



Studies of Visual Inspection

By J. W. SCHOONAHD, J. D. GOULD & L. A. MILLER

To cite this article: By J. W. SCHOONAHD, J. D. GOULD & L. A. MILLER (1973) Studies of Visual Inspection, Ergonomics, 16:4, 365-379, DOI: [10.1080/00140137308924528](https://doi.org/10.1080/00140137308924528)

To link to this article: <https://doi.org/10.1080/00140137308924528>



Published online: 24 Oct 2007.



Submit your article to this journal [↗](#)



Article views: 214



View related articles [↗](#)



Citing articles: 7 View citing articles [↗](#)

Studies of Visual Inspection

By J. W. SCHOONARD, J. D. GOULD and L. A. MILLER

IBM Thomas J. Watson Research Center, Yorktown Heights, New York,
10598, U.S.A.

This paper describes the results of four experiments in a series aimed at understanding and improving visual inspection in general and of small integrated circuits (i.e. 'chips') in particular. Stimuli consisted of chips that, although electrically sound, contained visual anomalies. The first experiment found that the modal duration of eye fixations of trained inspectors was about 200 msec. The most accurate inspectors made the fewest eye fixations and were the fastest. The second experiment evaluated the performance of inspectors at one of the many sequential stages of chip inspection and found that 23% of the chips containing anomalies were accepted, whereas only 2% of the chips without anomalies were rejected. When the same chip was judged more than once by an individual inspector the consistency of her judgment was very high whereas the consistency between inspectors was somewhat less. The third experiment showed that variation by a factor of six in inspection speed led to variation of less than a factor of two in inspection accuracy. The fourth experiment showed that inspection via a ground glass screen is only a little worse than the usual method of looking through a binocular microscope. This was true even though the inspectors had no previous experience with the screen.

1. Introduction

Visual inspection pervades the lives of all people today. From poultry, meat, and fish inspection, to drug inspection, to medical X-ray inspection, to production line inspection, to photo interpretation, the consequences of inspection directly affect people's lives through their effects on the quality and performance of goods and services.

In general, human visual inspection is characterized by three facts. First, inspectors look for many things at once. Second, they must do this very fast. Third, they are not very accurate.

Although human visual inspection is widespread and economically important (Harris and Chaney 1969), there exists little theoretical understanding of it. Fundamental laboratory studies, such as the influential visual search studies of Neisser *et al.* (e.g. 1963), have limited application to most actual inspection situations because they use discrete targets (e.g. alphabetic characters) on homogeneous backgrounds, whereas most inspection situations consist of poorly defined targets on non-homogeneous backgrounds (e.g. chest X-rays). Another characteristic of human visual inspection is that it is sufficiently complicated to rule out automatic inspection with today's state-of-the-art.

An integrated circuit chip is a tiny slice of silicon (often less than 0.1 inch on a side) containing circuits comprised of many transistors, resistors, and diodes. The manufacture of integrated circuit chips requires many stages of visual inspection. In one stage at IBM, inspectors using a microscope cull out chips which, although electrically sound, have visually detectable anomalies. These anomalies will be referred to as 'targets', chips containing them as 'positive' chips, and chips not containing them as 'negative' chips. In

actual practice the presence of such targets is neither a necessary nor sufficient condition for chip malfunction. The task of the inspector is to make a rapid visual scan of each chip, searching for six or more different types of targets at once, and to classify the chip as positive or negative. This situation offered an opportunity to investigate a problem with direct impact upon manufacturing while serving as a vehicle for fundamental work related to a variety of inspection situations.

2. Experiment I

This experiment investigated the way trained inspectors visually scan chips. The eye movements of eight women inspectors were recorded while they inspected coloured slides of one family of chips. These inspectors had inspected this family of chips for 3 to 13 months. Figure 1 is an example of two different types of chips (or part numbers) within this family. Inspectors looked for eight different targets as described below.

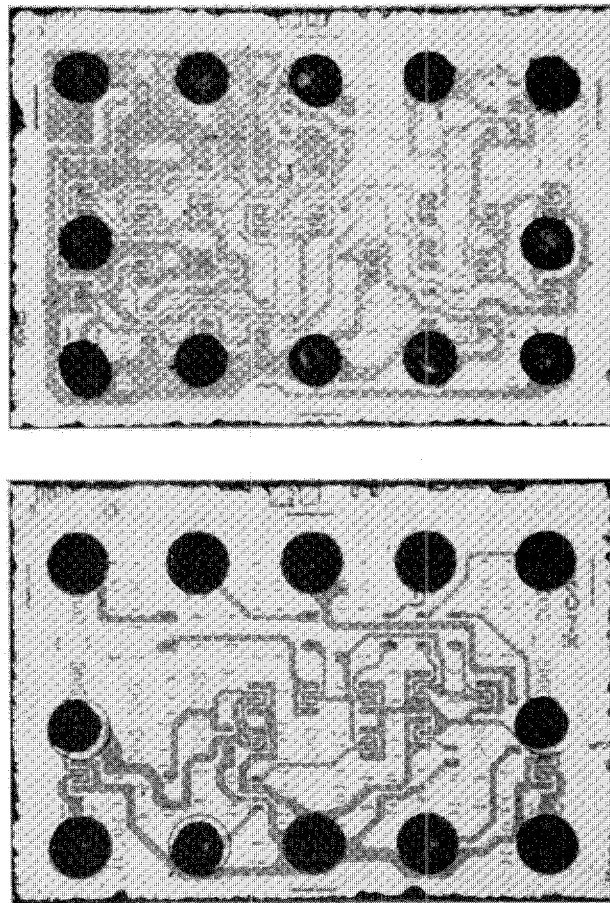


Figure 1. Examples of MST-47 chips enlarged approximately 50 times. These chips are roughly 40×60 mils. The 12 black spots around the periphery are hemispheres. The dark grey pattern represents the features.

