Clin. Vision Sci. Vol. 2, No. 3, pp. 187–199, 1988
Printed in Great Britain. All rights reserved

0887-6169/88 $3.00 + 0.00
Copyright © 1988 Pergamon Press plc

THE DESIGN OF A NEW LETTER CHART FOR MEASURING CONTRAST SENSITIVITY

D. G. PELLI,[1] J. G. ROBSON[2] and A. J. WILKINS[3]

[1]Institute for Sensory Research, Syracuse University, Syracuse, NY 13244-5290, U.S.A.,
[2]Physiological Laboratory, Cambridge University, Cambridge CB2 3EG and
[3]Medical Research Council, Applied Psychology Unit, Cambridge, England

(*Received* 31 *August* 1987; *in revised form* 9 *October* 1987)

Summary—1. A consideration of methods for assessing contrast sensitivity leads to the conclusion that, for a clinical test, letters are more suitable than gratings.

2. A letter chart is described in which letters decrease in contrast but not in size. The letters are arranged in groups of three; successive groups decrease in contrast by a factor of $1/\sqrt{2}$ from a very high contrast down to a contrast below the threshold of normal observers. A subject's threshold is taken to be the lowest contrast for which at least two letters in a group are correctly reported.

3. A mathematical model of the observer and the chart-testing procedure has been used to predict how the accuracy and repeatability of the test score depend on the parameters of the chart and observer. This reveals that even a low probability of misreporting suprathreshold letters will seriously bias the test score if the passing criterion is strict, requiring correct report of *all* letters in each group, but will have little effect if the passing criterion is less strict. This effect of the passing criterion may explain Rubin's [*Clin. Vision Sci.* **2**, No. 1 (1987)] finding that the new test, which uses a lenient criterion, has excellent test-retest reliability, much higher than the Ginsburg [*Am. J. Optom. Physiol. Opt.* **61**, 403–407 (1984)] chart with its strict criterion.

Key words—Contrast sensitivity; letter sensitivity; contrast sensitivity test; clinical testing; charts.

INTRODUCTION

Characterisation of visual function in clinical practice is still based primarily upon measurement of letter acuity, supplemented occasionally by other measures such as perimetry. While these measures are certainly very useful, there is still a great need for earlier detection of various diseases which impair visual function, as well as for a quick and simple method for characterising the level of visual function resulting from stable visual impairment. We believe that this need can best be met by supplementing the usual acuity test by a measurement of contrast sensitivity for a target of intermediate size. We shall describe the design considerations that have led us to develop a variable-contrast letter chart for the clinical determination of contrast-sensitivity*.

*The Pelli–Robson Letter Sensitivity Chart will be available from Clement Clarke International Ltd, 15 Wigmore Street, London W1H 9LA, England, and Clement Clarke Inc, 3128-D East 17th Ave., Columbus, Ohio 43219, U.S.A.

BACKGROUND

Although it has been realised for a long time that the measurement of increment thresholds for such test targets as discs on a uniform background can provide useful information about the visual system that acuity measurements fail to provide, past attempts to introduce such contrast-sensitivity tests into clinical practice (e.g. the George Young Threshold Test, Young, 1918) have had little lasting success. More recently, following the introduction of contrast-sensitivity measurements for sinusoidal gratings of different spatial frequencies as a general research method for the characterisation of human spatial vision (Schade, 1956; Campbell and Robson, 1968), there has been substantial interest in using such measurements as a basis for clinical visual assessment (e.g. see Wolkstein *et al.*, 1980). While sufficient work has now been done to suggest that contrast-sensitivity measurements made with gratings would be useful in the clinic, it seems to us unlikely that the full potential of contrast-sensitivity testing will be realised unless the

measurement can be made as simple and reliable as the letter-acuity test itself.

In research laboratories, contrast-sensitivity measurements for sinusoidal gratings have usually been made using patterns displayed on a cathode-ray tube, the electrical signals required often being computer generated and the measurement procedure computer controlled. With such equipment, the threshold contrast can be measured at as many closely-spaced spatial frequencies as desired over the visible range (typically 0.5–40 c/deg), and the result plotted as a contrast-sensitivity function. It is also possible, in a research setting, to use good psychophysical procedures (e.g. a two-alternative forced-choice staircase method with randomly interleaved stimuli of different types) despite the added complexity and extra time required. Measurements can certainly be made in this way in a clinical setting (e.g. see Wolkstein *et al.*, 1980) and apparatus for performing such measurements is commercially available (e.g. the Nicolet Optronix Vision Tester). However it is not clear that for routine clinical use, particularly for screening purposes, it is appropriate to use the same measures of contrast sensitivity or the same methods of measurement as have been adopted in the context of basic vision research.

There are three aspects of the standard research procedure whose appropriateness in a clinical context may be questioned. These are (1) the use of cathode-ray-tube displays to get test patterns of different contrasts, (2) the use of gratings as test patterns and (3) the measurement of contrast sensitivity for test patterns of several different spatial frequencies.

Methods for producing the test patterns of different contrasts

It is certainly true that cathode-ray-tube raster-display devices used in conjunction with analogue or digital electronic waveform generators provide the most satisfactory and convenient way of producing stimuli of different low contrasts. They are particularly satisfactory (in some cases essential) if continuously variable grey levels or continuous control of contrast is required, if very low contrast patterns are needed (they can provide very uniform backgrounds) or if the contrast must be temporally modulated or the pattern moving. However, they suffer from being relatively bulky and expensive, as well as requiring maintenance and routine calibration. These are not important considerations in a research context but in considering how to provide a routine clinical test which might be carried out in every optometrist's or physician's office, it is clearly desirable to find some simpler alternative. There can be no doubt that in this context what is needed is an easily used printed test chart rather than a complex apparatus requiring elaborate or time-consuming procedures. The chart should be durable and inexpensive to produce. Such a chart would display a small number of test patterns of different contrasts and the subjects would be required to identify each test pattern, thereby providing objective evidence that they could see it.

