Evaluating the performance of the quick CSF method in detecting CSF changes: an assay calibration study

Fang Hou¹, Luis Lesmes², Woojae Kim¹, Hairong Gu¹, Mark Pitt¹, Jay Myung¹, Zhong-Lin Lu¹

¹ Department of Psychology, The Ohio State University, Columbus, OH;
² Adaptive Sensory Technology, Boston, MA
Abstract

Purpose
The contrast sensitivity function (CSF) has shown promise for monitoring the changes in functional vision that accompany eye disease or its treatment. In this assay calibration study, we focus on the challenge of detecting CSF changes with precision and efficiency. We exploit the Bayesian foundation of the quick CSF 1 to develop and evaluate metrics for detecting changes in the contrast sensitivity function.

Methods
A 10-letter identification task was used to assess the systematic changes in the CSF measured in three luminance conditions in 112 naïve observers with self-reported normal vision. The reliability of CSF metrics was calculated both within- and between test sessions. In addition, the sensitivity, specificity and accuracy for detecting CSF changes in individuals were evaluated. Finally, we conducted empirical statistical power analyses for detecting CSF changes in groups of observers.

Results
We found that the standard error of the CSFs obtained with the quick CSF method was less than 0.1 log unit after 50 trials. The test-retest reliability was greater than 0.9 after 15 quick CSF trials, and reached 0.974 with 50 trials. In 50 trials, the quick CSF method can detect area under log CSF (AULCSF) changes caused by a 7.8 and 36.4 fold luminance change with 94.0% and 98.9% accuracy, respectively. A power analysis showed that a very small contrast sensitivity change (0.025 log unit or 6%) could be detected with the quick CSF method with 112 observers and 50 trials.
Conclusions

The quick CSF is very precise, highly reliable and extremely sensitive in detecting CSF changes at both individual and group levels. These advantages make it plausible to apply the method to monitor the progression of visual diseases or treatment effects on individual patients, and greatly reduce the time, sample size and costs in clinical trials.
Introduction

“... Without instruments that could measure such small improvements in performance, invention was doomed to be rare and erratic. ... the mania for precision that was Maudslay’s defining characteristic made it commonplace.”

Science and technology advance through gradual, quantitatively incremental improvements of design, which can eventually accumulate to qualitative revolutions. Precise instruments are needed to detect the small changes in performance that are needed to drive the iterative design process in the correct direction. The bench micrometer built by Henry Maudslay in the early 1800s provides a signature example of how accuracy and precision contributes critically to technological development. In addition to the technical innovation needed for its development, the micrometer represented a broader culture of precision that inspired the invention and development of the steam engine and the progress of the industrial revolution.

In clinical vision science, the lack of precise “instruments” for evaluating functional vision change has been a major limiting factor in the diagnosis and treatment of eye disease. Current shortcomings in functional vision endpoints include unsatisfactory precision, poor test-retest reliability, and insensitivity to visual changes caused by conditions/diseases. Best-corrected visual acuity (BCVA), the most common functional endpoint in clinical vision, may not be ideal in characterizing visual deficits and is insensitive to early stages of eye diseases and progressive disabilities in functional vision. The Humphrey visual field test, another widely used tool in
clinical vision, also exhibits poor precision. It has been reported that 85.9% of the initial
abnormal visual fields assessed in the Ocular Hypertension Treatment Study were not verified on retest\textsuperscript{13}.

As a fundamental assay of spatial vision that describes performance over a wide range of spatial frequencies, the contrast sensitivity function (CSF) exhibits promise as a functional vision endpoint\textsuperscript{14}. The CSF is closely related to daily functional vision, which can be used to characterize visual performance in both normal and impaired vision\textsuperscript{3, 4, 10, 14–19}. It has also shown promise for monitoring the progression of vision loss in eye diseases or their remediation with treatment\textsuperscript{20}.

Despite its clinical promise, the precise and efficient assessment of the CSF is challenging. The conventional CSF tests can be either fast or precise, but not both. The various CSF charts are fast and convenient to use, but the sparse sampling of both spatial frequency and stimulus contrast renders the tests rather imprecise\textsuperscript{21–25}. The laboratory CSF tests can be precise but may require 500–1000 trials and take about 30–60 minutes\textsuperscript{26, 27}. Such testing times might be acceptable for measuring a single CSF in the laboratory, but are prohibitive in situations requiring assessment of CSF in clinical settings, especially when multiple CSFs need to be assessed (e.g., for left and right eyes).

In an attempt to achieve both high precision and efficiency in CSF assessment, Lesmes et al\textsuperscript{1} developed the \textit{quick CSF} method, which applies a Bayesian adaptive algorithm to estimate the full shape of the contrast sensitivity function. They showed that CSFs measured with the \textit{quick CSF} method were precise and exhibited excellent agreement with CSFs obtained independently using a conventional method. Recently, the
efficiency of quick CSF method was further improved another 2.5 times by incorporating a 10-alternative forced choice (10AFC) letter identification task. The CSFs obtained with the quick CSF method have also been shown to be precise and accurate in many other situations, other experimental conditions, including peripheral vision, second-order perception, and in other clinical populations, such as patients with amblyopia, age-related macular degeneration, glaucoma, and normal age-related visual changes. The quick CSF method has been utilized in various empirical studies to investigate the dynamic effects of visual adaptation, and effects of emotional arousal on CSF. With the help of the quick CSF method, Kalia et al. found surprising visual development in a unique sample of patients who experienced extended early-onset blindness before removal of bilateral cataracts.

Although it has demonstrated great potential as a clinical tool, the quick CSF method has been tested with relatively small samples or under a single experimental condition. Its precision and test-retest reliability have yet to undergo more rigorous tests. Furthermore, the performance of the quick CSF method in detecting CSF changes between different conditions has not been extensively studied.