In visual scanning, the eyes do not move smoothly, but rather they stop or fixate, and then jump or move ballistically to the next point of fixation. There is little or no useful perception between jumps; all significant information is obtained during a fixation. Eye movements during scanning can, therefore, be described in terms of the number, duration, and location of eye fixations.

Several specific questions were studied. First, is there a difference in the way accurate and inaccurate inspectors look for targets (cf. Gould 1969, for a review of eye movements during visual search)?

Second, do inspectors make many, brief eye fixations or do they make fewer but longer fixations? Boynton (1960) cites findings which suggest that in a visual search task the more fixations a searcher makes per unit time (within limits), the more accurate his performance.

Third, where on the chips are most of the eye fixations occurring? Results of radiologists scanning X-rays (Llewellyn-Thomas and Lansdown 1963), operators scanning radarscopes (White and Ford 1960), and operators scanning maps (Enoch 1960) all show that some areas tend to be relatively neglected.

Fourth, do the sequences of eye fixations follow a set pattern for all chips, or are the optical characteristics of different types of chips so compelling that they determine how inspectors will scan them?

Fifth, is there a systematic relationship between how inspectors visually scan chips, their accuracy of detecting targets, and their speed of detection?

2.1. Method

Eight experienced female inspectors served as subjects. Each subject was instructed to inspect at her normal rate, to reject a slide if it had any of the eight targets, and to accept it if it did not. The targets were: (A) insufficient clearances at the edge of the chip; (B) mechanical damage altering the appearance of certain surface features of chips; (C) a significant reduction in size of or the total absence of one or more hemispheres (similar to tiny drops of solder) on the chip; (D) improperly located hemispheres; (E) flaws in the coating on the chip; (F) two adjacent features touching each other where they should not; (G) localized decrease in the width of a feature; (H) the presence of a chip with a different feature pattern than that which the inspector is inspecting.

Each subject was shown 168 35-mm coloured slides of chips. A slide subtended 18° vertically at an inspector's eye, which is somewhat less than the 23° that chips ordinarily subtended in a microscope. Twenty per cent of the slides contained targets. Positive and negative slides were arranged in a random sequence. Most of them were of one kind of chip; a few were of five other kinds of chips so as to assess detection of target H.

Eye movements were recorded by a system that produced filmed records of the stimulus field and the positions of the inspector's fixations within it. This system, described elsewhere (Gould and Peeples 1970) is a modified closed-circuit television, corneal reflectance, eye-marker system (Mackworth and Mackworth 1958). The duration of each fixation (recorded at 19 frames per second) was read from filmed records.

2.2. Results and Discussion

2.2.1. Eye-movements

The overall mean scan time per chip for a single subject varied from 2.7 sec. to 5.4 sec.

Most of the eye fixations occurred inside the boundary outlined by the hemispheres (cf. Figure 1), which is the area most likely to contain targets. The edges of a chip were relatively neglected even though there were some targets (e.g. target A) there. Consequently, these targets sometimes had to be detected from an image formed in the periphery of the retina, which has poorer acuity than does the fovea or central retina. Eye fixations were most frequent in areas of the chips where the features were the most complex. The frequency of re-fixation of any particular area was low.

Analysis of the scanning patterns (sequential position of fixations) was inconclusive. Scanning patterns of each individual subject showed a great deal of variability. Since eye scanning patterns are influenced by the patterning of the visual field (cf. Williams 1967), the variety of patterns seen may have obscured some consistency. There were no obvious differences among inspectors.

The distribution of fixation durations for chip inspection had a mode of approximately 200 milliseconds and was positively skewed. As can be seen in Figure 2, these are short fixation durations compared to those obtained by Ford *et al.* (1959) in a simple detection task and those reported by White and Ford (1960) for a radarscope application. This finding of short fixation durations for trained chip inspectors suggests, in view of Boynton's (1960) findings, that inspectors are using a better strategy than if they made fewer but longer eye fixations. Most of the variation in fixation durations was due to

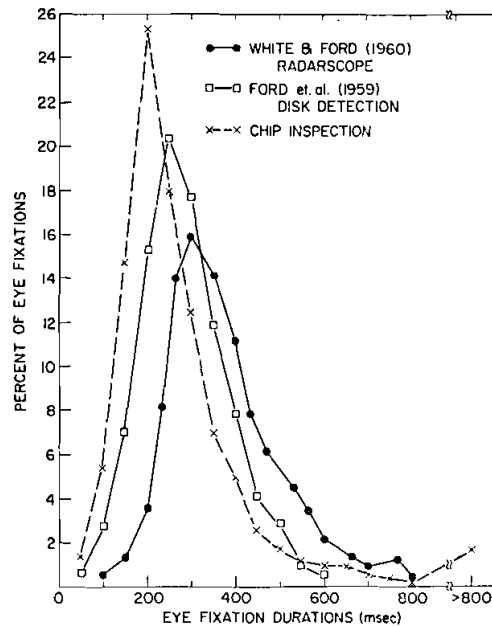


Figure 2. Distribution of eye fixations for three different search tasks.

variations 'within chips' (mean square within chip variance = 34 msec.) rather than to variations from one chip to another (mean square between chip variance = 3.6 msec). This large within chip variation reflects the large decision time component present for some fixations on many of the chips, as shown in Figure 2 by the 4% of eye fixations that exceeded 600 msec. Fixation of a suspicious area of a chip to determine whether or not it contained a target seemed to require considerably more time than did the fixations taken to locate the area in the first place.