One obvious method of producing a test chart with patterns of different contrasts is to photograph a cathode-ray-tube display. This is a very convenient way of obtaining static demonstration patterns and has been frequently used in recent years to provide illustrations for research publications (e.g. Robson's variable contrast and variable frequency grating reproduced as Fig. 4.12 in Ratliff, 1965) where exact rendering of the original is not really necessary. However, the problems encountered in obtaining exact masters from the original display and in printing multiple subsequent copies are formidable, whether the printing is by normal half-tone methods or is photographic. Despite this, two sinusoidal grating test charts for clinical use have been made photographically (Arden, 1978; Vistech Vision Contrast Test System, Ginsburg, 1984) and are commercially available. Photographic methods have also been previously used to produce other low-contrast test charts in limited numbers where the requirement for individual calibration was acceptable (e.g. the Landolt-C chart of Hecht *et al.*, 1949, used for demonstrating the effects of hypoxia, or the edge charts of Bailey, 1982, or Verbaken and Johnston, 1986).

Another alternative for producing test patterns is to use computer graphics methods to generate half-tones directly. Della Sala *et al.* (1985) have described the use of square-wave gratings made up of many closely-spaced parallel thin black lines which were varied in their separation (and thus their density) to vary the mean reflectance of the areas comprising the light and dark bars of the grating. The authors discuss the simplicity and robustness of this method of producing gratings of controlled contrast using a computer-driven XY plotter to produce the master patterns.

A variant of this method has now been de-

scribed by Wilkins *et al.* (1987) whose square-wave gratings* were printed using black dots produced by an impact matrix printer (rather than pen-plotted lines) to reduce the reflectance of the dark areas of the test chart. So long as the dots are all equal in size and blackness and do not overlap, the reduction in reflectance (and hence luminance) of a printed area is exactly proportional to the density (number per unit area) of the dots. This makes it easy to generate grating patterns with known contrast ratios and provides masters which can easily and satisfactorily be reproduced by standard printing methods. The ability to reproduce a test chart using standard printing methods is important if production costs are to be minimised.

There is a new kind of printing device, much better than impact printers or XY plotters. Pelli (1987) has discussed the use of standard (300 dots/inch) xerographic laser printers, as well as photographic laser typesetters with substantially greater resolutions, for the production of various vision test charts. The use of such devices is greatly simplified by the availablility of programming languages such as PostScript, specially designed for controlling them. There seems little doubt that programming a laser typesetter to produce variable-dot-density halftones is the best method currently available for originating test charts which are to be reproduced by standard printing methods.

Although it is not even potentially suitable for rendering test patterns involving continuously varying grey levels, there is one more method of printing low-contrast patterns which has been employed in producing clinical test charts. This relies on using pale inks to print areas of constant reflectance. The George Young Test† used water-based inks diluted by successively greater amounts to print progressively paler circular patches randomly placed on succeeding pages of a little book (Young, 1918). Diluting the ink by a further factor of two each time produced a test patch whose contrast was one half that of the previously printed patch.

Much more recently Bailey and Regan have used grey inks of different reflectances to print (silk-screen) a set of Snellen charts, each with letters of a different low contrast (Bailey, 1982, 1987; Regan and Neima, 1983, 1984). While this is evidently a satisfactory method for the production of charts with letters having a contrast of 0.1 (10%) or greater (as in Regan's charts), it would be more difficult to use this method to produce a chart with a much lower contrast or with many contrasts.

Sinusoidal and other gratings as test patterns

The use of sinusoidal gratings as optical test patterns was introduced by Selwyn (1948) primarily for studying camera optics and photographic films. However, since the overall resolution of a photographic system could only be understood by taking into account the characteristics of the human observer as well as those of the lens and film, Selwyn also made the first direct measurements of the visibility of sinusoidal gratings. Selwyn's main justification for using a sinusoidal grating as a test object was based upon the realisation that the light distribution in the image of such a target must necessarily also be sinusoidal. Selwyn was also aware that photographic emulsions as well as lens performance could usefully be examined using sinusoidal grating targets and was impressed that "the adoption of such . . . test objects results in considerable simplification of theoretical arguments".

A rather similar rationale lay behind Schade's (1956) adoption of sinusoidal gratings for characterising human vision in the context of providing a comprehensive description of the whole television chain; each component in the chain—camera lens, television pick-up tube, signal transmission system, cathode-ray-tube display and human observer—could be characterised in a consistent way in terms of its spatial-frequency response function. In many ways this unified approach to an electro-optical system was a spatial counterpart to the characterisation of electronic sound-recording and reproducing systems using pure-tone sounds (sinusoidal pressure fluctuations) and sinusoidally-varying voltages.

In both Selwyn's and Schade's work the use of sinusoidal gratings was a practical and theoretical convenience; it did not depend upon any deep knowledge of the visual mechanism or upon the visual system functioning in any particular way; essentially it represented these authors' response to a practical need to proceed with applied work involving vision, in the absence of detailed understanding of the visual system. Following Selwyn and Schade there have been a very large number of subsequent

*Cambridge Low Contrast Gratings, available from Clement Clarke International Ltd, 15 Wigmore Street, London W1H 9LA, England.

†Originally published by Raphael's Ltd, London.

studies of the visual system, both psychophysical and neurophysiological, which have also been based upon characterising the visual process by an examination of visual performance using sinusoidal gratings. Many of the psychophysical studies, starting with that of Campbell and Robson (1968), have concluded that the visual system operates as a system of parallel channels or mechanisms tuned to different bands of spatial frequencies, while neurophysiological studies (see Robson, 1983) have shown that neurones at various levels in the visual system are also more-or-less selectively responsive to restricted bands of spatial frequencies (as well as to restricted ranges of orientation). These findings can be taken to imply that sinusoidal gratings play a special role in the study of visual function since they are the elementary signals appropriate to the spatial-frequency domain.