The ability to detect CSF changes in individuals provides the important bases for diagnosing the progression of vision loss in eye disease or the treatment response. At the group level, the ability to detect CSF changes can provide decisive evidence to judge the safety and efficacy of treatments. As a test case for detecting CSF changes in different visual conditions, we applied the quick CSF method to assess contrast sensitivity changes in three different luminance conditions (2.62, 20.4 and 95.4 cd/m²). This manipulation
was motivated by several considerations. First, CSF changes in different luminance conditions may be important for diagnosis of age-related macular degeneration \(^{40-42}\). Second, CSF changes in low luminance conditions are especially informative for aging vision, due to the prominent high-spatial frequency sensitivity loss due to age \(^{43-45}\). Third, similar luminance manipulation has been used as a benchmark test for judging the performance of several CSF methods \(^{39, 46}\).

To comprehensively investigate the performance of quick CSF, we recruited 112 normal college students in our study. The large sample not only increased the external validity of our results, it also allowed us to analyze the performance of the quick CSF in change detection with minimal assumptions. With CSF data from this sample, we evaluated the precision, test-retest reliability of the quick CSF method as well as its sensitivity, specificity and accuracy in detecting individual CSF change. Power analyses for detecting CSF changes based on the empirical data were also conducted. We show that the quick CSF method is very precise, highly reliable and extremely sensitive. The high accuracy and great power of the method make it possible for vision researchers and clinicians to measure CSFs efficiently and detect CSF changes with greatly reduced sample sizes and costs in experiments.

**Methods**

**Observers**

112 college students from The Ohio State University participated in the study to obtain partial course credit in an Introductory Psychology course. All observers had normal or corrected-to-normal vision, and were naive psychophysical observers. Verbal
consent was obtained before the experiment. The study protocol was approved by the institutional review board of human subjects research of the Ohio State University and adhered to the tenets of the Declaration of Helsinki.

Apparatus

All programs were written in Matlab (The Mathworks Corp., Natick, MA) with Psychtoolbox subroutines (Kleiner, Brainard, & Pelli, 2007) and run on a PC computer. Stimuli were displayed on a gamma-corrected Samsung UN55FH6030 55” LED TV with a mean luminance of 95.4 cd/m². The resolution was a 1920 ×1080 pixels and the vertical refresh rate was 60 Hz. A bit-stealing algorithm was used to achieve 9-bit gray-scale resolution. Participants viewed the display binocularly from a distance of 4 m in a dark room.

Four different viewing conditions were tested in the experiment: low luminance (L), medium luminance (M), high luminance (H) and low pass (LP) conditions. In the H condition, subjects viewed the display through uncovered goggles. In the M condition, subjects viewed the display binocularly through goggles with neutral density filters with an attenuation factor of 0.67 decimal log units. In the L condition, subjects viewed the display through goggles fit with two neutral density filters with a total optical density of 1.56 log units. Bangerter occlusion foils were used as the low pass filter in the LP condition. The data in the LP condition are analyzed separately and not presented in this paper. The equivalent mean luminance in the L, M, H and LP conditions was 2.62, 20.4, 95.4 and 95.4 cd/m², respectively. Compared to the L condition, there were 7.8 and 36.4 folds of luminance change in the M and H conditions, respectively. The magnitudes of luminance change were comparable to those used in previous studies.
Stimuli

Ten filtered Sloan letters, C, D, H, K, N, O, R, S, V and Z, were used as stimuli. All letter stimuli had a center frequency of 3.3 cycles per object (cpo) and a full bandwidth of one octave. The pixel intensity of each filtered image was normalized by the maximum absolute intensity of the image. After normalization, the maximum absolute Michelson contrast of the image is 1.0. Stimuli with different contrasts were obtained by scaling the intensities of the normalized images with corresponding contrast values. Stimuli with different spatial frequencies were generated by resizing. There were 128 possible contrasts (evenly distributed in log space from 0.002 to 1) and 19 possible spatial frequencies (evenly distributed in log space from 1.19 to 30.95 cycles per degree, cpd). The narrow band filtered letters can provide assessment of contrast sensitivity in different central spatial frequencies that are equivalent to that with sinewave gratings.
Figure 1. a) Ten filtered Sloan letters. b) Illustration of filtered letter ‘C’ in several spatial frequency conditions. c) tri-letter stimuli

Implementation of the quick CSF method

In the quick CSF method, CSFs are characterized by a truncated log parabola (Figure 2a) with four parameters: peak gain $g_{\text{max}}$, peak spatial frequency $f_{\text{max}}$, bandwidth at half-height $\beta$ (in octaves), and low-frequency truncation level $\delta$. In the rest of the paper, we use “CSF parameters” interchangeably with “truncated log parabola parameters”, unless otherwise stated. Unlike many conventional methods that select stimuli adaptively only in the contrast space, the quick CSF method selects optimal stimuli in the two-dimensional contrast and spatial frequency space (Figure 2b) by
maximizing the information about the to-be-measured CSF in each trial. Using a Bayesian adaptive algorithm \cite{27, 52-61} to select the optimal test stimulus prior to each trial and update the posterior probabilities of CSF curve parameters following observer’s response, the quick CSF method directly estimates the entire CSF curve instead of contrast thresholds or contrast sensitivities at discrete spatial frequencies (See Appendix A for more details).

Figure 2. The quick CSF method adopts a four parameter truncated log parabolic functional form to describe the shape of the contrast sensitivity function (a) and selects the optimal test stimulus in each trial in the two-dimensional contrast and spatial frequency space (b).