2.2.2. Comparison of accuracy and eye-scanning parameters

It was possible to use the data of only five subjects to compare accuracy and eye scanning parameters. Table 1 shows that the two subjects who made the least errors were also the fastest inspectors, having an average scan time of 2.6 seconds. Error rate is the per cent of slides misclassified. (These error rates are higher than those obtained on chips themselves because of loss of resolution and loss of subtle colour and depth cues.) The two subjects who made the most errors, although neither was the slowest inspector, had an average scan time of 4.0 seconds. The difference in scan time is primarily due to a difference in mean number of fixations per chip (9.2 *v* 14.8) rather than due to any differences in fixation durations. These results, although limited by a sample of five subjects and a small range of inspector accuracies, suggest that the fastest inspectors also tend to be the most accurate inspectors. This is not an artifact of experience, since the subject with the least amount of experience was the most accurate.

Table 1. Comparison of scanning parameters for inspectors ranked according to accuracy

Inspector	Overall error rate	Overall mean scanning time	Mean no. fixations per chip	Modal fixation duration in 50 millisecond intervals
1	10%	2.7 sec	9.4	250
2	12%	2.5 sec	9	200
3	13%	5.4 sec	15.3	200
4	15%	3.6 sec	14.4	200
5	16%	4.4 sec	15.2	200

These results, which are in contrast to the negative correlation usually found between speed and accuracy in behavioural laboratory experiments (Fitts 1966), suggest that good inspectors are characterized by relatively high accuracy and relatively high speed, and make many brief eye fixations (as opposed to fewer longer ones) during the time they have to view the chip.

3. Experiment II

This experiment investigated the detailed performance of trained inspectors on actual chips in a throughput situation, i.e. with an emphasis upon speed. Several different lines of evidence led to the prediction that a significant number of positive chips would be accepted. Niesser *et al.* (1963) have shown that well practiced people who visually search, as fast as they can, lists of alphabetic characters for more than one target character miss about 25% of all targets. The results of experiments on simulated circuit board design (Badalamente

and Ayoub 1969), disc and line inspection (Sadler 1966, Teel *et al.* 1968) and geometric form inspection (Harris 1968) also show that inspectors miss at least 25% of all targets when searching for several targets at once, or when inspecting rapidly (Lion *et al.* 1968). Finally, informal observations in a variety of factory inspection situations suggested that pressures to meet manufacturing quotas can become intense enough to affect eventually even the decision processes of inspectors.

3.1. Method

Nine female inspectors, each with a minimum of 6 months experience inspecting chips served as subjects. Subjects performed essentially the same task as in Experiment 1, except they viewed chips directly through a microscope, as is normally done.

On four consecutive half-days, each subject was presented 198 chips of a particular kind. A special apparatus was built that allowed the presentation of chips of known characteristics in a pre-determined sequence. The apparatus consisted of a Leitz metallurgical microscope fitted with a Leitz brightfield 10 $\times$ objective (N.A. = 0.18) and Leitz 10 $\times$ periplanatic oculars. The microscope was mounted over a micro-positioner on which was mounted an X-Y stage driven by two stepping motors. Interchangeable cartridges, each holding a 33 $\times$ 6 matrix of 198 chips could be attached to the stage. A manually operated micro-positioner allowed the subject to move a chip slightly from side to side so as to see all parts of it, just as is necessary on some of the regular inspection units. The subject's accept and reject buttons activated the stepping motors which advanced the X-Y stage from chip to chip. The time for the apparatus to position a chip was set at 1 sec for these experiments. For each chip, the accept-reject decision of each subject as well as the latency of judgment was automatically recorded.

A pacing device was used to insure that each subject inspected at her normal rate, which was a few seconds per chip. One pointer on this device moved at a rate proportional to the particular subject's normal speed. Another pointer moved incrementally, one step for each response. When the two pointers were not aligned, a subject knew she was falling behind or getting ahead, depending upon the direction of misalignment.

A chip sample was prepared by selecting chips from all target categories. Eighteen per cent of the chips contained targets. Care was taken to select these chips so that each target category was represented with a range from obvious to subtle cases. Following completion of the experiment, the chips were re-classified to assess the reliability of the initial experimenter-determined classification as well to determine whether any were damaged during the experiment (one was damaged). Most of the chips (98.2%) were classified the same way both times.

3.2. Results and Discussion

The accuracy of classification for the 792 decisions made by each subject is summarized in Table 2. Overall, approximately 6% of the judgments were erroneous. Only about 2% of all negative chips were rejected, which seems to be a reasonable performance level given the inspection rate of only a few

seconds per chip. However, about 23% of all positive chips were accepted, which is indicative of the inherent difficulty of such inspection tasks. In general, the accuracy of performance was similar to that found in other visual inspection situations previously mentioned.

