann Arbor.

Svetkoff, D., Rohrer, D., Doss, B., Kelley, R., and Jakimcius, A. (1989), "A High Speed 3-D Imager for Inspection and Measurement of Miniature Industrial Parts," in *SME Vision '89 Conference Proceedings,* (Chicago), pp. 10.1–10.9.

Taguchi, G., and Wu, Y. (1979), *Introduction of Off-Line Quality Control,* Meieki Nakamura-Ku Magaya, Japan.

Taylor, J. C. (1991), "Maintenance Organization," in *Human Factors in Aviation Maintenance Phase 1: Progress Report,* DOT/FAA/AM-91/16, Office of Aviation Medicine, National Technical Information Service, Springfield, VA, pp. 15–43.

Taylor, J. C., and Felten, D. F. (1993), "Designing for Purpose: Understanding the Business We're In," in *Performance by Design,* J. C. Taylor, and D. F. Felten, Eds., Prentice Hall, Upper Saddle River, NJ, pp. 35–50.

Thomas, L. F., and Seaborne, A. E. M. (1961), "The Sociotechnical Context of Industrial Inspection," *Occupational Psychology,* Vol. 35, pp. 36–43.

Thornton, D. C., and Matthews, M. (1982), "An examination of the Effects of a Simple Task Alternation in Quality Control of Cookware," in *Proceedings of Human Factors Association of Canada/Association Canadien d'ergonomie* (*HFAC/ACE*) (Toronto), pp. 134–137.

Umbers, I. G. (1981), "A Study of Control Skills in an Industrial Task, and in a Simulation, Using the Verbal Protocol Technique," *Ergonomics,* Vol. 24, pp. 275–293.

Vardeman, S. V., and Jobe, J. M. (1998), *Statistical Quality Assurance Methods for Engineers,* John Wiley & Sons, New York.

Ventura, J. A., and Chen, J-M. (1994), "Automated CAD-based Vision Inspection," in *Handbook of Design, Manufacturing and Automation,* R. C. Dorf and A. Kusiak, Eds., John Wiley & Sons, New York.

Wang, M. J., and Drury, C. G. (1989), A Method of Evaluating Inspector's Performance Differences and Job Requirements," *Applied Ergonomics,* Vol. 20, No. 3, pp. 181–190.

Wang, M-J., Drury, C. G., and Sharit, J. (1986), "An Application of Fuzzy Set Theory for Evaluation of Human Performance on an Inspection Task," in *Applications of Fuzzy Set Theory in Human Factors,* W. Karwowski and A. Mital, Eds., Elsevier, Amsterdam, pp. 257–268.

Watson, S. R., Weiss, J. J., and Donnell, M. L. (1979), "Fuzzy Decision Analysis," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 19, No. 1, pp. 1–9.

Wenner, C. (1999), "The Role of Instructions in the Aircraft Maintenance Domain," Master's Thesis, State University of New York at Buffalo.

Wenner, C. L, Wenner, F., Drury, C. G., and Spencer, F. (1997), "Beyond 'Hits' and 'Misses': Evaluating Inspection Performance of Regional Airline Inspectors," *Proceedings of the 41st Annual Human Factors and Ergonomics Society Meeting* (Albuquerque, NM), pp. 579–583.

Whitman, M. V. (1999), "Global Competition and the Changing Role of the American Corporation," *The Washington Quarterly, Vol. 22, No. 2, pp. 59–82.*

Wickens, C. D. (1991), *Engineering Psychology and Human Performance,* 2nd Ed., Harper Collins, New York.

Zadeh, L. A. (1965), "Fuzzy Sets," *Information and Control,* Vol. 8, pp. 338–353.