EMPIRICAL BASIS FOR AN OCCUPATIONAL VISION STANDARD

Bettina L. Beard¹, Karen M. Jones², Cynthia Chacon² and Albert J. Ahumada¹, Jr.

¹NASA Ames Research Center, Moffett Field, CA 94035-1000,
²San Jose State University Foundation, San Jose, CA

ABSTRACT

There are no government mandated vision standards for aviation maintenance inspectors. Empirically derived vision standards for other occupations cannot be extended to this very different occupation. We apply a psychophysical human-in-the-loop methodology toward defining an empirically-based visual acuity standard for a representative task performed by aircraft maintenance inspectors. Visual acuity declines are simulated using a Gaussian blur function on airframe images. Psychophysical data were collected in non-inspectors and in highly experienced aviation maintenance inspectors. The data may be used to construct an empirically-based visual acuity standard.

INTRODUCTION

It is difficult, if not impossible, to eliminate human error in the process of inspection. Interventions must be developed to reduce these errors and make the process more error-tolerant. Since visual inspection represents a large part of aviation maintenance inspection, one mitigation strategy is to define vision standards for this vision-intensive, safety-critical occupation. A fine-tuned ability to localize, detect, discriminate, and identify job-relevant stimuli can bring cost savings and safety benefits to industry.

In 2001, an FAA Advisory Circular (AC No: 65-31) recommended examination guidelines for the vision of non-destructive inspection (NDI) personnel. It was suggested that near and far vision in at least one eye must be 20/25 and 20/50, respectively. Both near and far requirements could be with corrected or uncorrected vision. This FAA recommendation was based on acuity standards defined in other NDI/NDT occupations. Reviewing the occupational vision standards literature, Beard et al. (2002) found no studies that allow generalization of standards to aircraft maintenance inspection. It is unknown how similar tasks must be to validly borrow standards from another occupation without being subject to compromise. What is needed is a rapid, empirically-based methodology for defining occupational vision standards.

No current general standard exists in the aviation industry for the visual qualifications of aircraft maintenance inspectors. Some maintenance facilities use the visual acuity and color vision standards suggested in the FAA Advisory Circular, while other facilities have defined their own vision requirements. This illustrates the need for a uniform and universally accepted set of vision standards that would apply to all aircraft non-destructive inspection and testing (NDI/NDT) personnel.

There are several broad steps that should be taken toward setting an objective, empirically-based occupational vision requirement. The first step is a thorough vision task analysis. In the current context, the FAA commissioned CAMI to perform this analysis focusing on the role of visual processes. Next, to see if a rigorously defined standard can be borrowed from a similar occupation, a review of the literature should be undertaken. Beard et al. (2002) compiled a review of a text and WEB-based search for occupational vision requirements, knowledge gained from site visits to major aircraft maintenance facilities, relevant information from technical, mechanical, and inspection textbooks, the FAA maintenance human factors web-site¹, and the human vision literature.

If the standard cannot be legitimately borrowed from a previous standard, an objective research methodology should be followed. In their review of the vision standards literature, Beard et al. (2002) identified four occupations that had empirically derived standards. These empirical methodologies ranged from mathematically measuring the size and working distance of the critical visual details (Sheedy, 1980) to psychophysical measurements with blurring lenses placed in front of the eye on a single task (Good & Augsburger, 1987; Good et al., 1996) or multiple tasks (Padgett, 1989).

Here we present a strategy for defining a visual acuity standard that permits increased experimental control by blurring the image before presenting it to the observer, within a computer program. In this way what is done to the signal is exactly known. On the authors WEB page

¹ http://hfskyway.faa.gov
we provide the software so that this methodology may be used toward setting standards in other visually intensive occupations.

The primary objective of this research is to aid in the development of recommendations for visual acuity requirements for aviation inspection personnel. Specifically we determine that visual acuity deficits reduce critical task performance and show in graphical form the relationship between acuity decline and performance.