Table 2. Classification of chips at normal inspection rates

		True state		
		Negative	Positive	
Inspectors decision	Accept	5743	291	6034
	Reject	129	965	1094
		5872	1256	7128

It appears that the payoff matrix of the subjects emphasized minimizing the error of rejecting negative chips, even though the avowed purpose of visual inspection is to cull out positive chips. It is well known that the error of missing a target is greater than the error of reporting a false positive, even when people know that a target is present on each stimulus (Neisser *et al.* 1963).

The subject's mean acceptance rate (84.7%) was about the same as the percentage of negative chips in their input (82.4%). The similarity of these two figures cannot simply reflect intentional probability matching (cf. Thomas and Legge 1970) on the part of the subjects as the probability of a target was not specified. This finding most likely resulted from the fact that 94% of all judgments were correct, and therefore, the two percentages approximated each other.

The first row of Table 3 shows the errors of accepting positive chips for each target type. The process of correctly rejecting a positive chip involves both detection and judgment. Relatively few A targets (6.4%) were erroneously accepted. These targets were quite visible and required little judgment. The eye-movement results of Experiment 1 suggest that they were detected with peripheral vision. About three times as many B targets (18.8%) and about four times as many C and D targets (24.1%) were erroneously accepted. This appears to reflect an increasing difficulty of judgment. E targets (40.2% error) not only presented problems of judgment as to whether they met the target criteria, but they were often difficult to detect. The error rates for F and G targets were highest (56%). These targets were simply hard for subjects to detect. Once detected, they required little judgment.

The differences in accuracies among the nine inspectors, contrary to general assumption, were not great. The distribution of errors of rejecting negative chips had a mean of 2.2%, and a standard deviation of 1.2%. The distribution of errors of accepting positive chips had a mean of 23.1% and a standard deviation of 5.6%. The consistency (without regard to accuracy) of judgments of individual subjects are reported in Experiment III.

There were no changes in inspection accuracy within each half-day session or over the four half-days.

Some of the chips used in this experiment (27 of 990) were marginal in the sense that it was agreed *a priori* that their classification was dubious. Twenty of the 27 were initially classified as borderline positive chips and the remainder

were initially classified as borderline negative chips. When these 27 chips were eliminated from the analyses of the results, inspection performance improved considerably. Errors on negative chips decreased from 2.2% to 1.8% and errors on positive chips decreased from 23.2% to 16.5%. The first row of Table 3 contains these values in parentheses. The errors on each target type were reduced about 25–35% of their original values, except on F and G targets. The latter still had an error rate in excess of 50%.

An important point is that the exclusion of a small percentage of the stimuli (2.7%), designated as borderline independent of performance results, led to a distinct improvement in accuracy (29% reduction in overall error rate). As a general question, it is not known what impact acceptance of marginal items has upon subsequent performance of products in real-life. It obviously depends upon the stringency of inspection requirements in quality control. If marginal stimuli can be accepted with little deleterious effect, then the meaningful performance figures would be those in parentheses in Table 3. On the other hand, the various experimental demands of the study may have tended to reduce error rates relative to what can be expected in practice. In either case, the absolute accuracy values obtained in this and other experiments do not reflect the final quality control levels obtained in manufacturing plants because chips are not an end product, and many products are inspected more than once at different stages of assembly.

There are several possible factors that might lead to an improved accuracy of chip inspection. One of these is the way in which a target is defined. This is illustrated by the definition of target C, which states that if a specific proportion of the volume of any hemisphere is missing relative to that of other hemispheres on the chip, then the chip is positive. In an informal demonstration, black and white photographs of 38 chips were given to nine subjects who were told to inspect them for target C only. All chips had borderline target C's, but some of these did not meet the reject criterion. Although only 16% of the photographs representing negative chips led to a reject decision, 86% of the photographs representing positive chips led to an accept decision. These error rates may be inflated due to the difficulty of inspecting via photographs. The point of interest, however, is that an analysis showed subjects were basing their decisions on the diameter of the hemispheres, without taking into account the fact that a reduction in diameter of a given proportion represents a larger proportional reduction in volume.

4. Experiment III

Another factor that was thought to affect inspection accuracy was the average rate of inspection. Given the complex nature of chips and targets, the results of related experiments, and the fact that an inspector must look for several targets at once, it did not seem surprising to us that nearly one in four positive chips was erroneously accepted. We were not at all sure, however, whether reasonable decreases in inspection rate would result in proportional gains in inspection accuracy, even though a speed/accuracy tradeoff typically occurs in laboratory experiments. We reasoned that if an experiment showed that a decrease in the rate of inspection led to a distinct improvement in the accuracy of inspection then this would provide an immediate recommendation

on how to improve inspection accuracy. If, on the other hand, this result did not occur, then a more radical redesign of inspection (e.g. training, optics, and procedures) would have to be studied to improve inspection accuracy. Accordingly, the purpose of this experiment was to assess the effect of inspection rate on the accuracy of inspection.

