

Contrast Sensitivity Function as a Screening Test: A Critique

GORDON E. LEGGE*
GARY S. RUBIN*†

Department of Psychology, University of Minnesota, Minneapolis, Minnesota

ABSTRACT

Currently, there is intense clinical, commercial, and academic interest in the potential value of the contrast sensitivity function (CSF) for detecting eye disease. This paper contains an evaluation of the CSF as a screening test. Questions are raised concerning its scoring, accuracy, reliability, and robustness in screening situations. We conclude that the CSF cannot be of much value in visual screening until these questions are answered.

Key Words: contrast sensitivity function, visual screening, visual acuity

The CSF is a vision test that uses sine wave grating stimuli rather than letters or disks. As illustrated in Fig. 1, sine wave gratings can vary in spatial frequency (bar width) and contrast. The test consists of reducing the contrast of a grating until the threshold of visibility is reached. Contrast thresholds of this sort are obtained for gratings of different spatial frequencies, e.g., 0.5 to 30 c/deg. Contrast sensitivity is defined to be the reciprocal of threshold contrast. This means that a low threshold contrast corresponds to a high sensitivity.

Fig. 2 shows the CSF for a normal, adult subject. Spatial frequency is plotted along the horizontal axis on a logarithmic scale. Contrast thresholds are plotted on the right vertical axis and corresponding contrast sensitivities on the left vertical axis. Both vertical scales are logarithmic. For normal observers, the CSF peaks at an intermediate spatial frequency and declines for both high and low spatial frequencies.

Typically, the highest spatial frequency that can be seen at maximum contrast lies between 30 and 60 c/deg. This "high-frequency cutoff" is a measure of visual resolution and is closely related to conventional acuity. In a study of 93 patients with various vision disorders, Virsu et al.¹ reported a correlation of 0.84 between the high-frequency cutoff and Snellen acuity.

The CSF has been of great value in laboratory studies of vision. The way in which the shape of the CSF changes, depending on stimulus factors such as luminance or flicker, has informed us about the normal visual process. Differences between the CSF's of adults and infants tell us about the development of pattern vision. Comparisons of CSF's across species ranging from the eagle to the goldfish have shown us how pattern vision varies in the animal kingdom.

It has been suggested that the CSF offers promise as a clinical test because it provides information not available from traditional tests. For example, abnormalities in the CSF have been found in some patients with multiple sclerosis² and cerebral lesions³ who appear normal by acuity tests and by ophthalmoscopic examination. The CSF may be useful as (1) a diagnostic aid, (2) a means for evaluating the nature and severity of visual impairment, and (3) a screening test for ocular disorders. This critique deals only with the suitability of the CSF as a screening test.

CSF AS A SCREENING TEST

A visual screening battery consists of tests of several aspects of visual function. Its purpose is to separate people with normal vision from those with visual disorders. Screenings may be administered in schools, residences for the elderly, or public places such as shopping malls. People who fail are referred to eye care specialists for full examinations. A practical screening battery must consist of tests that are quick, cheap, easy to score and administer, and sensitive.

Should the CSF be included in screening pro-

Received July 17, 1985; revision received November 7, 1985.

* Ph.D., Member of Faculty.

† Current address: Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205.