**METHODS**

*Choice of a critical vision task for inspectors*

The central question that must be addressed is “At what level of visual deficit would a maintenance or inspection worker become unable to safely and efficiently perform the critical visual tasks required by the job?” Aircraft inspection is a complex process, requiring many tasks, skills, and procedures. There are multiple critical vision tasks that the workers are required to perform. One purpose of inspection is to detect surface discontinuities such as cracks within the airframe and powerplant regions of the aircraft. Cracks are typically caused by two surfaces being overlaid at a boundary (Hellier, 2001). Since these cracks may be very small and of low contrast, adequate visual acuity is likely to be involved in their detection. After consultation with domain experts, crack detection was chosen as the representative task in order to ultimately set a visual acuity standard for aircraft maintenance inspection.

*Psychophysical Experiments*

**Observers**

Two female non-inspection personnel (age range from 23-30) and seven male maintenance inspectors (age range from 35-58 years) participated in the study. Maintenance inspectors were actively employed and had from 10-18 years on the job. All wore corrective lenses, though not always while inspecting. Near and far visual acuity, stereo vision, and color vision tests revealed that all had at least 20/20 acuity, good color vision, however one inspector lacked stereo vision. The inspector lacking stereo vision did not differ significantly from the other inspectors in overall detection performance (data shown below).

**Stimuli**

Airframe and powerplant crack images were obtained from various sources. Color images were converted to 8 bit black-and-white images to delete any color cues. Before the experiment, “crack removed” stimuli were generated. Using Photoshop™, the crack was deleted from the image while maintaining the integrity of the background image. The 15 images used in the current experiment are provided in Appendix A.

A “background-with-crack” image at a particular contrast level was generated by multiplying the full contrast difference image (the crack itself) by a multiplicative factor (<=1) and adding it back to the background image. The contrast in dB is 20 times the log to the base 10 of the factor. An image with the contrast of 0 dB has the original crack. An image with a crack contrast of 6 dB has the difference image reduced by a factor of 0.5. This logarithmic scale keeps the variation in the results more constant over different threshold levels.

**Crack length estimation**

To accurately determine the crack length and width, estimates of the magnification in each photo had to be determined. Each photo included a circular label or ‘sticky’ whose diameter is a known 0.75 in. To estimate the image magnification, Photoshop™ was used to identify the coordinates of six points along the perimeter of the sticky. These estimates of the perimeter were taken by eye; therefore the error in these judgments was also determined. A computer program took these data and computed a magnification value estimating the diameter of the sticky. When the sticky was on a flat surface, the image is an ellipse and the estimates were very accurate. Some of the stickies were located on an edge or curved surface. In these cases, coordinates were identified only on the flat portion of the sticky and the ellipse estimated based on this flat portion.

---

Figure 1: Crack length and width estimates. Each photo included a circular label or ‘sticky’ whose diameter is a known 0.75 in. A magnification value...
estimating the diameter of the sticky were computed from six points along the perimeter of the sticky (shown in the figure).

At first, images were adjusted so that all the stickies had a diameter of 0.75-in on the experimental display screen (the same as the sticky’s actual size), but these images were so coarse because of display resolution limitations that features of the fine cracks disappeared. Images were then adjusted to a screen sticky size of 3-in, resulting on average image width reduction from 1500 pixels to 800 pixels. Some of the images were still larger than the screen resolution of 1024 by 768 and so were cropped to 990 x 660.

**Apparatus**

Photographs of large engine airframe cracks were presented on a 1024x758-pixel display screen (SONY Trinitron). Viewing was binocular with natural pupils. From observations of aircraft inspectors performing primary inspections, Goode (personal communication) found that the majority of visual observations were done in the distance range from 34 to 40 cm. Because of screen resolution limitations, images were magnified by 4 as discussed above and so the experimental distance was comparably increased to 160 cm. From this distance each pixel subtended 0.31 arc min. The display background screen had a mean luminance of approximately 40 cd/m². Three lights illuminated a gray wall behind the monitor. Another lamp illuminated the ceiling behind the observer to achieve ambient lighting. Photometric measurements of the SONY monitor revealed that screen luminance values remained constant only after it was turned on for at least 45 minutes.