On the other hand, it must be remembered that the visual nervous system is comprised of neurones which, although they may be orientation and spatial-frequency selective, have receptive fields of limited extent. For such neurones the most appropriate stimulus will be one which is restricted to the area of the neurone's receptive field. This suggests that the appropriate test stimulus would be a grating patch of rather small extent, a stimulus which is believed to be visually optimal (Watson *et al.*, 1983). The desirability of using stimulus patterns of limited extent for clinical vision testing may also be argued on the basis of the well-known non-uniformity of acuity and sensitivity over the visual field which are confounding factors when large area gratings are used. However, while a case could certainly be made for using small grating patches as the basis of a standard test of contrast sensitivity, there are various reasons why, in clinical practice, letters are more desirable than grating patches. The principal advantage of letters is that they lend themselves to the adoption of particularly good psychophysical procedures (to be discussed below). But as well as this, letters are familiar and important in everyday life. Letters have also been perceived to have the advantage of including "vertical, horizontal, oblique and curved contours" (Sloan, 1959).

Letter charts

The high-contrast letter-based acuity chart was first described in very nearly its modern form by Snellen in 1862. It is a printed chart which can be inexpensively reproduced and whose relevant characteristics (contrast and size) are very stable. Because letters have distinct forms with which most subjects will be very familiar, it is easy to use correct identification as a criterion for determining whether or not the letters have been seen; this makes the test objective and more reliable than it would be if a subjective detection criterion were used (Vaegan and Halliday, 1982; Higgins *et al.*, 1984). While subjects can be asked to identify the orientation of gratings (as for Ginsburg's chart, 1984), for most subjects this is not as straightforward, or as quick, as naming letters; nor is it possible to provide as many different, but readily-named, orientations as letters. With a letter chart, the testing procedure—identifying letters in a group—is easily explained, easy to score, and very quick.

A subtle virtue of traditional acuity charts is that the test is relatively error-tolerant since the subject is typically required to read only 4 out of 5 letters in each line to pass that line, so that up to one mistake per line has no effect on the test result.

Given that the standard letter-based acuity chart is, by its very nature, a most satisfactory way of determining a subject's high-contrast acuity, and one with which both patients and examiners are already very familiar, it seems particularly desirable to adopt as similar a chart as possible for measuring contrast sensitivity. This argument has been put forward by Regan and Neima (1983, 1984) to support the use of Snellen charts printed with low-contrast leters to determine low-contrast acuities. As noted earlier, Regan and Neima's charts have, like the regular Snellen chart, letters of various sizes but all of the same contrast. Using a single chart of this kind, it is therefore possible to determine for what size of letter a subject's contrast sensitivity has some particular value; by using a number of charts with different contrasts it is possible to map out a contrast-sensitivity function for letters.

The main reason for clinical contrast sensitivity measurements is to determine if subjects with normal acuity have abnormally low contrast sensitivity at lower spatial frequencies. Because contrast sensitivity falls steeply both at high spatial frequencies and at small letter size (Legge *et al.*, 1987), the Regan–Neima low-contrast acuity charts are less sensitive to changes in contrst sensitivity than is a direct

measure of contrast sensitivity by a chart in which fixed-size letters vary in contrast.

How many sizes of letters do we need?

As explained above, the recent interest in visual contrast-sensitivity measurements as a clinical tool has followed from the introduction of the concept of the contrast-sensitivity function (contrast sensitivity for sinusoidal gratings as a function of spatial frequency) as a fundamental description of normal visual performance. Initially it was unclear how this function might be affected by clinically significant disorders, and at that stage it was obviously necessary to examine the whole contrast-sensitivity function. Now, however, it is fairly clear that one of four kinds of distortion of the contrast-sensitivity function can occur (Regan and Neima, 1983). There may be a loss of sensitivity at high spatial frequencies only (resulting in a reduction in high-contrast acuity), a general loss at all frequencies (resulting in a reduction in high-contrast acuity and intermediate-frequency contrast sensitivity), a loss mainly at low and intermediate spatial frequencies or a loss restricted to intermediate frequencies. While detailed assessment of the visual defect in any particular subject will require the subject's whole contrast-sensitivity function to be measured, it should be possible to determine if a subject's vision is abnormal by making only two measurements. One measurement is required to reveal the existence of reduced acuity at high contrast and a second to reveal a contrast-sensitivity loss at intermediate spatial frequencies. It should be possible to provide the first of these measurements using a standard high-contrast acuity chart and the second using a variable-contrast letter chart with letters of appropriate size.

Ginsburg (1978) has shown that the recognition of letters is possible using just the spatial frequencies of between 1.5 and 2.5 cycles per letter width. Similarly, Legge *et al.* (1985) have shown that spatial frequencies higher than two cycles per letter width are unnecessary for achieving optimum reading rates. Since the contrast energy of letters falls off rapidly with spatial frequency, it is likely that at threshold it is frequency components below about 2.5 cycles per letter width which provide the visual signals used by a subject for identifying letters. Thus a variable-contrast letter chart with letters that subtend an angle of one half degree at the eye can be presumed to measure a subject's contrast-sensitivity at spatial frequencies between 3 and 5 c/deg. This frequency range corresponds reasonably well with the optimum frequency for normal subjects and is arguably the best range for determining whether or not a subject has an intermediate spatial-frequency sensitivity loss. The choice of spatial frequency at which it is best to estimate contrast-sensitivity loss is in many cases probably not very critical as Pelli *et al.* (1986, 1988) have shown that the upper part of the contrast-sensitivity function of normal and low-vision observers has a fairly stereotyped form whose parameters can be quite well estimated from measurements of high-contrast acuity and contrast sensitivity for 3 deg letters. On the basis of the above discussion we believe that a single chart with letters all of one size could be used to obtain as much information on a subject's contrast sensitivity as would be clinically useful.