**Design**

At the start of the experiment, observers were given five minutes to dark-adapt to the test environment. The experiment consisted of six blocks of quick CSF measurements, each in one luminance/filter condition. The order of the test blocks was L, L, M, H, LP and H. The first L condition was used for observers to dark-adapt and practice the quick CSF test, data from which were not analyzed here. The two H conditions were included
to assess the test-retest reliability of the quick CSF method, and are labeled as H1 and H2 in the rest of the paper. In each test block, the quick CSF procedure with a 10AFC letter identification task was used to measure the CSF in 50 trials. Each observer finished one experimental session, which included six distinct quick CSF runs, in approximately 70 minutes.

**Procedure**

At the beginning of each trial, the quick CSF method selects the optimal stimulus (contrast and spatial frequency) by maximizing the expected information gain in that trial. To improve observers’ experience in CSF testing, two stimuli with higher contrasts in addition to the optimal stimulus were also presented in each trial. The three letters were independently chosen from the 10 Sloan letter set and presented in a row with a center-to-center distance of 1.1 times letter size (Figure 1c). The three letters had the same spatial frequency, but differed in contrast. From left to right, the contrast of the letter stimuli was 4, 2 and 1 times the optimal contrast respectively, with the maximum contrast capped at 0.9. Observers were asked to verbally report the identities of the letters presented on the screen to the experimenter, who used the computer keyboard to enter the observers’ verbal responses. The stimuli disappeared after all responses were entered. Observers were given the option to report “I don’t know”, upon which the response was coded as “incorrect”. No feedback was provided. All three responses were used to update the posterior distribution of the CSF curve parameters (Appendix A). A new trial started 500 ms after the responses.
**General analysis**

For each quick CSF assessment, the posterior distribution of CSF parameters was numerically converted to the posterior distribution of the contrast sensitivities that define the CSF curve. The procedure automatically takes into account the covariance structure in the posterior distribution of the CSF parameters. Specifically, 1,000 sets of truncated log-parabola parameters were sampled from the posterior distribution, $p_t(\theta)$, and used to construct 1,000 CSF curves. Each CSF curve was represented by contrast sensitivities sampled at 19 spatial frequencies ranging from 1.19 to 30.95 cpd, evenly distributed in log space. We then obtained the empirical distributions of the CSF from these 1,000 CSF curves.

In addition to the CSFs, the posterior distribution of the area under the log CSF (AULCSF) was converted from $p_t(\theta)$ in the same way. The AULCSF is a summary measure of spatial vision and was calculated as the area under log CSF curve (and above zero) in the spatial frequency range of 1.5 to 18 cpd.

The estimated CSF and AULCSF metrics were calculated as respective means of their posterior distributions in a quick CSF run. The within-run variability of CSF and AULCSF estimates can be evaluated as the half width of the 68.2% credible interval (HWCI) of their respective posterior distributions: (Figure 3a).

$$\text{HWCI}_{68.2} = \frac{P^{-1}(0.841) - P^{-1}(0.159)}{2},$$  \tag{1}
where $P^{-1}(\cdot)$ is the empirical inverse cumulative distribution function of the posterior for CSF or AULCSF. Because CSF was sampled at 19 spatial frequencies, the average HWCI of CSF across 19 spatial frequencies was reported for each individual.

The variability of CSF estimates can also be computed in another way, i.e. the standard deviation of the estimated CSFs in repeated measurements. Given that there were only two repeated CSF measurements in the H condition, we computed the mean distance between two CSF measures for each participant:

$$D = \frac{\sqrt{2} \sum_{i=1}^{19} |\tau_{i1} - \tau_{i2}|}{19},$$

where $\tau_{i1}$ and $\tau_{i2}$ are estimated CSFs (i.e., posterior means) at the $i$th ($i = 1, 2, \ldots, 19$) spatial frequency in the H1 and H2 conditions, respectively (Figure 3b). $\sqrt{2}/2$ is a correction coefficient to make the distance essentially the same as the standard deviation of estimated CSF in two repeated measurements.
Figure 3. a) Illustration of the mean and 68.2% credible interval of a posterior distribution.

b) The distance between two sensitivity measurements, \( \tau_1 \) and \( \tau_2 \), at a given spatial frequency. Red and blue curves are the posterior distributions of sensitivity at a given spatial frequency in the H1 and H2 conditions respectively.

**Results**

Figure 4 shows the estimated CSFs and posterior distributions of AULCSF of several representative observers (S14, S26, S86 and S107) after 50 trials in the L, M, H1, and H2 conditions. As seen in the figure, increasing luminance led to increased contrast sensitivity; the posterior distributions of AULCSF in the L, M and H conditions were well separated; the posterior distributions of AULCSF in the H1 and H2 conditions overlapped with each other.

The observations were confirmed by a within-observer ANOVA based on the data from all the 112 observers: Increasing luminance significantly increased the CSF (\( F(2, 222) = 774.6, p < 0.001 \)). The AULCSFs were significantly different in the L, M, and H conditions (\( F(2, 222) = 1517.2, p < 0.001 \)), but the CSFs and AULCSFs in the H1 and H2 conditions overlapped and were not different (\( F(1, 111) = 0.670, p = 0.415 \); \( F(1, 111) = 0.646, p = 0.423 \)). These results show that the quick CSF method was able to capture the CSF differences induced by the luminance manipulation and also suggested the high test-retest reliability.
Figure 4. Top row: Estimated CSFs in the L, M, H1, and H2 conditions of S14, S26, S86 and S107. Shaded regions indicate the 68.2% HWCI. Bottom row: Posterior distributions of estimated AULCSF in the L, M, H1, and H2 conditions of the four observers.