**Simulating Visual Acuity Decline**

Although the shape of the human blur function differs between individuals and changes for different optical conditions, it can be approximated by a Gaussian blur function. An observer with 20/20 visual acuity was assumed to have a Gaussian blur spread of 2 arc min (Barten, 1999; Ahumada, 1996). A person is said to have 20/40 visual acuity if they see at 20 ft what a 20/20 person sees at 40 feet. If we assume that the 20/40 person has the same contrast sensitivity as the 20/20 person, then the blur for the 20/40 person must be twice the blur of the 20/20 person. Therefore, to simulate 20/40 visual acuity the combined blur of the image and the observer should be 4 arc min. The combination rule for Gaussian blur is the Pythagorean rule, so, for example, to obtain an acuity value of 20/40, the image blur spread was set to 3.46 since the sqrt(3.46^2 + 2^2) is 4. To obtain an acuity of 20/A where A = the desired acuity level, then the blur in minutes = 2 sqrt((A/20)^2 – 1). Figure 2 presents example “crack removed” and background-with-crack images with and without blur.

![Figure 2: Examples of the crack-removed (upper left panel) and background with crack (lower left panel) images. The two right panels demonstrate these images after they have been blurred to simulate visual acuity decline.](image)

**Procedures**

**Crack Contrast Detection Thresholds**

To increase the number of images tested and the range of conditions, the two non-inspector observers collected data on a large set of crack images at a greater number of blur levels, while the NDI/NDT inspectors were run on subsets of crack images and blur levels.

Contrast detection thresholds were obtained using a two interval forced choice staircase method. The background airframe image remained on during the duration of the block of trials. On a single trial, observers were presented with the background alone in one 500 msec time interval and the background with crack in another 500 msec time interval. The interval containing the crack was randomized. The two time intervals were demarcated with a simultaneous tone. Interval one contained one tone burst, while interval two contained two tone bursts. Only one of the time intervals contained the crack stimulus. The observer’s task was to choose which interval contained the crack stimulus. The inter-stimulus interval was 500 msec. The sequence of each block of trials and the crack with background image were randomly chosen.

A different airframe image was presented in each

---

2 The spread is the distance from the center to where the blur amplitude is 1/e (0.3679) of the center amplitude.
block of trials, selected by a random permutation of all of the images and blur levels to be presented in a replication, for at least three replications. To help the observer find the crack, in initial practice trials the crack position was indicated to the observer by surrounding the crack with a rectangle. After localizing the crack, the observer could then practice the crack detection task without the surrounding rectangle before continuing on to the experiment.

On the first trial of a block of trials, the crack stimulus was presented above threshold. Estimates of these supra-threshold contrast levels were determined from model predictions (see Ahumada & Beard, 1998) and pilot data. The contrast was adjusted by a staircase procedure. On each trial, if the observer correctly responded as to which interval the crack was shown, then the response was tallied as correct. After three consecutive correct responses, the crack contrast was decreased by a specified amount (step factor). If the observer chose the interval that did not contain the crack stimulus, then a brief feedback tone would sound, the response was tallied as incorrect, and the crack contrast increased by a specified amount on the next trial. To more rapidly converge to threshold, initially the contrast step factor was 2 dB, but was reduced to 1 dB after a change in the direction of the staircase (a reversal), and then reduced to 0.5 dB after the second reversal. After eight reversals in contrast and at least 30 trials, but no more than 50 trials, the block of trials was terminated and the detection threshold calculated by a probit analysis for that crack with background image.

The two non-inspectors collected data on 10 images. The seven highly experienced aircraft maintenance inspectors collected data on either a subset of these same 10 images or on 5 different images. Observer CA collected data on images that represented six levels of blur or acuity levels: 20/20, 20/25, 20/30, 20/35, 20/40, and 20/50. Observer KJ ran on this same set of acuities plus an acuity level of 20/45. The 7 maintenance inspectors collected data on 4 acuity levels: 20/20, 20/30, 20/40, and 20/50. To evaluate the effect of viewing distance on the detection thresholds, one NASA observer was run on a subset of her conditions at a farther viewing distance of 267 cm.

**Contrast Sensitivity Functions**

To estimate the observer’s internal blur and screen resolution limitations, each observer’s contrast detection thresholds were measured for a range of stimuli. The Contrast Sensitivity Function (CSF) provides an estimate of visual acuity because an individual’s resolving power is indicated by the intersection of the curve on the abscissa of the graph. Horizontal and vertical thresholds were obtained to estimate meridional differences in the amount of blur within the experimental display.