A NEW LETTER CHART

We have now produced such a chart with letters of various contrasts (Fig. 1) as a computer-generated variable-dot-density halftone originated on a phototypsetter (Linotronic 300) having a resolution of 2540 dots/inch (100 dots/mm). Full-size copies of the original chart have been printed using offset lithography.

The size of the letters on this new chart is such that they subtend 0.5 deg at 3 m. While it is supposed that the chart will normally be used at a distance of 3 m, it can be used at much nearer distances for assessment of low vision.

For the new chart we have chosen to use the Sloan letter set, mainly because we suppose that this chart will be used in conjunction with a standard high-contrast acuity chart using the same letter set (National Academy of Sciences–National Research Council, 1980). Although the Sloan letters are said to be "about as nearly equal in legibility as can be obtained with simple capital letters" (Sloan, 1959), we do not know whether this holds for large, low-contrast letters. It is desirable to use letters which are equally well identified and it is possible that an alternative letter set or letter forms would be better in this respect. We intend to study this.

The chart is read in the conventional way from left to right and from top to bottom. On each line of the chart there are two groups, each of 3 letters. The letters in each group have the same contrast and the contrast of the letters in each successive group is less than that of the

0.05	O S N Z C N	0.20
0.35	S H O C H V	0.50
0.65	K D R Z K D	0.80
0.95	H C D S N O	1.10
1.25	O V S D R H	1.40
1.55	D S N H R K	1.70
1.85	D N Z N V H	2.00
2.15	R D H H K Z	2.30

Fig. 2. Key to the Pelli–Robson chart as used for scoring a subject's performance. The marginal numbers give the log contrast sensitivity corresponding to the neighbouring group of three letters. For example, the number 2.00 appears next to a group of letters with a contrast of 1/100 (i.e. 1%) indicating a log contrast sensitivity of 2.00.

letters in the preceding group by a factor of $1/\sqrt{2}$. The correct relationship between the contrasts of the different groups of letters is assured by the use of the variable-dot-density half-tone technique, while the correct absolute contrast level is adjusted during printing and checked afterwards photometrically.

The size of the letters on this new chart is such that they subtend 0.5 deg at 3 m, the distance at which we suppose the chart will normally be used. However, in order to detect a loss of contrast sensitivity due to wide-angle light-scatter (e.g. from cataract) it is necessary for a large part of the subject's visual field to be approximately as bright as the chart itself, so the chart should be mounted on a white wall. If this cannot be provided, then it would be preferable to use a chart with smaller letters and large white margins at a near-vision distance. For assessment of low vision the standard 3 m chart should be used at a shorter viewing distance, so that the letters will be well within the acuity limit.

The subject reads the letters on the chart, starting with those of highest contrast, and continues until two or three of the letters in one group are incorrectly named. The examiner scores the subject's performance on a key like that in Fig. 2, which shows all the letters at full contrast, and gives the log contrast sensitivity corresponding to each group. The subject's score is then determined by the previous group (the last group in which two or three letters were correctly named).

The order in which the letters appear on the chart is randomised but, in order that subjects should be aware of the character set being used in the lower-contrast area of the chart, all ten of the Sloan letters appear in the top three lines of the chart. For most subjects these first high-contrast characters will be readily visible and will serve not only to familiarise them with the characters but also with the nature of the chart and the test procedure.

PARAMETERS OF THE NEW CHART

In deciding on the parameters of the new chart, i.e. the number of letters per line, the decision rule for finding the subject's threshold contrast and the magnitude of the contrast change from line to line, we developed a statistical model for the testing procedure. This model was designed to enable us to choose the optimum parameters that would maximise the accuracy of the measurements provided by the chart. The model has two parts, the observer and the chart (with test procedure).

The chart

Each letter on the chart is assumed to be a random sample from an alphabet of N letters, usually 10. The letters are arranged in groups of m letters, all of equal contrast. The first group is assumed to have unit contrast (zero log contrast; somewhat greater than the maximum contrast on a printed chart) and each subsequent group to be at a lower contrast, the log contrast being reduced by a constant decrement, δ. The subject starts reading at the first group and continues reading subsequent groups until failing a group. The criterion for passing a group (i.e. not failing) is to name correctly at least k of the m letters. The raw score assigned to the subject is the log contrast of the last group passed (i.e. for which the subject identified at least k of m). If the observer fails the first group then the raw score is the log contrast of the first group plus δ.

The observer

The goal of the test is to measure the observer's threshold log contrast, t. We assume that the observer's probability of correctly naming a letter at log contrast x is given by a psychometric function, $P(x-t)$

$$P(x-t) = (1-\epsilon)W(x-t) + \epsilon g \qquad (1)$$

where $W(x-t)$ is the probability that the subject will correctly see the letter, ϵ is the probability of the subject making a "misreporting" error, i.e. that the subject will report a random letter instead of the one actually "seen", and the