Figure 5a shows AULCSF as a function of background luminance level for all 112 observers. The curves that connect AULCSFs in different luminance conditions generally exhibit an apparent laminar structure: the curves have approximately the same shape but are shifted vertically relative to each other. The pattern of results suggests that changing the mean luminance had similar effects across observers. The mean AULCSF for L, M, H1 and H2 was 1.29 ± 0.15, 1.58 ± 0.13, 1.71 ± 0.12 and 1.71 ± 0.13 log units, respectively. Again, the estimated AULCSFs from the two H conditions were essentially identical. The amount of change across these conditions was calculated as the AULCSF difference between two conditions for each participant. Comparisons that were made included H1 - L, M - L, H1 - M, and H1 - H2. In the rest of the paper, the H1 - M and H1 - L comparisons are denoted as H - M and H - L and used in further analysis. The mean AULCSF difference for the H - L, M - L, H - M, and H1 - H2 comparisons was 0.43 ±
0.10, 0.29 ± 0.07, 0.14 ± 0.07, and 0 ± 0.06 log units, respectively. These mean AULCSF differences were served as the “true” magnitude of change in the later power analysis.

Figure 5. AULCSF as a function of luminance level for all 112 observers;

**Precision and test-retest reliability**

The precision of a test is inversely proportional to the variability of its measures. A more precise test would deliver less variable result. The variability of the estimated CSF can be represented by the 68.2% HWCI of the posterior distribution of CSF estimates as well as the mean distance between two repeated CSF measures. Figure 6 shows histograms of the HWCI of the CSFs of all the 112 observers in the L, M, H1, and H2 conditions after 5, 10, 20 and 50 trials. The mean of the distributions of the HWCI was 0.36, 0.20, 0.13 and 0.08 after 5, 10, 20, and 50 trials in the L condition, 0.36, 0.21, 0.13 and 0.08 in the M condition, 0.32, 0.20, 0.13 and 0.08 in the H1 condition, and 0.34, 0.19, 0.13 and 0.08 in the H2 condition, all in log units. The HWCI of the CSFs in the
four conditions were about the same. At the same time, the width of all the histograms narrowed as trial number increased: the standard deviation of HWCI decreased by an order of magnitude from 0.07 to 0.007, 0.075 to 0.007, 0.067 to 0.007, and 0.071 to 0.007 in the L, M, H1, and H2 conditions, respectively, as trial number increased from 1 to 50. The result suggests the confidence of the estimated HWCI increases as trial number increase. Although it widely varied across different observers at trial 1, the HWCI values stabilized across observers after 10 trials.

Figure 6. Histograms of the 68.2% HWCI of the posterior distribution of CSF in the L, M, H1 and H2 conditions after 5, 10, 20 and 50 trials for all the 112 observers.

Figure 7 shows histograms of the mean distances $\bar{D}$ between the two repeated measures of the CSF measured in the H1 and H2 conditions after 5, 10, 20 and 50 trials for all 112 observers. The mean of the distribution of $\bar{D}$ was 0.22, 0.16, 0.11 and 0.07 log
units after 5, 10, 20, and 50 trials. The width of the distributions narrowed as trial number increased: the standard deviation of the mean distance decreased from 0.23 to 0.03 as trial number increased from 1 to 50. It should be noted that the variability of the estimated standard deviation was greater than that of HWCI. This is likely because there were only two repeated measurements of the CSF in the H condition. Running more repeated tests would reduce this variability.

Figure 7 Histograms of the mean distance, as defined in Eq. (2), of the two estimated CSFs in the H1 and H2 conditions after 5, 10, 20 and 50 trials for all the 112 observers.

Figure 8 shows the average HWCI across subjects (in the H1 condition) and average mean distance as functions of trial number. Both measures of variability decreased quickly with number of trials. Most importantly, the two measures of variability were very similar, suggesting that the HWCI from a single quick CSF assessment is an adequate estimate that reflects the precision of the quick CSF method.
Figure 8. The average (over observers) HWCI (blue) and mean distance (red) of measured CSF are plotted as functions of trial number. Shaded area indicates ± 1 standard deviation for HWCI. The standard error of the estimated CSF is also plotted in green color.

To gauge the test-retest reliability of the quick CSF method, we computed the Pearson correlation between the CSFs obtained in the H1 and H2 conditions. If a test is extremely unstable, the correlation between two repeated measurements would be very poor. Otherwise, the correlation coefficient should be close to 1. One could argue that due to the underlying assumption of log parabola CSF model, the CSFs measured by the quick CSF method are highly constrained. In order to eliminate such constraint, we developed the following procedure: (1) randomly divide the 112 observers into 19 groups such that there are 6 observers in each of the first 18 groups and 4 observers in the last group; (2) for each of the 19 spatial frequencies, randomly select a group (without replacement) and obtain contrast sensitivities of H1 and H2 at that spatial frequency from
the 6 or 4 observers in that group; (3) run correlation analysis on the 112 pairs of CSF
(H1 and H2) constructed in step 2; and (4) repeat steps 1 to 3 one thousand times and
calculate the average correlation coefficient. By using this procedure, the sensitivities at
difference spatial frequencies were from completely different observers and were not
constrained by the truncated log-parabola model.

In Figures 9abcd, we show contrast sensitivities constructed in one instance of the
500 permutations. The CSFs measured in H1 and H2 after 5, 10, 20, and 50 quick CSF
trials are shown as scatter plots in Figures 9a, b, c and d, respectively. The average
correlation coefficient across 500 permutations is plotted as a function of trial number in
Figure 9e. The average correlation coefficient was 0.751, 0.836, 0.924 and 0.974 after 5,
10, 20 and 50 quick CSF trials, respectively. After 15 trials, the average correlation
reached 0.9. This result shows that the quick CSF is highly reliable across repeated tests.
Figure 9. (a), (b), (c), and (d) scatter plots of the CSFs measured in the H1 and H2 conditions after 5, 10, 20 and 50 trials. (e) Average Pearson correlation coefficient as a function of trial number.
The variability of the estimated CSF can also be derived from the correlation coefficient:

$$SE = SD_{sample} \sqrt{1 - r},$$

(3)

where $SE$ is the standard error of the estimated CSF, $SD_{sample}$ is the standard deviation of the pooled CSFs of all observers in the H1 and H2 conditions, and $r$ is the average correlation coefficient between the two repeated measures. The standard error is also plotted as a function of trial number in Figure 8. The standard error is virtually identical to the other two variability measures. Again, the result supports that the HWCI in a single quick CSF measurement could be an adequate estimate of the variability.