Much like the experimental task, the observer had to decide in which of two 500 msec intervals the stimulus was presented and respond accordingly (i.e., they responded by pressing ‘1’ if they thought the stimulus was presented in interval one, and ‘2’ if they thought the stimulus was presented in interval two.) There was a 300 msec gap in between the presentation of the two stimulus images. Instead of cracks, however, the target stimuli for this experiment were a square, line and a dipole. Observers completed this experiment while sitting 273 cm from the screen.

**RESULTS**

Probit analyses were done on each block of trials to estimate the contrast threshold, the value at which the probability of correctly identifying the interval was 75%. The median of the scores replicating a particular condition was then computed.

In Figure 3, detection thresholds are presented across blur or simulated acuity levels. Each symbol represents a different airframe image. The data for each image were fit with linear functions with slopes ranging from –1.3 to -2.9 (median slope = -2.2) for Observer CA and from –2.0 to –3.3 (median slope = -2.9) for Observer KJ.
Figure 3: Contrast thresholds are presented across blur or simulated acuity levels. Each symbol represents a different airframe image. The results for observer CA and KJ are shown.

Figure 4 presents contrast thresholds for the different images as a function of blur averaged over the nine observers (inspectors and non-inspectors). Each symbol represents a different airframe image. There is a general tendency for the effect of blur to be larger as the thresholds increase. The two images with the highest thresholds could not even be run at the higher blur levels. Again, the data for each image were fit with linear functions with slopes ranging from -1.5 to -2.8 (median slope = -2.3).

Figure 5 shows the effects of blur on observers averaged over images. There is a general tendency for the effect of blur to be greatest for the observers with the lowest thresholds. The two non-inspectors (CA and KJ) showed lower detection thresholds than did the experienced aircraft inspectors. The reason for this is that observers had participated in a study of practice effects on contrast thresholds in a complex scene (Beard, et al., in preparation) and therefore are highly experienced psychophysical observers.

All data presented thus far were collected at a distance of 160 cm. Because not all inspections are done from one single distance, thresholds were measured from a second distance of 266.8 cm. Thresholds were elevated at a further distance, and show a similar increase in threshold with increases in blur.
Figure 6 shows the effect of increasing the viewing distance. Thresholds for the far distance are consistently higher than those for the nearer distance. If the detection were simply a function of target contrast energy, the threshold would be expected to increase by $20 \log_{10}(267/160) = 4.4 \text{ dB}$. Attenuation of the high spatial frequency energy should cause an additional increase in the threshold, which should be greater for the less blurred stimuli and the higher threshold stimuli.

![Detection Threshold vs. Viewing Distance](image1)

Figure 6: Viewing distance effect. Data were collected in one non-inspector observer.

To foster translation of these data into an occupational visual acuity standard, in Figure 7 we have transformed the data from Figure 5 into Probability of Detection (PoD) curves. The data were converted back to probability of Yes/No detection after being normalized by setting the unblurred probability of detection to 0.99. This calculation depends strongly on the assumed slope of the psychometric function. Here we assume the standard deviation of the cumulative Gaussian to be 4 dB, but the actual value could be anywhere from 1 dB to 6 dB.

**DISCUSSION**

Although good vision is a vital qualification for aircraft maintenance inspectors, no general standards for visual acuity currently exist for this occupation. Vision standards from other occupations cannot be “borrowed” to set a standard for maintenance inspectors because the visual demands between occupations are dissimilar and the majority of occupational vision standards are not empirically based (Beard et al., 2002).

One way to look at the effect of not having 20/20 vision is to say the an inspector with 20/40 vision sees at 20 feet what the 20/20 inspector sees at 40 feet. That is to say that the 20/40 inspector has to be twice as close as the 20/20 inspector to make the same discriminations. When the viewing distance is halved, the foveal search area is reduced by a factor of 0.25, so it would take about 4 times as much time to search the same area with the same discriminative ability.

In this project we measured detection performance on a representative task performed by aircraft maintenance inspectors as a function of image blur. These measurements allow predictions of the amount the probability of detection could change as a function of blur. As shown in Figure 7, cracks whose detection was initially at 99% could be greatly reduced by blur corresponding to only 20/30 if the inspection situation was kept constant in all other respects.