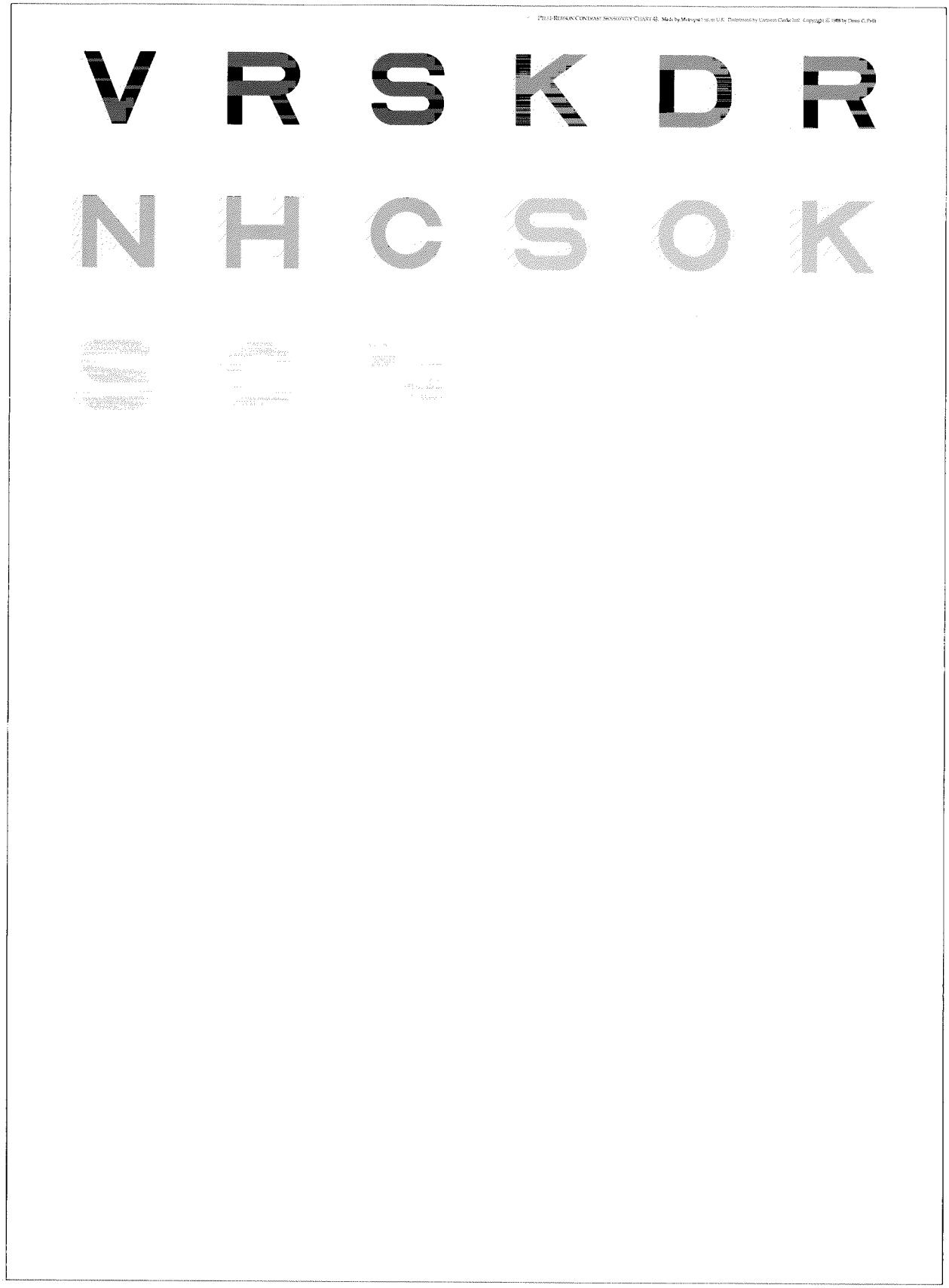

Fig. 1. A miniature Pelli–Robson Letter-Sensitivity Chart. It should be noted that the contrast levels of the letters in this miniature reproduction of the chart are not exactly those on the full-size chart as listed in Fig. 2.

guessing rate g is the probability of correct response at zero contrast. We set $g = 1/N$, where N is the number of letters in the alphabet. $W(x-t)$ is assumed to be a Weibull (1951) function

$$W(x-t) = 1 - (1-g)\exp[-10^{\beta(x-t)}], \quad (2)$$

where β is a parameter controlling the steepness of the Weibull function. The psychometric function is sigmoidal with a lower asymptote g, a transition section with steepness controlled by β, and an upper asymptote $1-\epsilon$. The lower asymptote is the guessing rate g which we set to $1/N$.

The calculation

Let the log contrasts of the groups on the chart be $x_0, x_1, x_2, \ldots$, where $x_i = x_0 + i\delta$ is the log contrast of the ith group, δ is the step size, and x_0 is the log contrast of a nonexistent zero-th group, i.e. the score assigned when the subject fails the first group. The probability of passing the ith group is given by the probability of correctly identifying at least k of m letters.

$$P_{\text{pass}}(i) = \sum_{j=k}^{m} \binom{j}{m} P(x_i - t)^j [1 - P(x_i - t)]^{m-j} \quad \text{for } i > 0 \quad (3)$$

where $\binom{j}{m}$ is the binomial coefficient

$$\binom{j}{m} = \frac{m!}{j!(m-j)!}. \quad (4)$$

The probability of passing the nonexistent zero-th group is 1

$$P_{\text{pass}}(0) = 1. \quad (5)$$

The probability P_i of obtaining score x_i is given by the product of the probability of passing all groups up to and including group i and the probability of failing group $i+1$

$$P_i = [1 - P_{\text{pass}}(i+1)] \prod_{j=0 \text{ to } i} P_{\text{pass}}(j). \quad (6)$$

The mean score will be

$$\mu_t = \sum_{i=0 \text{ to } \infty} x_i P_i. \quad (7)$$

The variance of the score, will be given by

$$\sigma_t^2 = \sum_{i=0 \text{ to } \infty} (x_i - \mu)^2 P_i. \quad (8)$$

Obviously the mean score, μ_t, will depend upon the threshold, t. It might seem that the ideal chart would yield $\mu_t = t$, but that is not strictly necessary. For a fixed set of parameters one can always print a scoring sheet that assigns modified scores x_i' so that the mean modified score equals threshold. However, in practice the examiner cannot correct for parametric variation from subject to subject and this is an unnecessary refinement.