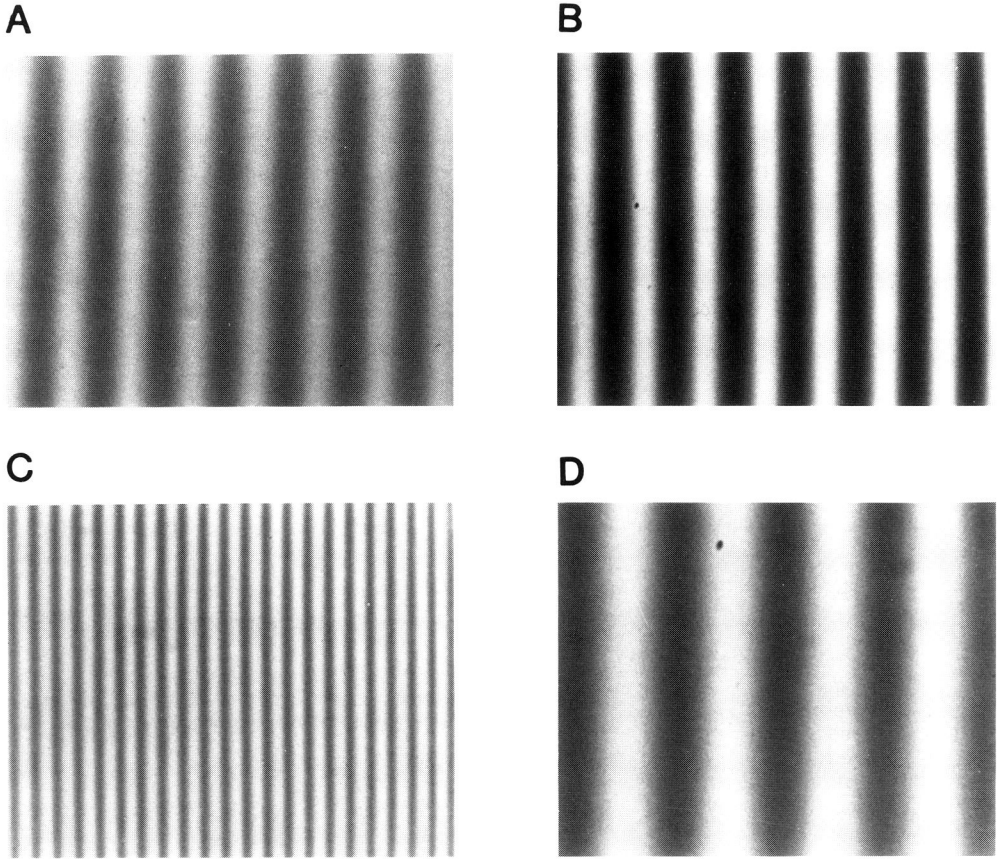


FIG. 1. Four examples of sine wave gratings are shown. Patterns A and B differ only in contrast, whereas C and D differ only in spatial frequency. The contrast of the pattern is defined to be $(L_{max} - L_{min}) / (L_{max} + L_{min})$ where L_{max} and L_{min} are the maximum and minimum luminances of the wave-form. Contrast ranges from a minimum of 0 to a maximum of 1.0. The pattern's spatial frequency is the number of light and dark cycles per degree of visual angle.

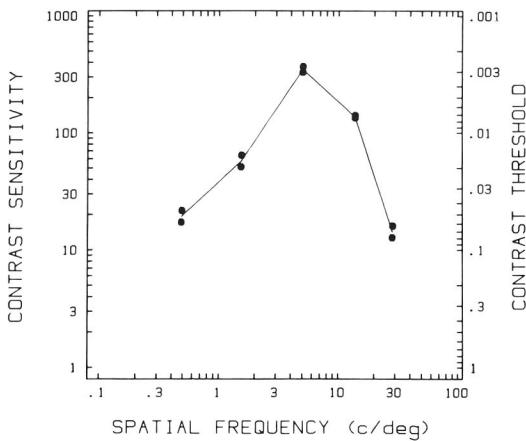


FIG. 2. CSF for a normal observer. Contrast sensitivities, on the left axis, are plotted as a function of spatial frequency. Corresponding contrast thresholds are shown on the right axis. The solid line connects the mean of two contrast sensitivity measurements collected using a forced-choice procedure.

grams? The only scientifically valid reason for including the CSF would be that it markedly increases the accuracy of the screening battery.

ACCURACY OF A SCREENING TEST

We may think of a screening test as having four possible outcomes. The presenting subject has either normal or abnormal vision and he or she either passes or fails the test. Abnormal subjects who fail the test are "true positives" and those who pass are "false negatives." Normal subjects who fail are "false positives" and those who pass are "true negatives."