4.1. Method

The same nine subjects and the same 990 chips, 198 of each of five part numbers, were used as in Experiment II. Eighteen per cent of the chips were positive. In addition to each subject working at her normal speed (reported as Experiment II), each subject worked at 1.5, 2.0, and 3.0 times slower than her own normal speed as well as 0.5 times as slow (i.e. twice as fast). Each subject was run in four half-day sessions. In each session a subject looked at five different types of chips, one at each of the five different speeds. The first five subjects were assigned to a chip type and speed combination for the first half-day session on the basis of a 5×5 Greco-Latin square in which each row was the sequence of chip type and speed combination received by a particular subject. To obtain an estimate of the reliability (consistency) of judgments, each subject received the same chip type and speed combination on the second and fourth half days. Different Greco-Latin squares were used for the other two sessions, each square balancing the order of presentation of both chip type and speed. Thus each subject inspected 792 chips at each speed. It was intended to replicate each square twice. However, it was only possible to obtain nine subjects and thus the last row of one square on each half day was not used.

The pacing device described in Experiment II was used to control the rate of inspection. Subjects were told at the beginning of each speed condition whether that speed was a fast or a slow one and to inspect as accurately as possible in the time allotted.

4.2. Results and Discussion

The percentage of errors at each of the five speeds is shown in Table 3. Subjects rejected about the same number of negative chips regardless of inspection speed (Friedman test, $p > 0.05$; Hays 1963). Fewer positive chips were accepted at the slower speeds ($p < 0.001$). When subjects slowed down, the percentage of positive chips accepted was reduced from 23.2% to 16.5%. Given that they were going slower, it did not matter much whether they went 1.5, 2.0 or 3.0 times slower ($p > 0.05$). The net effect of slowing the rate of inspection was to reduce the percentage of positive chips in the output of visual inspection from 4.8% (at the normal rate) to 3.6% (mean of the three slow speeds). This is a 25% increase in quality of inspection with increases in inspection time ranging from 50% to 200%. Thus, although, inspection accuracy improved, it did not improve in proportion to the decrease in rate of inspection.

Subjects reported that they found the task tedious and boring at the slower inspection speeds. This finding may reflect their reaction to a departure from what they were trained and accustomed to doing. Similar reports were obtained in another experiment (Schoonard and Gould 1973) with much less experienced subjects.

Table 3. Percentage of errors for different targets (A through G) at each inspection rate. Rate N is based on the normal inspection rate for each subject. Figures in parentheses are based on analyses which excluded borderline chips

Inspection rate	A	B	C and D	E	F and G	All positive chips	All negative chips
N	6.4 (4.4)	18.8 (12.5)	24.1 (18.5)	40.2 (29.5)	56.0 (52.6)	23.2 (16.5)	2.2 (1.8)
0.5	5.9 (5.7)	23.0 (17.4)	23.3 (18.5)	51.8 (43.8)	86.4 (80.0)	29.6 (23.3)	1.8 (1.4)
1.5	3.4 (2.6)	14.3 (9.1)	14.0 (9.3)	28.1 (17.9)	46.0 (38.0)	16.5 (10.6)	2.1 (1.7)
2.0	2.9 (2.0)	12.8 (8.8)	22.6 (17.6)	34.1 (22.2)	46.5 (38.0)	18.0 (12.0)	3.1 (2.8)
3.0	4.6 (2.8)	12.4 (7.2)	17.4 (12.0)	28.3 (17.3)	42.6 (35.0)	16.4 (10.2)	2.6 (2.2)
Weighted mean of all rates	4.6 (3.5)	16.3 (11.0)	20.3 (15.2)	36.6 (26.2)	55.7 (48.7)	20.8 (14.5)	2.4 (2.0)
Weighted mean of three slow rates	3.6 (2.5)	13.2 (8.4)	18.0 (13.0)	30.2 (19.1)	45.0 (37.0)	17.0 (10.9)	2.6 (2.2)

Doubling the normal rate of inspection (0.5 condition) led to a significant increase in the percentage of positive chips accepted from 23.2% to 29.6% ($p < 0.05$). The net effect was to increase the percentage of positive chips in the output of visual inspection from 4.8% (at the normal rate) to 6.1% (at twice the normal rate). Nevertheless, the fact that some targets (e.g. A and C) were inspected just as accurately at a higher than normal speed has direct implications for a serial inspection system in which each subject might look for only a subset of targets. Each subset would be looked for with microscopy techniques optimized for that subset and perhaps at different speeds. Yonas and Pittenger (1970) have shown that when subjects search for multiple targets, each individual target may have a different speed/accuracy trade-off.

The differences between the error rate at the normal speed and the mean error rate of the three slower speeds (cf. Table 3) was computed for each target. Each difference was expressed as a percentage decrease from the error rate at the normal speed. The largest percentage decrease in error rate was obtained on the easiest targets, i.e. A (44%). Foveal fixation of the edge of the chip is unlikely at normal speeds (cf. Experiment I) whereas at slower speeds there is ample time to fixate all portions of the chip. Targets that were the most difficult to inspect (F and G) showed the least improvement in performance with a decrease in inspection rate.

Inspection performance appears better by about the same amount at all speeds when borderline chips (i.e. the 27% cited above) are excluded (cf. parentheses in Table 3). Elimination of these marginal chips did not affect the rank order of inspection accuracy across target categories.