Equations (7) and (8) converge quickly, so that the mean score and variance are easily determined by computer calculations. Initial calculations showed that the variance σ_t^2 and mean error $\mu_t - t$ are remarkably insensitive to t. However, we did consider more than one value of threshold, in case it matters whether threshold contrast is equal to the contrast of a group or to an intermediate level between two groups. Therefore we assumed threshold was equally likely to take one of two values, a log contrast of $t_1 = -2$ (i.e. 1% contrast), or a log contrast of $t_2 = -2 + \delta/2$, i.e. half a step higher. The overall mean error, e, and variance, σ^2, are computed from the separate mean errors and variances

$$e_1 = \mu_{t_1} - t_1$$
$$e_2 = \mu_{t_2} - t_2$$
$$e = (e_1 + e_2)/2$$
$$\sigma^2 = (\sigma_{t_1}^2 + \sigma_{t_2}^2)/2 + (e_1 - e_2)^2/4.$$

The parameter β was fixed at 3.5. Increasing or reducing β is equivalent to making a reciprocal change in the step size δ (and the relative threshold $t - x_0$, but $x_0 = 0$ and the results depend only weakly on t).

RESULTS

Figure 3 shows the overall standard deviation σ of the test score as a function of the step size δ for various passing criteria k and numbers m of letters per group with the initial assumption that there is no misreporting error, i.e. $\epsilon = 0$.

As one would expect, reducing the step size reduces the standard deviation. However, note that as the step size is made smaller and smaller the benefit of further reduction wanes. For most of the conditions, when the step size is larger than 0.1 (i.e. a tenth of a log unit) the standard deviation is smaller than the size of a single step. When the step size is smaller than 0.1 the standard deviation is mostly larger than a step. Note that the various passing criteria (various k of m) affect standard deviation in the direction one would expect, with more letters per group

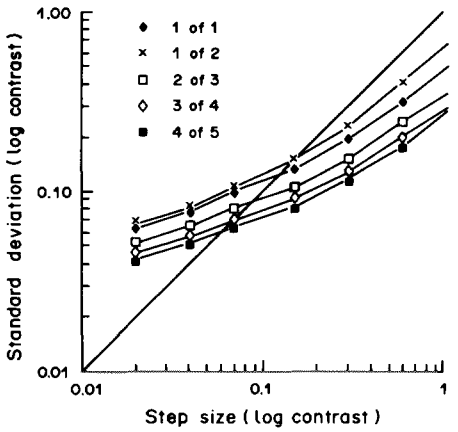

Fig. 3. Standard deviation of estimated log contrast sensitivity vs step size, for various passing criteria, assuming no misreports, $\epsilon = 0$.

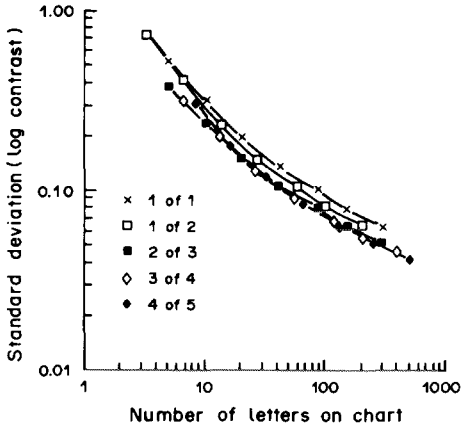

Fig. 4. Standard deviation of estimated log contrast sensitivity vs number of letters on the chart, for various passing criteria, assuming no misreports, $\epsilon = 0$.

tending to reduce the standard deviation. However, the effect is modest, and two of the conditions cross, so that for the smallest step size the 1 of 2 rule yields the largest standard deviation.

In a contrast-sensitivity chart all the letters are the same size, so a chart of fixed area will accommodate a fixed number of letters. Doubling the number of letters per group will require halving the number of groups on the chart. The proper comparison then, among rules with different numbers of letters per group, is to examine the standard deviation as a function of the total number of letters on the chart. Figure 4 replots the data of Fig. 3 in this way—standard deviation vs number of letters in the chart. The number of letters per chart is calculated as $2m/\delta$.

Surprisingly the lower three conditions are nearly superimposed, showing that, by this measure, it makes no difference whether we have steps of $2.0/9 = 0.22$ among 9 groups of 5 letters (and require 4 of 5 to pass) or steps of $2.0/15 = 0.13$ among 15 groups of 3 letters (and require 2 of 3 to pass). The upper two conditions, 1 of 1 and 1 of 2, yield higher standard deviations over the entire domain (i.e. charts of less than 10 to more than 100 letters). Thus we should use at least 3 letters per group to obtain the highest efficiency, i.e. the lowest standard deviation for any given number of letters on the chart.

At this point we must consider the appropriateness of our model of human performance. It is important that the test be robust, so that the mean and standard deviation of the score be at most slightly affected by variations in the parameters of the observer, or behavior different from that of our model observer. There are an unlimited number of ways in which real observers can deviate from the simple model assumed here, but we believe that as a practical matter the performance of experienced observers is reasonably well modeled by the psychometric function in equation (1), and that the main difference between experienced and inexperienced observers is that the inexperienced observers have a higher (and more variable) probability ϵ of misreporting their observations.

Figure 5 shows the standard deviation of the test score as a function of the probability ϵ of misreporting. A logarithmic step size of 0.15

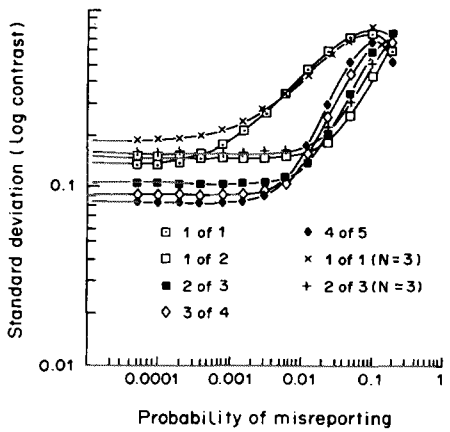

Fig. 5. Standard deviation of estimated log contrast sensitivity vs probability of misreporting, for various group sizes and passing criteria. The alphabet size N is always 10, except as indicated. The Ginsburg chart has a 1 of 1 passing criterion and an alphabet size of $N = 3$.