**Detecting CSF changes in individuals: Sensitivity and specificity**

The Bayesian parametric nature of the quick CSF method makes measurement very information-rich. It provides not only a (conventional) point estimate, but also the posterior distribution of the estimate of interest that can be exploited to test and detect CSF changes of an individual in different conditions. In this analysis, we focus on detection of AULCSF changes based on the posterior distribution of AULCSF measured in different luminance conditions.

First, the distribution of AULCSF difference was derived from the posterior distributions of the AULCSFs in the two to-be-compared conditions, i.e., M - L, and H1 - H2, of an individual observer (see Figure 4 bottom row):

$$p_{difference}(\Delta a) = \int_{a=\infty}^{\infty} p_1(a)p_2(a - \Delta a),$$

(4)
where $a$ and $\Delta a$ represent AULCSF and AULCSF difference, respectively, $p_{\text{difference}}()$ is the distribution of AULCSF difference, and $p_1()$ and $p_2()$ are the posterior distributions of AULCSF in the two conditions. The distributions of AULCSF difference for H1-H2, H-L, M-L and H-M after 50 trials of observer S14, S26, S86 and S107 were shown in Figure 10.

The distribution of AULCSF differences between two repeated measures in the same conditions, H1 - H2, obtained for each individual, was used as a reference. We assumed that the shape of the distribution of AULCSF difference doesn’t change with its mean. This assumption is reasonable because the HWCLs of the posterior distributions of AULCSF differences between different test conditions were almost the same (Figure 6). The 95% credible interval of the AUFCSSF difference distribution between H1 and H2 was set as the change criterion (Figure 10). Any AULCSF difference within the criterion...
was classified as no change, while any AULCSF difference outside the criterion was
classified as a change.

The concepts of sensitivity and specificity were used to evaluate the performance
of the quick CSF method\textsuperscript{68,69}. \textit{Sensitivity} is defined as the probability of reporting a
change when there is a real condition change as indicated by the red area under the
distribution of AULCSF difference between different test conditions. \textit{Specificity} is
defined as the probability of declaring no change when there is no change, as indicated by
the blue area under the distribution of AULCSF difference between repeated measures of
the same condition, i.e., $H_1 - H_2$. By definition, the specificity corresponding to the 95%
criterion credible interval is 95%.

![Credible interval as change criterion](image)

Figure 10. The blue curve represents the distribution of AULCSF difference between two
repeated measures in the same condition, i.e., $H_1 - H_2$. The red curve represents
distribution of AULCSF difference between two different conditions, e.g., $H - L$. The
criterion of change is based on the credible interval of AULCSF difference between
repeated measures. Specificity and sensitivity are indicated by the blue and red area, respectively.

The sensitivity for detecting an AULCSF change for observer S14, S26, S86 and S107 is plotted in Figures 11a, b, c and d, respectively. Figure 11e contains the plots of the average sensitivities of the 112 observers for the H - L, M - L and H – M changes.

Generally, the sensitivity increased with trial number. The average sensitivity for detecting an AULCSF change (0.43 log unit) between H and L conditions was 19.5%, 42.5%, 71.6% and 97.2% with 5, 10, 20 and 50 quick CSF trials. The average sensitivity for detecting an AULCSF change (0.29 log unit) between M and L conditions was 19.5%, 32.0%, 45.1% and 81.0% with 5, 10, 20 and 50 quick CSF trials. The average sensitivity for detecting an AULCSF change (0.11 log unit) between H and M conditions was 11.0%, 15.6%, 19.5% and 38.6% with 5, 10, 20 and 50 quick CSF trials.
Figure 11. Sensitivities of the quick CSF method in detecting an AULCSF change for observer S14 (a), S26 (b), S86 (c) and S107 (d) as functions of trial number. e) The average sensitivities across all 112 observers as functions of trial number. Different colors represent the sensitivities of detecting different changes.

Because the width of the criterion region could affect both sensitivity and specificity (Figure 10), we varied the change criterion across the credible interval range (0% to 100%) and computed the corresponding average sensitivity for the changes of H-L, M-L and H-M. The average sensitivity can be plotted against 1-specificity to yield the receiver operating characteristic (ROC). The area under the ROC curve provides a measure of the accuracy of correctly classifying change vs no change for the quick CSF method. Figure 12 shows the curves after 5, 10, 20 and 50 quick CSF trials.

With 50 quick CSF trials, the accuracy of the quick CSF method in detecting AULCSF changes in H-L and M-L was very close to 100% (Figure 12ab), while it was lower in detecting changes in H-M. In Figure 12, the accuracy of CSF change detection is plotted as functions of quick CSF trial number for difference changes. The accuracy of detecting an AULCSF change (0.43 log unit) between the H and L conditions was 65.6%, 79.3%, 91.4% and 98.9% with 5, 10, 20 and 50 quick CSF trials. The accuracy of detecting an AULCSF change (0.29 log unit) between the M and L conditions was 66.3%, 71.3%, 78.4% and 94.0% with 5, 10, 20 and 50 quick CSF trials. The average accuracy of detecting an AULCSF change (0.11 log unit) between the H and M conditions was 57.2%, 60.2%, 63.4% and 76.0% with 5, 10, 20 and 50 quick CSF trials.
Figure 12. The receiver operating characteristic (ROC) curves for detecting AULCSF change between the H and L conditions (a), between the M and L conditions (b) and between the H and M conditions (c). The curves in different colors represent results obtained with different number of quick CSF trials.