Figure 3 shows accuracy as a function of rate of inspection for the individual subjects. The differences in overall accuracy among the nine subjects were relatively small. For each, the errors on positive chips tended to be inversely related to rate of inspection. For seven of the subjects, errors on negative chips were directly related to each one's overall rejection rate. In addition, there was a tendency to reject more chips overall at the slower speeds (Friedman test; $p < 0.01$).

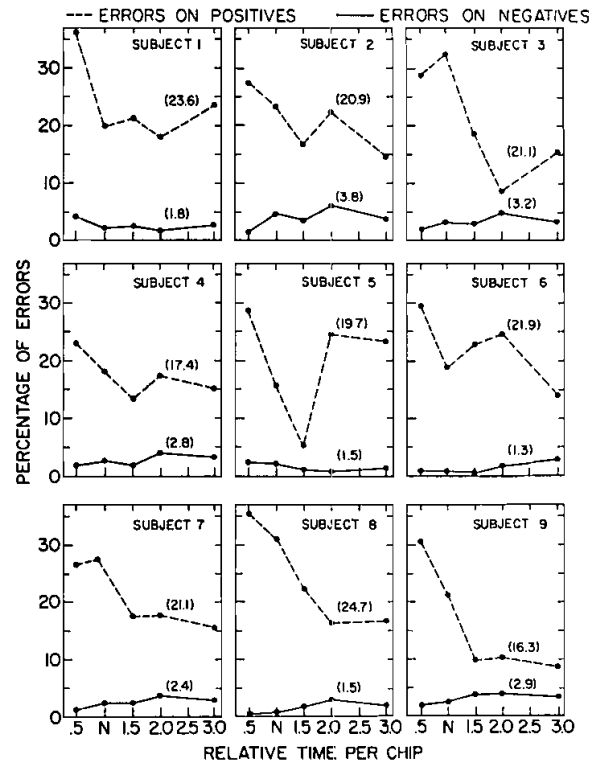


Figure 3. Percentage of errors for each inspector at each inspection rate. Mean of all inspection rates shown in parentheses.

Inspection accuracy was better on some types of chips than on others ($p < 0.05$). The easiest chip pattern was inspected almost twice as accurately as the most difficult. This suggests the possibility of predicting inspection accuracy for a particular type of chip prior to its manufacture by measuring the detection of a target placed on a picture of the chip pattern.

The data of the second and fourth half-day sessions were used to calculate both the 'within-subject' consistency or reliability (i.e., the consistency of judgments of a particular subject on the same chip when seen more than once) and the 'between-subject' consistency or reliability (i.e., the consistency of judgments across all subjects on the same chip).

The within-subject consistency was extremely high. On the average, each subject judged 97% of the chips the same way in classifying them twice. The range of within consistencies for individual subjects was 0.95 to 0.99.

To calculate the between-subjects consistency on each chip, the most frequent (of 3 possible) outcomes (without regard to accuracy) of the pair of judgments by each subject on that chip was determined. Then the between-subject consistency was expressed as a ratio of the maximum number of subjects whose judgments were in agreement on that chip to the total number of subjects (nine). The cumulative percentage for the 990 chips is shown in Figure 4. As can be seen, all nine subjects were in agreement on 79% of the chips. Eight of nine subjects were in agreement on another 10% of the chips, yielding a cumulative percentage of 89% for eight or nine subjects in agreement. The

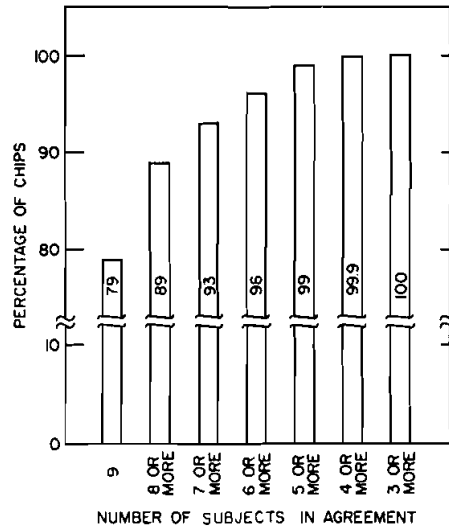


Figure 4. Between inspector consistency expressed as the percentage of chips classified the same way by x inspectors, where x varies from three to nine.

weighted average of this distribution, 95%, summarizes the between consistency. This result cannot be accounted for simply in terms of individual acceptance rates (χ^2 test; $p < 0.001$).

Inconsistency between subjects can arise from several sources. Perhaps some subjects had different subjective criteria, concentrated primarily on certain targets, had different expectations as to the probability of occurrence of particular targets, perceived the payoff matrix to be different, or differed in visual capacities (acuity, depth perception, etc.). Training techniques might reduce some of this inconsistency.

The between-consistency results have a direct implication for one possible method of improving visual inspection, i.e., a method in which each chip is independently viewed by more than one subject. The assumption underlying this method is that each subject is independent of the other one and that the second subject detects a significant amount of what the first subject misses. For example, if each subject misses $q\%$ of positive chips, then two subjects together would miss only $q^2\%$ of them. The present between-inspector consistency results are, however, too high to support this independence assumption. The question was asked as to whether most inspection errors occurred on only a few chips or whether they were spread across most of the chips. For some targets most of the inspection errors occurred on a small per cent of the chips. For other targets, however, errors were spread across a great percentage of chips containing them.