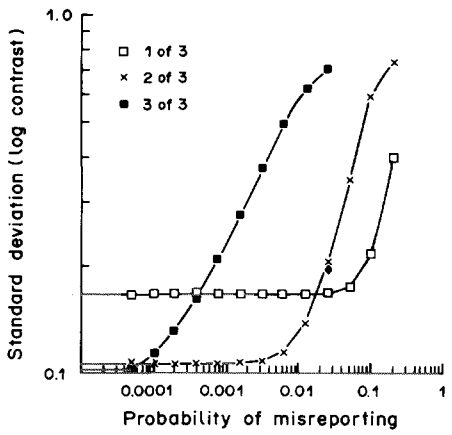

Fig. 6. Standard deviation of estimated log contrast sensitivity vs probability of misreporting, for groups of three letters and various passing criteria.

(i.e. a factor of $\sqrt{2}$ in contrast) is assumed; both our new chart and the Ginsburg chart use this step size. An alphabet size, N, of ten letters is assumed for our new chart, and three for the Ginsburg chart (gratings of three different orientations). The main feature of this graph, which shows the effect of adopting various decision rules, is that for each rule the standard deviation of the test score is essentially unaffected by infrequent misreporting errors, but beyond some critical probability of misreporting the standard deviation rises disastrously. Of the curves relating to a ten-alternative chart, the least satisfactory is clearly that for the 1 of 1 rule for which the standard deviation is more than doubled (to 0.4 log units) if the probability of misreporting is only as high as 1%. A similar unsatisfactory behaviour is seen with the three-alternative Ginsburg chart which also uses a 1 of 1 rule. The Ginsburg chart differs from our chart in two ways, a smaller alphabet of 3 instead of 10, and a 1 of 1 rule instead of a 2 of 3 rule. Figure 5 includes graphs for the 1 of 1 and 2 of 3 rules with the smaller alphabet, $N = 3$, as well as the usual $N = 10$. Comparing these graphs shows that reducing the alphabet only increases the standard deviation somewhat, and does not affect the degree of tolerance to misreporting. Thus the use of only three test patterns in the Ginsburg chart is only a minor drawback, increasing the standard deviation slightly, but the use of a strict 1 or 1 criterion is a major flaw, resulting in high susceptibility to misreporting errors. This difference in susceptibility to misreporting errors may explain why Rubin (1988) found poor test–retest reliability using the Ginsburg chart (52% intraclass correlation for normal subjects and 60% for patients) and high test-retest reliability using the Pelli–Robson chart (98% and 86%).

We suspect that even experienced observers have a probability of misreporting of around 1% and that inexperienced observers may have a higher probability, perhaps occasionally as high as 5%. Note that the left-to-right ordering of the curves is in order of leniency; the more forgiving the rule, the more misreporting can be tolerated without loss of accuracy. It is clear that the 2 of 3 rule compares favorably with the other rules, tolerating more misreporting than all but the 1 of 2 rule, but having nearly half the standard deviation of that rule when misreporting is rare.

Figure 6 compares the 2 of 3 rule with other 3-letter rules: 1 of 3 and 3 of 3. As in Fig. 5, the more lenient rules provide more tolerance for misreporting. However, the 2 of 3 rule is a good compromise, having the lowest standard deviation for misreporting rates in the range 0.007–2%.

Figure 7 shows that the misreporting rate also affects the mean score. This is because misreporting may cause a subject to fail one of the many groups before reaching threshold. The plotted bias is the difference between the mean score at a particular misreporting rate and the mean score when the misreporting rate is zero. The graph is qualitatively similar to Fig. 6, but the critical misreporting rates are higher, so that this bias is not likely to be important in practice, as the standard deviation with high rates of misreporting would already be prohibitively large. However, it is worth pointing out that

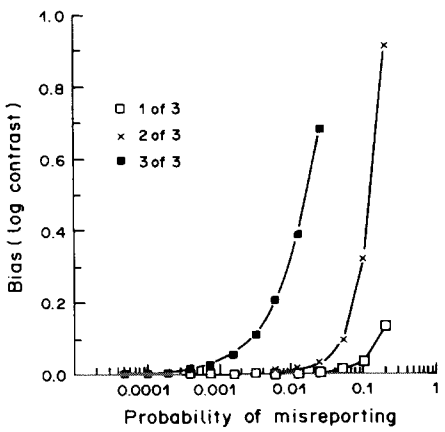

Fig. 7. Bias in the estimated log contrast sensitivity vs probability of misreporting, for various passing criteria.

the graphs of standard deviation eventually fall (when the bias rises) at high misreporting rates (Fig. 5). This is because the misreporting rates eventually become so high that the observer regularly fails the first group. Thus empirical evidence of a consistently high test score does not necessarily imply accurate measurement of a high threshold, as it could also be the result of a high misreporting rate.

Conclusions from modelling

Statistical modelling of the testing process indicates that for minimum standard deviation of the final threshold estimate there should be at least 3 letters per group. To minimise the effect of misreporting it is essential that the rule be lenient. A 2 of 3 rule is essentially immune to misreporting at rates up to about 1%. If misreporting rates are higher than 2% then an even more lenient rule should be used. A 1 of 3 rule is immune up to about 10%, but the cost is that at low misreporting rates its standard deviation is nearly twice that of the 2 of 3 rule. These conclusions have been incorporated in the design of the new chart described earlier.

The poor test–retest reliability of the Ginsburg chart (Rubin, 1988) may be due to its 1 of 1 criterion, partly because having only one letter per group is inefficient (see Fig. 4), but mostly because a strict criterion makes the test extremely vulnerable to misreporting errors. Figure 5 indicates that if the observer misreports only 1% of the trials the standard deviation will be increased from 0.2 to 0.5 log units.

GENERAL DISCUSSION

Clinical contrast-sensitivity measurements promise to detect and assess visual diseases which do not affect visual acuity. However, none of the previously available methods seem to be sufficiently quick and reliable for it to be expected that they will be widely used. It is certainly possible to make measurements, using currently available methods, for assessing the progress and disabling effect of known disease, though we believe that use of our new chart could make this easier. One potential benefit of contrast-sensitivity measurement is the early detection of such disorders as diabetic retinopathy and glaucoma. As Hyvärinen *et al.* (1983) have written of glaucoma, "The large interindividual and possible interocular variations in contrast sensitivity make it difficult to assess the effect of glaucoma until we have routine measurements of contrast sensitivity in a standardised situation throughout life."