Fig. 3 presents results from a hypothetical screening based on acuity testing only. Suppose that the test is given to 1000 subjects, 850 of whom are normal and 150 of whom are abnormal. Suppose that the failure criterion is an acuity of 6/15 or worse, and that the corresponding true-positive and false-positive rates are 90% and 1%, respectively. These two percentages jointly provide a measure of the test's accuracy.⁴

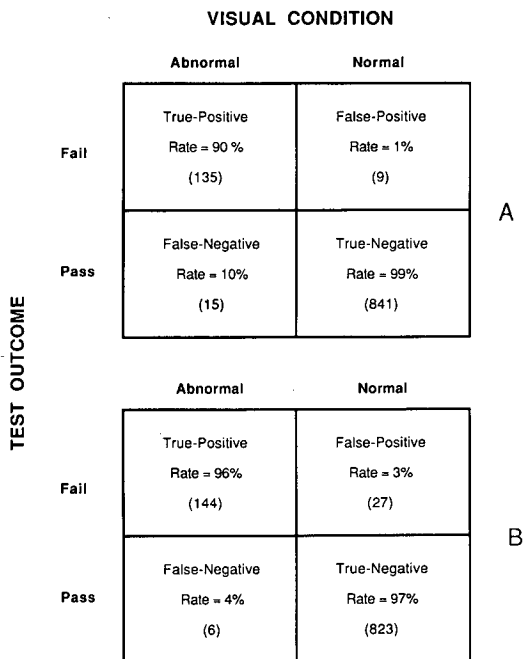


FIG. 3. Results from a hypothetical screening test using two different referral criteria. In the example, 1000 subjects were screened; 850 were normal and 150 were abnormal. The cells of the table indicate the true- and false-positive rates, and the true- and false-negative rates. The numbers in parentheses show the number of subjects falling in each of the four outcome categories. The true- and false-positive rates together determine the accuracy of the test. In panels A and B, test accuracy is the same (according to the d_a' measure of accuracy defined by Swets and Pickett⁴). However, the referral criterion is more inclusive in panel B than panel A, so there are more true positives and false positives.

They indicate that 135 people will be correctly referred, and about 9 people will be incorrectly referred for eye examinations. On the other hand, 10% of the abnormal group were undetected by the test, that is, 15 subjects passed who should have failed.

In order to catch some or all of the 15 missed subjects, we may modify the screening procedure in one of two ways: by changing the referral criterion or by somehow increasing the accuracy of the test.

Suppose we change the referral criterion from an acuity of 6/15 to 6/12. Suppose this increases the true-positive rate to 96% and the false-positive rate to 3%. The test now correctly identifies 144 of the 150 abnormal subjects, and misses only 6. However, it also falsely refers 27 normal subjects rather than 9 (see Fig. 3B).

The examples summarized by Fig. 3 indicate that for a test of fixed accuracy, we may increase the true-positive rate to whatever high value we wish, but at the expense of increasing the false-

positive rate as well. A much more appealing approach would be to increase the true-positive rate while keeping the false-positive rate at a low, acceptable level. Such an improvement requires an increase in the accuracy of the screening test. This cannot be accomplished simply by changing the referral criterion. Instead, we must modify the screening test in some way that increases its accuracy.

Inclusion of the CSF as part of a screening battery might increase accuracy by raising the true-positive rate while keeping the false-positive rate at a low, acceptable value. This would be the case if the CSF itself has high accuracy and if it provides nonredundant information. To date, no one has determined whether this would be the case or not. We know neither the screening accuracy of the CSF alone nor its effect on the accuracy of screening batteries where it is included as a component.

SCORING A CSF

The CSF is a multivalued test. Typically, contrast sensitivities are measured at four to eight spatial frequencies. How should these measures be combined to form an appropriate screening criterion?