5. Experiment IV

Besides inspection rate, another factor that might affect inspection accuracy is the optical condition under which inspection occurs. This experiment compared the accuracy of visual inspection on a viewing screen with the accuracy through a binocular microscope (Experiment II). The viewing screen has potential advantages in terms of a possible reduction in inspection fatigue (e.g. from eye strain or lack of freedom of head movement), an opportunity

for simultaneous viewing by more than one person (for training and consultation), and facilitation of the use of overlays and reticles. On the other hand, a primary drawback of a viewing screen is the loss of image resolution.

5.1. Method

Five of the nine subjects who participated in Experiments II and III were available for this experiment. The ground-glass viewing screen, a standard attachment to a Leitz metallurgical microscope, was 155 mm in diameter. It was a directional screen in that it directed a relatively high proportion of its light through a restricted viewing angle. The same $10\times$ objective (N.A. = 0.18) as used in Experiments II and III, together with a $10\times$ lens in the screen assembly, provided the same $100\times$ magnification as in Experiments II and III. In comparing the resolution of the screen to the microscope, we determined that human observers were able to resolve 287 lines pairs per millimeter on the screen, compared with 406 line pairs on the microscope. A slight mechanical scan was necessary in order to see the entire chip, just as in Experiments II and III. Each of the five chip types (198 chips of each) used in Experiments II and III were used here. Eighteen per cent of the chips were positive. All subjects worked at their normal inspection rate. A 5×5 Latin square was used to present the part numbers in a different order to each subject.

5.2. Results and Discussion

Errors on positive chips were greater on the display screen (29%) than in the microscope (23%). Errors on negative chips were the same (about 2%) in both cases.

Table 4 shows inspection performance on each target for both the display and the microscope. With the exception of A targets, performance was

Table 4. Percentage errors on positive chips for inspection via binocular microscope and for inspection via display screen at normal inspection rates. Figures in parentheses are based on analyses which excluded borderline chips

Defect classification	Microscope	Display	Difference (microscope - display)
A	6.4 (4.4)	4.2 (3.7)	+2.2 (+0.7)
B	18.8 (12.5)	20.4 (16.4)	-1.6 (-3.9)
C and D	24.1 (18.5)	36.2 (32.0)	-12.1 (-13.5)
E	40.2 (29.5)	53.5 (45.8)	-13.3 (-16.3)
F and G	56.0 (52.6)	77.5 (72.7)	-21.5 (-20.1)
Weighted mean of all targets	23.2 (16.5)	28.8 (23.4)	-5.6 (-6.9)

poorer on all targets when the display was used. Those targets which had the highest error rates for the microscope showed the greatest decrement in performance when inspection was done via the viewing screen.

The somewhat poorer performance on the viewing screen is open to at least two interpretations. First, it may be due to the loss of resolution. The targets that seem to require the most resolution show the greatest decrement in performance on the screen. Second, the decrement in performance may have arisen from lack of subjects' experience in using the screen. The subjects had

been trained on a microscope and had inspected with it exclusively over a period of months. In this respect it should be noted that only a modest increase in accuracy with the screen, over that found in this experiment, would suggest that it is as good as the binocular microscope. Thus, it is not possible to conclude unequivocally that the display screen is inferior to the microscope for the task performed by the subjects.

6. Conclusion

Although the variables studied in this series of experiments were shown to be relevant to inspection accuracy, it is concluded that even if the optimal level is selected for each variable the accuracy of inspection will not go up dramatically. It appears that if substantial improvement in human inspection accuracy can occur it will depend upon the study of three basic aspects of the inspection system: training, inspection procedures, and apparatus (optics, lighting, etc.). To this end, the effects of restricting the inspectors' field of view has been evaluated (Schoonard and Gould 1973) and a variety of common microscopy techniques have been experimentally evaluated to determine which ones best enhance each target.

We thank the Fishkill laboratory of the Components Division of IBM under the direction of Mr. Theodore J. Lakoski for jointly supporting this work. We thank Mr. Emil Cohen, IBM Research, for his valuable assistance in helping debug our data analysis computer programs, Mr. Kurt Leimbrock, IBM Research, for technical work on the experimental apparatus used in Experiments II-IV, and Mr. Peter Moczarski, a summer high school student, for help with data analysis in Experiment I.

Cet article rapporte les résultats de quatre expériences dont le but était de mieux comprendre et d'améliorer l'inspection visuelle, en général et celle des circuits intégrés de petites dimensions (les 'chips'), en particulier. Les stimulus étaient constitués de 'chips' qui, bien que normaux du point de vue électrique, présentaient quelques anomalies à la vue. Au cours de la première expérience, on a montré que la durée de la fixation oculaire, pour un contrôleur entraîné était d'environ 200 ms. Les contrôleurs les plus efficaces présentaient le moins de fixations oculaires et étaient les plus rapides. Dans la deuxième expérience, on a évalué la performance des contrôleurs durant l'une des nombreuses étapes du contrôle des 'chips'. Il s'est avéré que 23% des 'chips' présentant des défauts étaient acceptés, alors que 2% seulement des 'chips' sans défauts étaient rejetés. Lorsque le même 'chip' était examiné plus d'une fois par le même contrôleur, la cohérence entre ses jugements était très élevée, alors que la cohérence entre les contrôleurs était moindre. La troisième expérience a montré qu'en faisant varier dans un rapport de 6 le facteur de vitesse, le facteur de précision variait à peine dans un rapport double. La quatrième expérience a montré que le contrôle au moyen d'un écran en verre dépli était un peu moins bon que celui effectué au moyen du microscope binoculaire, même si le contrôleur n'avait pas une expérience préalable avec l'écran.