We hope that the new chart whose design has been described and explained here may make the routine measurement of contrast sensitivity a real possibility.

Acknowledgements—This research was supported by National Eye Institute grant EY04432 to DGP.

REFERENCES

Arden G. B. (1978) The importance of measuring contrast sensitivity in cases of visual disturbance. *Br. J. Ophthal.* **62,** 198–209.

Bailey I. L. (1982) Simplifying contrast sensitivity testing. *Am. J. Optom. Physiol. Opt.* **59,** 12P.

Bailey I. L. (1987) Mobility and visual performance under dim illumination. In *Night Vision: Current Research and Future Directions* (Symposium Proceedings). National Academy Press, Washington, DC.

Campbell F. W. and Robson J. G. (1968) Application of Fourier analysis to the visibility of gratings. *J. Physiol., Lond.* **197,** 551–566.

Della Sala S., Bertoni G., Somazzi L., Stubbe F. and Wilkins A. J. (1985) Impaired contrast sensitivity in diabetic patients with and without retinopathy: a new technique for rapid assessment. *Br. J. Ophthal.* **69,** 136–142.

Ginsburg A. P. (1978) Visual information processing based upon spatial filters constrained by biological data. Ph.D. dissertation, Cambridge Univ. Reprinted as AFAMRL Tech. Rep. 78–129, Library of Congress 79–600156.

Ginsburg A. P. (1984) A new contrast sensitivity vision test chart. *Am. J. Optom. Physiol. Opt.* **61,** 403–407.

Hecht S., Hendley C. D., Frank S. and Schlaer S. (1949) Contrast discrimination charts for demonstrating the effect of anoxia on vision. *J. opt. Soc. Am.* **39,** 922–923.

Higgins K. E., Jaffe M. J., Coletta N. J., Caruso R. C. and deMonasterio F. M. (1984) Spatial contrast sensitivity: importance of controlling the patient's visibility criterion. *Arch. Ophthal.* **61,** 403–407.

Hyvärinen L., Rovamo J., Laurinen P., Saarinen J. and Näsänen R. (1983) Contrast sensitivity in monocular glaucoma. *Acta ophthal.* **61,** 742–750.

Legge G. E., Pelli D. G., Rubin G. S. and Schleske M. M. (1985) Psychophysics of reading—I. Normal vision. *Vision Res.* **25,** 239–252.

Legge G. E., Rubin G. S. and Leubker A. (1987) Psychophysics of reading—V. The role of contrast in normal vision. *Vision Res* **27,** 1165–1178.

National Academy of Sciences–National Research Council (1980) Recommended standard procedures for the clinical measurement and specification of visual acuity. Report of working Group 39. *Adv. Ophthal.* **41,** 103–148.

Pelli D. G. (1987) Programming in PostScript: imaging on paper from a mathematical description. *BYTE* **12**(5), 185–202.

Pelli D. G., Rubin G. S. and Legge G. E. (1986) Predicting the contrast sensitivity of low-vision observers. *J. opt. Soc. Am.* **A3,** P56.

Pelli D. G., Legge G. E. and Rubin G. S. (1988) Clinical measurement of contrast sensitivity. *Invest. Ophthal. visual Sci.* (Submitted).

Ratliff F. (1965) *Mach Bands: Quantitative Studies on Neural Networks in the Retina*. Holden-Day, San Francisco, Calif.

Regan D. and Neima D. (1983) Low-contrast letter charts as a test of visual function. *Ophthalmology* **90,** 1192–1200.

Regan D. and Neima D. (1984) Low-contrast letter charts in early diabetic retinopathy, ocular hypertension, glaucoma and Parkinson's disease. *Br. J. Opthal.* **68,** 885–889.

Robson J. G. (1983) Frequency domain visual processing. In *Physical and Biological Processing of Images*, edited by Braddick O. J. and Sleigh A. C. Springer, Berlin.

Rubin G. S. (1988) Reliability and sensitivity of clinical contrast sensitivity tests. *Clin. Vision Sci.* **2**(3), 169–177.

Schade O. (1956) Optical and photoelectric analog of the eye. *J. opt. Soc. Am.* **46,** 721–739.

Selwyn E. W. H. (1948) The photographic and visual resolving power of lenses. *Photographic J.* **88B,** 6–12.

Sloan L. (1959) New test charts for the measurement of visual acuity at far and near distances. *Am. J. Ophthal.* **48,** 807–813.

Snellen H. (1862) *Scala tipographia per mesurare il visus*. Utrecht.

Vaegan and Halliday B. L. (1982) A forced-choice test improves clinical contrast sensitivity testing. *Br. J. Ophthal.* **66,** 477–491.

Verbaken J. J. and Johnston A. W. (1986) Population norms for edge contrast sensitivity. *Am J. Optom. Physiol. Opt.* **63,** 724–732.

Watson A. B., Barlow H. B. and Robson J. G. (1983) What does the eye see best? *Nature, Lond.* **302,** 419–422.

Weibull W. (1951) A statistical distribution function of wide applicability. *J. appl. Mech.* **18,** 293–297.

Wilkins A. J., Della Sala S., Somazzi L. and Nimmo-Smith I. (1988) Age-related norms for the Cambridge Low Contrast Gratings, including details concerning their design and use. *Clin. Vision Sci.* **2**(3), 201–212.

Wolkstein M., Atkin A. and Bodis-Wollner I. (1980) Contrast sensitivity in retinal disease. *Ophthalmology* **87,** 1140–1149.

Young G. (1918) Threshold tests. *Br. J. Ophthal.* **2,** 384–392.