The amount of visible contrast energy in the crack correlated well with the contrast thresholds for the crack ($r = -0.89$). However, the effect of the blurring on the thresholds was much greater than the loss in visible contrast energy. For the two images with the greatest loss in visible contrast energy (4.7 dB) at the 20/40 blur level, the average threshold loss was 10 dB. Although this may be in part due to a lack of experience with these blurred images, it is also possible that the blur causes more problems with crack detection than predicted by contrast energy loss alone, such as affecting the extraction of edges. The loss in visible contrast energy can be thought of as a lower limit for the effect of blurring.

Blurring is only one possible cause of lowered acuity. Another possible cause is decreased overall contrast sensitivity. In this case, the predicted effects are
expected to follow more closely the rule that a cutting
the viewing distance in half will compensate for a 6
dB loss in sensitivity.

Methodological Limitations and Strengths
The experimental image generation procedure was
only an approximation of actual visual inspection.
Inspectors were able to use only one very relevant
strategy (contrast detection) to look for the defect
embedded within a number of realistic aircraft
locations. Although the cracks were positioned on
actual aircraft structures, inspectors could not use
many of the common strategies used in their work
environment, such as tribal knowledge (knowing
where to look), moving closer, use of shadows (i.e.,
changing the angle of light from their flashlight),
touching the crack. But there is a trade-off between
being able to use these techniques and the time it
takes to do a search. Differences between the
background conditions indicate the effect of
background variations on performance and will
reduce the importance of decision strategies on defect
detection. This methodology permits manipulation of
defect absence, length, color, and other attributes. It
is important to be capable of manipulating the
absence of a defect since uncertainty plays a large
role in maintenance inspection (i.e., there is no prior
knowledge that a defect will be present). In fact, it is
only occasionally that a defect is actually present.

Vision is a fundamental component of effective
aircraft inspection. All the same, so too are other
cognitive factors such as attention, memory, and
experience. Inspectors are knowledgeable about
individual components as well as the overall aircraft
being inspected, thus they possess the background to
properly locate, identify, and evaluate aircraft
defects. Often NTSB accident reports will point at
visual deficits as contributors to accidents because a
-crack went undetected, or a worker failed to detect
fatigue damage. However, it may not be that vision
led to these overlooks. Other cognitive factors may
have played major roles in the lack of detection: job-
related stress, worker fatigue, multi-tasking, or
memory effects of interruptions. The proposed
research isolates vision requirements on these duties.
Because the job entails much more than vision, these
results may not relate to how well the inspector will
do on the job. Therefore, although vision is a critical
component in inspection, other factors weigh in
heavily in the naturalistic task.

Other requirements should address the effects of
other cognitive contributors. These data can then be
used by the FAA to write acceptable cognitive and
perceptual standards and procedures for inspectors
including the type and frequency of vision testing
necessary to ensure the safe and effective
performance of current employees and job applicants
who will perform a particular inspection procedure.

Although psychophysical human-in-the-loop
experiments can provide accurate and objective data
toward setting a standard, it would be optimal to be
able to predict performance using a computational
model. Ahumada & Beard (in preparation) show that
a model of image discrimination does predict similar
blur effects as reported for model predictions of
simulated crack stimuli (Beard et al., 2003) but
under-predicts the blur effects seen in psychophysical
data using these actual crack stimuli.

Guidance toward the setting of a standard
These measurements do not provide a standard, but it
converts the problem to specifying a desired physical
limitation in performance. The final step in the
process of defining a visual acuity standard lies in the
hands of the FAA. Using the data in Figure 7, the
FAA must decide which stimulus characteristics and
what margin of error (e.g., 1 error in one million) will
define where to draw the line for the standard.

Recruitment, testing, selection, and training costs are
high. The rejection of qualified persons imposes an
unnecessary cost on maintenance facilities. While
the failure of proper performance on visual tasks
could be catastrophic, persons with refractive errors
such as correctable myopia who can perform the job
should be permitted to do so. Vision requirements
should be based on a demonstration that, for
example, 20/25 near or 20/50 distance visual acuity is
actually needed to perform the essential task. If the
task is not generally performed alone (i.e., there are
several people in close proximity who provide
assistance) then these tasks should not be imposed
with a vision requirement for all the individuals. In
addition, vision requirements must be based on tasks
that cannot be modified by current available
technology to assist the vision of the worker.