Consider the CSF's shown in Fig. 4. The solid curve connecting the circles represents average data from a group of normal subjects. The bars represent ± 1 SD. The squares are contrast sensitivities for subjects suffering from Huntington's disease.⁵ There is no question that the CSF's for subjects MB and CJ are subnormal. All data points lie more than 1 SD below the normal mean. The case is not quite so clear for KD. Some of his sensitivities lie within 1 SD of normal. Subject LB is even more problematic. His CSF appears to be subnormal throughout, but his mean values only lie outside the error bars once. Should his vision be classified as normal or abnormal?

We are faced by the problem posed by a multivalued test. What sort of screening criterion should we choose? We might begin by collecting a large sample of normative data (see, e.g., Ginsburg et al.⁶). We could use these data to establish lower bounds on normal contrast sensitivity—say, for example, values of contrast sensitivity at each spatial frequency exceeded by 90% of normal subjects. Consider a CSF based on seven spatial frequencies for a test subject. Suppose, for the sake of illustration, that the subject is normal, the seven measurements are independent, and that for each measurement he has a 90% chance of having a contrast sensitivity within the normal category. The probability that at least one of the seven measurements will lie below the lower bound is $1 - (0.9)^7 = 0.52$. In other words, in this simple example, a normal

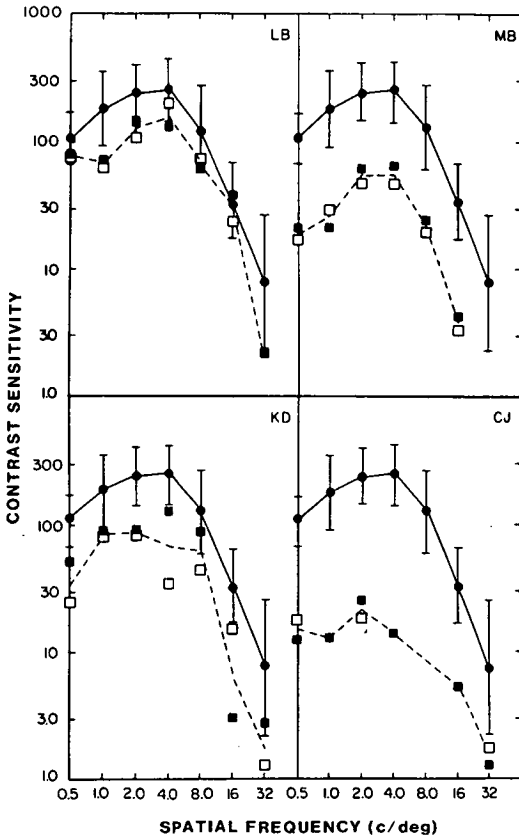


Fig. 4. CSF's for subjects with Huntington's disease.⁵ In each panel, the filled circles, connected by solid lines, represent data for a normal control group. The error bars indicate ± 1 SD. The open and closed squares represent repeated measurements of contrast sensitivity for individual patients. The dashed lines connect mean values. Each sensitivity was estimated using a forced-choice procedure.

subject has about a 50% chance of appearing "abnormal" for at least one spatial frequency. Even if we increase the norms to the 99th percentile, there remains a 7% chance for the normal subject to appear abnormal on at least one measurement. Gathering reliable norms out to the 99th percentile or more is likely to be impractical. It seems we must turn to some alternative means for assessing abnormality of the CSF.

This problem also confronts other multivalued tests of vision. For example, the Farnsworth Panel D-15 test of color vision requires a subject to order 15 color chips serially according to their similarity of color. The designation of color abnormality is based on the number and type of errors. It is possible that screening criteria of this sort could be established for the CSF, e.g., three of seven measures lying more than 1 SD

below the normal mean, or two lying more than 2 SD's below the mean, etc. Whatever choices of screening criteria are made, they will require: (1) suitable normative data, and (2) evaluation of the true-positive and false-positive rates they produce. Although several studies have compared CSF's of normal and control groups with those of clinical populations, no one has fully addressed this scoring problem.