Figure 13. Accuracy of the quick CSF method in detecting different AULCSF change between the H and L conditions (red), between the M and L conditions (green) and between the H and M conditions (blue).
Detecting CSF changes between groups: Statistical Power

The large sample size used in the current study made it possible for us to evaluate the performance of the quick CSF method in detecting changes in the group mean between two conditions. We conducted an empirical power analysis by using the average effect size of all observers as the true effect size, rather than a hypothetical effect size in the conventional power analysis. This way, we were able to provide a realistic assessment of the detectability of group mean changes when the quick CSF method is deployed in the laboratory or clinic.

The empirical power analysis was performed with the following procedure. For a given number \( N (N = 2, 3, \ldots, 112) \), we randomly selected \( N \) observers from the total sample of 112 observers with replacement, and performed power analyses for paired \( t \)-test based on the observed standard deviation of the AULCSF of the subset of observers. The average effect sizes of AULCSF change (over 112 observers), 0.43, 0.29 and 0.14 for H - L, M - L and H - M respectively were considered as the true effect sizes. The statistical power for detecting AULCSF change for H - L, M - L and H - M, with \( \alpha = .05 \) was calculated. The procedure was repeated 1000 times and the average power was taken as the estimated power of the quick CSF method in detecting respective group mean changes. Figure 13 shows the estimated power as a function of the number of observers and the number of quick CSF trials as heat maps for different changes.
Figure 13. The power of the quick CSF method to detect group AULCSF changes is presented as functions of the observer and quick CSF trial numbers, for changes between (a) the H and L conditions, (b) between the M and L conditions, and (c) between the H and M conditions.

After 10 quick CSF trials, we needed only 8, 7, and 60 observers to detect an AULCSF change in the H - L, M - L and H - M comparisons with 95% power, respectively; after 20 trials, we needed only 5, 5 and 20 observer to detect an mean AULCSF change in all comparisons, respectively; after 50 trials, we needed only 4, 4 and 10 observer to detect a mean AULCSF change in all comparisons, respectively;

We can also assess statistical power in terms of the number of trials instead of the number of observers. For example, with 10 observers, we needed 7, 7, and 35 quick CSF trials to detect an AULCSF change in the H - L, M - L, and H - M comparisons with .95 power, respectively; with 20 observers, we needed only 5, 4, and 22 trials to detect respective AULCSF changes; with 112 observers, we needed only 1, 1, and 6 trials to detect respective AULCSF changes.
We also evaluated the effect size that is required for the quick CSF method to
detect a mean AULCSF change for given numbers of observers and trials. We assumed
that the distribution of AULCSF differences between the two conditions is Gaussian and
the shape or standard deviation of it doesn’t change with its mean. The former
assumption was validated by the Kolmogorov-Smirnov test. The histograms of AULCSF
difference after 50 quick CSF trials for H-L, M-L, H-M and H1-H2 were shown in
Figures 14abcd, respectively. One-sample Kolmogorov-Smirnov tests were applied to the
AULCSF differences of all observers. The results showed that the distributions of
AULCSF changes were normal for all comparisons (all \( p > 0.05 \)). We used the Brown-
Forsythe test for variance equality for the distribution of AULCSF differences in M-L,
H-M and H1-H2 comparisons (Figures 14acd). There was no difference found in
standard deviation of AULCSF differences in H1-H2, M-L and H-M (all \( p > .05 \) except
for trial 1; Figure 15a). It suggests that the standard deviation of AULCSF difference was
independent of the baseline level of AULCSF differences observed between conditions.
We calculated the average standard deviation across the three comparisons and plotted it
against trial number in Figure 15b. This value was used in the following analyses.

For any given number of trials, the effect size of change that the quick CSF
method can detect with \( \alpha = .05 \) and power = .95 was computed for given number of
observers, \( N (N = 2, 3, \ldots, 112) \) based on the estimated standard deviation in Figure 15b.
In Figure 15c, the effect size is plotted as a function of number of observers and trials.
Generally, with more observers and trials, the effect size that is required for the quick
CSF method to detect an AULCSF change shrinks. To detect 0.2, 0.1 and 0.05 log units
of AULCSF changes with 20 quick CSF trials, we would need 8, 25 and 93 observers,
respectively. To detect the same amounts of difference in 50 trials, only 4, 9 and 28
observers would be needed. With 20 observers, we would need 7, 23 and more than 50
trials to detect 0.2, 0.1 and 0.05 log units AULCSF changes, respectively. With 112
observers, we needed only 1, 6 and 15 trials to detect the same amount of difference,
respectively. In fact, we could detect a difference less than 0.025 log units (6% difference)
with 112 observers and 50 trials.

Figure 14. (a), (b), (c) and (d) Histograms of AULCSF difference in comparisons for H1-
H2, H-L, M - L, and H- M, respectively. These histograms show normality and equal
variance.
Figure 15. a) The $p$-value of the Brown-Forsythe test as a function of trial number. All $p$-values $>.05$ except for trial 1, signifying the same standard deviation of AULCSF difference in all AULCSF comparisons. b) The empirical standard deviation of population AULCSF difference as a function of trial number. c) The effect size that can be detected by the quick CSF method with $\alpha = .05$ and power $= .95$ as a joint function of observer and trial numbers.