Diese Arbeit beschreibt die Resultate von vier Versuchsserien, die beabsichtigen, visuelle Inspektion im allgemeinen und an kleinen integrierten Schaltkreisen ('Chips') im besonderen zu verstehen und zu verbessern. Die Reize bestanden aus Chips, die, obwohl elektrisch intakt, visuelle Anomalien enthielten. Der erste Versuch zeigte, dass die Moduldauer der Augenfixation trainierter Inspektoren ungefähr 200 msec betrug. Die genauesten Inspektoren machten die wenigsten Augenfixationen und waren die schnellsten. Der zweite Versuch bewertete die Leistung der Inspektoren bei einer der vielen aufeinanderfolgenden Chip-Inspektionen und fand, dass 23% der anomalen Chips akzeptiert wurden, während 2% der einwandfreien Chips zurückgewiesen wurden. Wurde derselbe Chip mehr als einmal von demselben Inspektor geprüft, so war die Beständigkeit des Urteils sehr hoch, während die Beständigkeit verschiedener Inspektoren etwas niedriger lag. Der dritte Versuch zeigte, dass eine Änderung der Inspektionsgeschwindigkeit um einen Faktor sechs zu einer Änderung der Inspektionsgenauigkeit um den Faktor zwei führte. Der vierte Versuch ergab, dass die Inspektion durch einen angerauhten Glasschirm nur wenig schlechter ist als die übliche Methode des Sehens durch ein binokulares Mikroskop. Das galt obwohl die Inspektoren keine vorausgehende Erfahrung mit dem Glasschirm hatten.

References

- BADALAMANTE, R. V. and AYOUB, M. M., 1969, A behavioral analysis of an assembly line inspection task. *Human Factors*, **11**, 339-352.
- BOYNTON, R. M., 1960, Summary and Discussion. In *Visual Search Techniques*. (Edited by A. MORRIS and E. P. HORNE) (Washington, D.C.: NATIONAL ACADEMY OF SCIENCES).
- ENOCH, J. W., 1960, Natural tendencies in visual search of a complex display. In *Visual Search Techniques*. (Edited by A. MORRIS, and E. P. HORNE) (Washington, D.C.: NATIONAL ACADEMY OF SCIENCES).
- FITTS, P. M., 1966, Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, **71**, 849-857.
- FORD, A., WHITE, C. T., and LICHTENSTEIN, M., 1959, Analysis of eye movements during free search. *Journal of the Optical Society of America*, **49**, 287-292.
- GOULD, J. D., 1969, Eye movements during visual search. *IBM Research Report*, RC 2680.
- GOULD, J. D., and PEEPLES, D. R., 1970, Eye movements during visual search and discrimination of meaningless, symbol, and object patterns. *Journal of Experimental Psychology*, **85**, 51-55.
- HARRIS, D. H., 1968, Effects of defect rate on inspection accuracy. *Journal of Applied Psychology*, **52**, 377-379.
- HARRIS, D. H. and CHANEY, F. B., 1969, *Human Factors in Quality Assurance* (New York: WILEY).
- HAYS, W. L., 1963, *Statistics for Psychologists* (New York: HOLT, RINEHART and WINSTON).
- LION, J. S., RICHARDSON, E. and BROWNE, R. C., 1968, A study of the performance of industrial inspectors under two kinds of lighting. *Ergonomics*, **11**, 23-34.
- LIEWELLYN-THOMAS, E. and LANSDOWN, E. L., 1963, Visual search patterns of radiologists in training. *Radiology*, **81**, 288-292.
- MACKWORTH, J. F. and MACKWORTH, N. H., 1958, Eye fixations recorded on changing visual scenes by the television eye-marker. *Journal of the Optical Society of America*, **48**, 439-445.
- NEISSER, U., NOVICK, R. and LAZAR, R., 1963, Searching for ten targets simultaneously. *Perceptual and Motor Skills*, **17**, 955-961.
- SADLER, E. E., 1966, The effect of an overlay field on visual inspection judgments. *Paper presented at Western Psychological Association Meeting, Long Beach, California*.
- SCHOONARD, J. W. and GOULD, J. D., 1973, Field of view and target uncertainty in visual search and inspection. *Human Factors*, **15**, 33-42.
- TEEL, K. S., SPRINGER, R. M., and SADLER, E. E., 1968, Assembly and inspection of micro-electronic systems. *Human Factors*, **10**, 217-224.
- THOMAS, E. A. C., and LEGGE, D., 1970, Probability matching as a basis for detection and recognition decisions. *Psychological Review*, **77**, 65-72.
- WHITE, C. T. and FORD, A., 1960, Eye movements during simulated radar search. *Journal of the Optical Society of America*, **50**, 909-913.
- WILLIAMS, L. G., 1967, The effects of target specification on objects fixated during visual search. *Acta Psychologica*, **27**, 355-360(a).
- YONAS, A. and PITTENGER, J., 1970, Learning to look for four things at once: A test of simultaneous processing. *Paper presented at the Forty-First Annual Meeting of the Eastern Psychological Association held in Atlantic City, New Jersey*.