WOULD A SIMPLIFIED TEST OF CONTRAST SENSITIVITY WORK?

One way of dealing with the multiple measures problem would be to reduce the CSF test to a single measurement. For example, we might measure contrast sensitivity for a single sine wave grating. Alternatively, we might assess contrast sensitivity by measuring the lowest contrast at which subjects can identify letters on a test chart. With only a single measure to worry about, the problem of scoring is simplified. We need only set some criterion and proceed to evaluate the accuracy of the test as described above. Some studies have taken this approach (see, e.g., Atkin et al.⁷).

Whether one or many measures of contrast sensitivity should be taken is controversial. It may be the case that some visual disorders that have no effect on acuity produce sensitivity loss in a narrow band of spatial frequencies.^{2,8} A priori, we have no idea where to look for, a narrow band sensitivity loss, so we would have to measure contrast sensitivities at suitably small steps of spatial frequency across a wide range. An alternative view holds that most eye diseases will cause at least some sensitivity loss across a fairly broad range of spatial frequencies, usually with greater loss at high spatial frequencies than at low. In addition to acuity, only a single measure of contrast sensitivity, possibly near the peak of the CSF, may be sufficient to identify such losses. We are left with an unanswered question. At how many spatial frequencies should contrast sensitivities be measured if the test is to be useful for screening?

Sometimes, subjects with normal acuity complain of disturbed vision. It has been reported^{2,3,8,9} that some of these subjects exhibit sensitivity losses in the CSF. This suggests a sequential screening strategy in which the CSF test is administered only to subjects with normal acuity who complain of disturbed vision.^a To validate such an approach, we would need to measure the correlation between the existence of subjective visual complaints and presence of

^a The sequential screening strategy was suggested to us by Jonathan Wirtschafter.

abnormalities of the CSF in cases of normal acuity.

ROBUSTNESS OF CSF MEASUREMENTS

Procedural differences can affect the reliability of CSF measurements. In some versions of the test, subjects must select their own criterion of visibility. The test administrator (or possibly the subject) adjusts the contrast of the grating up or down until the subject declares that threshold has been reached. Such a procedure is fraught with response bias. A conservative subject will require much higher contrast before declaring the presence of the pattern than a subject who wants to "beat the test." Moreover, if a subject's criterion changes between two administrations of the test (or within one testing session), differences in sensitivity will be obtained that are unrelated to his or her visual condition. A preferable test design would use a procedure that is less dependent on subjective visibility criteria. The forced-choice method has become the procedure of choice. Typically, the subject is shown two displays, one blank and the other containing the grating. The subject is not required to judge the visibility of the grating, but only to choose in which of the two displays it appeared. The subject's response can be scored right or wrong by the experimenter. Contrast of the grating is reduced according to some algorithm until a contrast level is reached at which the subject makes some specified proportion of errors. This contrast level is taken to be the threshold contrast. Comparison of the forced-choice procedure with the method of adjustment has revealed that the former exhibits less inter-subject variability among normals and better repeat reliability for individual subjects.¹⁰

Many studies have shown that the shape and overall height of the CSF vary with test conditions. Some of these conditions are inherent in the test itself, e.g., number of cycles in the grating patterns, orientation of patterns, and time course of presentation. Some of the conditions may depend on the environment in which the test is administered, e.g., nature of illumination. Finally, subject variables also play a role, even among normal populations, e.g., practice, monocular vs. binocular viewing, pupil size, and state of refraction. Age variation is particularly important. There is evidence that properties of CSF's change with age.^{11,12} In a practical screening context, some of these variables will be difficult to measure or control. Any deviation from controlled conditions will introduce variability into the data and reduce the accuracy of the screening test. We simply do not know how much variability in CSF measurements will be introduced by the lack of control inherent in

practical screening situations. We do not know how robust the CSF test is.