The governing body, here the FAA, should clearly
define the purpose of any vision test and not provide
medical examiners considerable latitude when
conducting visual acuity testing and evaluation. An
interesting case where this was not done, highlights
the importance of this recommendation. In a Safety
Advisory entitled ‘Determination of Vision
Impairment among Locomotive Engineers” (SA-98-
1) published by the Federal Railroad Administration
(FRA) and the Department of Transportation (DOT),
a lesson can be learned for the current purpose. The
FRA’s expectation was that the physicians who
would be designated as railroad medical examiners
would be trained to competently administer color
vision examinations. Thus, they did not anticipate
that it would be necessary to specify for the medical
examiners the test procedure to be employed when
testing for whether a person meets the standards
specified in this rule. That assumption has been
called into question under tragic circumstances. If
the current rule had been implemented as the FRA
expected, the rule would have adequate to prevent a
major railway accident involving the fatal collision
between two New Jersey transit commuter trains
(NTSB/RAR-97/01). The NTSB report found that
the medical history of the suspect engineer showed
that he had been administered an acceptable test
annually by the same contract physician for over 10
years. In the tenth year, the test results showed a
deterioration of the engineer’s ability to distinguish
among some colors. The engineer was then given a
Dvorine Nomenclature Test to further evaluate his
color vision. Many color weak individuals can
identify the names of colors by their brightness
instead of their hue. The examiner failed to
administer the accompanying Dvorine Second edition
color vision test, which measures color
discrimination abilities and therefore the results of
the first test suggested that the engineer did not have
a problem. It was ruled likely that the accident was
preventable if the physician had used a sound
approach to measure the person’s ability to
distinguish colors.

Self-monitoring

Aircraft maintenance inspectors as a group take great
pride in their ability to detect defects. In addition,
they care deeply about the safety implications of their
job. Many environmental and developmental
variables can affect visual sensitivity. Changes in
vision are typically slow and subtle and therefore not
easily identified by the individual. Long work shifts
or age-related accommodative changes can lead to
eye strain, headaches, excessive rubbing of the eyes,
esotropia or exotropia, and reduced efficiency on the
job. Without an objective measuring tool, workers
will not detect gradual changes in their vision. If you
don’t see something, you don’t know that you can’t
see it (self-awareness). Providing the workers with a
method to self-monitor their visual acuity would
enhance occupational safety and safety in the NAS.

ACKNOWLEDGMENTS

Supported by FAA Agreement DTFA-2045 to B. L.
Beard.

We thank the inspectors who participated in the
study. They were wonderful psychophysical
observers who always showed great professionalism
and devotion toward this effort.

REFERENCES

Ahumada, Jr., A.J. (1996) Simplified vision models
for image quality assessment, in SID International
Symposium Digest of Technical Papers, ed J.
Morreale, Vol. 27, pp. 397-400, SID, Santa Ana, CA.

Ahumada, A.J., Beard, B.L. (1997). Image
discrimination models predict detection in fixed but
not random noise. Journal of the Optical Society of
America A 14, 2471-2476.

Barten P.G.J. (1999) Contrast sensitivity of the
human eye and its effects on image quality. Kneisel:
HV Press.

model for inhomogeneous image quality assessment,
in SID Digest of Technical Papers, ed. J. Morreale,
vol. 29, Paper 40.1, Santa Ana, CA.

Using vision modeling to define occupational vision
standards, Human Factors and Ergonomics Society
Annual Meeting (October 15, Denver, CO).

Occupational vision standards: A review. Report for
FAA AAR-100.

Uncorrected visual acuity standards for police
applicants. Journal of Police Science and
Administration, 15(1), 18-23.

Good, G. W., Weaver, J. L., & Augsburger, A. R.
(1996). Determination and application of vision
standards in industry. American Journal of Industrial
Medicine, 30(5), 633-630.


validation of firefighter vision requirements: four
field experiments. Paper presented at the
International Personnel Management Association
Assessment Council 13th Annual Conference held in
Orlando, FL. Retrieved July 11, 2002, from
http://home.earthlink.net/~medtox/firevis.html

of Police Science and Administration, 8(3), 275-285.