Discussion

In this study, we collected CSFs from 112 college students under different mean luminance conditions and performed comprehensive analyses on the data to evaluate the performance of the quick CSF method. We found that the HWCI of the CSFs obtained with a single quick CSF measurement faithfully captured the measurement error, which typically need to be estimated by multiple measurement; and the HWCIs in different conditions were less than 0.1 log unit after 50 trials. This result demonstrates the impressive precision of the quick CSF method and also suggests that the precision of the
quick CSF won’t change with conditions, at least when luminance changes from 2.62 to 95.4 cd/m$^2$. The test-retest reliability was larger than 0.95 after 20 quick CSF trials, and reached 0.974 with 50 trials, which suggests the quick CSF delivers stable measurements. The high precision and reliability make it possible to apply the quick CSF method to detect CSF changes with high sensitivity and specificity. The method can be used to detect individual AULCSF changes caused by 7.79 and 35.7 folds luminance change with 94.0% and 98.9% accuracy in 50 trials. A power analysis showed that an effect size of 0.025 log units (6% difference) can be detected with the quick CSF method with 112 observers and 50 trials. The high accuracy and great power we demonstrated in this study manifest the quick CSF an efficacious tool for monitoring or detecting subtle CSF changes in both research and clinic.

Sloane, Owsley & Jackson measured contrast sensitivity as a function of luminance and showed that the log sensitivity vs log luminance function had a slope of 0.5 for spatial frequencies range from 2 to 4 cpd. In our study, the slope of the log sensitivity vs log luminance function was 0.19 at same range and the slope of the AULCSF vs log luminance function was 0.275 (Figure 4). We attribute the discrepancy in the slope for log sensitivity to different testing conditions. While Sloane et al used brief displays and stimuli with a fixed size across all spatial frequencies, we used stimuli that decreased in size as a function of spatial frequency and lasted until response. Sloane, et al. found that the slopes of the log CSF vs log luminance function were steeper for older adults than those for younger adults, which means that same amount of luminance change would induce greater CSF change in older than younger adults. In our case, if the quick CSF could detect the AULCSF change (M-L, 0.29 log unit) which induced by 7.8 fold
luminance change with the accuracy of 94% for our young observers, it could certainly
detect the AULCSF change induced by the same luminance change for older people with
an even higher accuracy.

The dark-adapted CSF is very informative in the diagnosis of AMD. Because
there is a substantial loss of parafoveal rod photoreceptors and small-scale, randomly
scattered lesions across the macula associated with drusen, the retina illumination is
attenuated compared to young people. The time course of dark adaptation of both cones
and rods in AMD were found to be delayed relative to young people. Dark
adaptation commonly lasts 30 to 40 minutes. In order to assess the CSF during the
course of dark adaptation, the testing procedure must be as short as possible. The quick
CSF method shows promise because it is rapid (10 trials, ~ 2 minutes) and precise (~ 0.2
log units).

Several quick CSF studies on patients with AMD, amblyopia and glaucoma
showed that similar or slightly greater number of trials (< 25%) was required to achieve
the same precision on clinical populations as normal subjects, and the test precision did
not depend on patient’s overall level of visual deficits. It suggests great external
validity of the quick CSF method. However, because only normal participants were
included in this study, the performance of the quick CSF method in clinical populations
should be examined in future studies.

Here, we conducted another statistical power analysis with the conservative
assumption that the standard deviation of AULCSF changes in patients is twice that of
the normal observers in the current study. The results are shown in Figure 15. To detect
0.2 and 0.1 log units of AULCSF changes with 20 quick CSF trials, we would need to run 25 and 93 observers, respectively. To detect the same amounts of difference in 50 trials, 9 and 28 observers are needed. With 20 observers, we would need 23 and more than 50 trials to detect 0.2 and 0.1 log units of difference, respectively. With 112 observers, we needed only 6 and 15 trials to detect the same amount of difference, respectively. The effect size changes with trial and observer numbers in a very similar way as that of the normal group.

Figure 15. The effect size that can be detected by the quick CSF method with $\alpha = .05$ and power = .95 as a joint function of both observer and trial numbers for a group having AULCSF changes with $2 \times$ standard deviation.

The quick CSF method has also been evolving in different forms. It was implemented on a tablet device, and the resulting CSF estimates had comparable precision to those obtained on laboratory equipment. A hierarchical adaptive design optimization (HADO) procedure was developed to achieve even greater efficiency by
exploiting the knowledge of population distributions of CSFs inferred from prior studies.

In summary, this assay calibration study demonstrates that the quick CSF method is very precise and sensitive to CSF changes in both individual and group levels. The method shows promise as a tool to monitor the progression of vision loss in eye disease or its remediation with treatment. It could also greatly reduce the sample size and costs in clinical trials.

Acknowledgement

Supported by the National Eye Institute (EY021553 to ZLL) and by the National Institute of Mental Health (MH093838 to JM and MP). LL & ZLL own intellectual property rights on quick CSF technology and have equity interests in Adaptive Sensory Technology, LLC. LL holds an executive position at AST.

Appendix A: The quick CSF algorithm

The quick CSF used the following steps to estimate $\tau(f)^{1}$:

1. **Define a CSF functional form.** $\tau(f)$ is the reciprocal of contrast sensitivity $S(f)$:

$$\tau(f) = \frac{1}{S(f, \theta)}$$

which is described by the truncated log parabola with four parameters $^{1}$:

$$\log_{10}(S(f, \theta)) = \begin{cases} 
\log_{10}(g_{\text{max}}) - \delta, f < f_{\text{max}} & \land & S_0 < \log_{10}(g_{\text{max}}) - \delta \\
\log_{10}(g_{\text{max}}) - \frac{4}{\log_{10}(2)} \left( \frac{\log_{10}(f) - \log_{10}(f_{\text{max}})}{\beta} \right)^2, f > f_{\text{max}} \end{cases} , \quad (A1)$$
where \( \theta = (g_{\text{max}}, f_{\text{max}}, \beta, \delta) \) represents the four CSF parameters: peak gain \( (g_{\text{max}}) \), peak spatial frequency \( f_{\text{max}} \), bandwidth at half-height \( (\beta, \text{in octaves}) \), and low-frequency truncation level \( \delta \). 