CONCLUSIONS

Before the CSF can be recommended as a practical screening test, we need answers to the following questions:

1. How accurately does the CSF distinguish subjects with abnormal vision from those with normal vision, either on its own or in conjunction with conventional test measures?
2. How do we score a CSF? What criteria do we use to separate abnormal from normal CSF's?
3. How many measures of contrast sensitivity are necessary to make the test accurate enough to be of use in screening? At what spatial frequencies should these measurements be taken?
4. How robust are measurements of contrast sensitivity to the types of unavoidable variability in testing conditions typical of screening contexts? What sort of repeat reliability is expected of the CSF?

It is our view that the benefits of the CSF as a screening test will remain unknown until these questions are answered.

ACKNOWLEDGMENTS

Preparation of this paper was supported by U.S. Public Health Service Grant EY02934 from the National Eye Institute, National Institutes of Health, Bethesda, Maryland. We thank Denis Pelli, Jonathan Wirtschaffer, George Woo, and Ivan Bodis-Wollner for helpful comments. We thank Maureen Karpan and Mary Schleske for help with manuscript preparation.

REFERENCES

1. Virsu V, Lehtiö P, Rovamo J. Contrast sensitivity in normal and pathological vision. *Doc Ophthalmol Proc Ser* 1981;30:263-72.
2. Regan D, Silver R, Murray TJ. Visual acuity and contrast sensitivity in multiple sclerosis-hidden visual loss. *Brain* 1977;100:563-79.
3. Bodis-Wollner I, Diamond SP. The measurement of spatial contrast sensitivity in cases of blurred vision associated with cerebral lesions. *Brain* 1976;99:695-710.
4. Swets JA, Pickett R. *Evaluation of Diagnostic Systems*. New York: Academic Press, 1982.
5. Pieper CM. Visual contrast sensitivity of patients with Huntington's disease. 1980. Undergraduate Honors Thesis. College of Biological Sciences, University of Minnesota.
6. Ginsburg AP, Evans DW, Cannon MW, Owsley C, Mulvanny P. Large-sample norms for contrast sensitivity. *Am J Optom Physiol Opt* 1984;61:80-4.
7. Atkin A, Bodis-Wollner I, Wolkstein M, Moss A, Podos SM. Abnormalities of central contrast sensitivity in glaucoma. *Am J Ophthalmol* 1979; 88:205-11.
8. Bodis-Wollner I. Visual acuity and contrast sensitiv-

ity in patients with cerebral lesions. *Science* 1972;178:769-71.

9. Arden GB. Visual loss in patients with normal visual acuity. *Trans Ophthalmol Soc UK* 1978;98:219-31.
10. Higgins KE, Jaffe MJ, Coletta NJ, Caruso RC, de Monasterio FM. Spatial contrast sensitivity. Importance of controlling the patient's visibility criterion. *Arch Ophthalmol* 1984;102:1035-41.
11. Owsley C, Sekuler R, Siemsen D. Contrast sensitivity throughout adulthood. *Vision Res* 1983;23:689-99.
12. Sokol S, Domar A, Moskowitz A. Utility of the

Arden grating test in glaucoma screening: high false-positive rate in normals over 50 years of age. *Invest Ophthalmol Vis Sci* 1980;19:1529-33.

AUTHOR'S ADDRESS:

Gordon E. Legge
Department of Psychology
University of Minnesota
Elliott Hall
75 East River Road
Minneapolis, Minnesota 55455

Coming Event

A program titled *Macular Degeneration: Lasers, Light Toxicity, and Low Vision*, will be held on May 16, 1986, at The Columbia Inn, Columbia, MD. The program is sponsored by The Wilmer Institute of The Johns Hopkins Medical Institutions.

For further information, contact: Diane Heydinger, Office of Continuing Education, The Johns Hopkins Medical Institutions, Turner 22, 720 Rutland Ave, Baltimore, MD 21205 (301) 955-6046.