2. **Define the stimulus and parameter spaces.** The application of Bayesian adaptive inference requires two basic components: (1) a prior probability distribution, \( p(\theta) \), defined over a four-dimensional space of CSF parameters \( \theta \), and (2) a two-dimensional space of possible letter stimuli with contrast \( c \) and spatial frequency \( f \). In our simulation study, the ranges of possible CSF parameters were: 2 to 2000 for peak gain, 0.2 to 20 cycles per degree (cpd) for peak frequency, 1 to 9 octaves for bandwidth and 0.02 to 2 for truncation. The ranges for possible grating stimuli were 0.2% to 100% for contrast \( c \) and 1.19 to 30.95 cpd for frequency \( f \). Both parameter and stimuli spaces were sampled evenly in log unit.

3. **Priors.** Before the beginning of the experiment, an initial prior, \( p_{t=0}(\theta) \), which represents the knowledge about the observer’s CSF before any data is collected, was defined by a hyperbolic secant function with the best guess of parameters \( \theta_{i,\text{guess}} \) and width of \( \theta_{i,\text{confidence}} \) for \( i = 1, 2, 3 \) and 4 \(^{1,76}\).

\[
p_{t=0}(\theta) = \prod_{i=1}^{4} \text{sech}(\theta_{i,\text{confidence}} \times (\log_{10}(\theta_i) - \log_{10}(\theta_{i,\text{guess}}))), \quad (A2)
\]

where \( \text{sech}(x) = \frac{2}{e^x + e^{-x}} \), \( \theta_i = g_{\text{max}}, f_{\text{max}}, \beta \) and \( \delta \) for \( i = 1, 2, 3 \) and 4, respectively,
\[
\theta_{i,\text{guess}} = 100, 2, 3 \text{ and } 0.5 \text{ for } i = 1, 2, 3 \text{ and } 4, \text{ respectively. } \theta_{i,\text{confidence}} = 2.48, 3.75, 7.8 \text{ and } 3.12 \text{ for } i = 1, 2, 3 \text{ and } 4, \text{ respectively.}
\]
4. **Bayesian adaptive inference.** After observer’s (three) responses is collected in trial $t$, knowledge about CSF parameters $p(\theta)$ is updated, given the evidence provided by the observer’s response $r_x$ = “correct” or “incorrect” to the stimulus $x_i = (c_i, f)$ with contrast $c_i$ and spatial frequency $f$ in the trial, where $i = 1, 2$ and $3$ represents the $i$th letter from left to right. The outcome of trial $t$ is incorporated into a Bayesian inference step that updates the knowledge about CSF parameters $p_{t-1}(\theta)$ prior to trial $t$,

$$p_t(\theta) = p_t(\theta | r_{x_1}, r_{x_2}, r_{x_3}) = \frac{p_{t-1}(\theta) \prod_{i=1}^{3} p(r_{x_i} | \theta)}{\sum_{\theta} \left[ p_{t-1}(\theta) \prod_{i=1}^{3} p(r_{x_i} | \theta) \right]}, (A3)$$

where $p_t(\theta | r_{x_1}, r_{x_2}, r_{x_3})$ is the posterior distribution of parameter vector $\theta$, after obtaining a response $r_x$ at trial $t$; $p(r_{x_i} = \text{correct} | \theta) = \Psi(x_i, \theta)$ is the percent correct psychometric function given stimulus $x_i$, and $p(r_{x_i} = \text{incorrect} | \theta) = 1 - \Psi(x_i, \theta)$; $p_{t-1}(\theta)$ is our prior about $\theta$ before trial $t$, which is also the posterior in trial $t - 1$.

5. **Stimulus search.** To increase the quality of the evidence obtained on each trial, the quick CSF calculates the expected information gain for all possible stimuli $x$,

$$I_t(\theta; r_x) = h\left( \int p_t(\theta) \Psi(x, \theta) d\theta \right) - \int p_t(\theta) h(\Psi(x, \theta)) d\theta, (A4)$$

where $h(p) = -p \log(p) - (1 - p) \log(1 - p)$ is the information entropy of the distribution $p$. Before each trial, we find out the candidate stimuli that correspond to the top 10% of the expected information gain over the entire stimulus space. Then we randomly pick one among those candidates as $x_t = (c, f)$ for presentation. In this way, the quick CSF avoids large regions of the stimulus space that are not likely to provide
useful information to the current knowledge about $\theta$. To improve observers’
experience in CSF testing, two additional letters with higher contrasts are presented
alongside with the optical test letter $x_t$. From left to right, their contrasts are $4c$, $2c$ and$c$, respectively. The maximum contrast is capped at 90%. Their spatial frequency is $f$.

6. **Reiteration and stopping rule.** The quick CSF procedure re-iterates steps (4) and (5)
until 50 trials are run.

7. **Analysis.** After step 6, we obtain the posterior distribution in numerical form of CSF
parameters $p_t(\theta)$ (see Figure A1 for the marginal prior and posterior distributions for
the four CSF parameters). A re-sampling procedure is used that samples directly from
the posterior distributions of the CSF parameters and generates the CSF estimates, i.e.
CSF and AULCSF based on all the CSF samples. The procedure automatically takes
into account the covariance structure of the CSF parameters in the posterior
distribution and allows us to compute the credible interval of the estimates derived
from CSF functions.
Figure A1. An illustration of the marginal distributions of four parameters before (prior: red) and after (posterior: blue) measurement. The plot was based on the simulation of a single quick CSF run with 100 trials.

Reference

19. Arden GB, Jacobson JJ. A simple grating test for contrast sensitivity: preliminary results 
20. Barnes RM, Gee L, Taylor S, Briggs MC, Harding SP. Outcomes in verteporfin 
photodynamic therapy for choroidal neovascularisation--‘beyond the TAP study’. *Eye (Lond)* 
21. van Gaalen KW, Jansonius NM, Koopmans SA, Terwee T, Kooijman AC. Relationship 
between contrast sensitivity and spherical aberration: comparison of 7 contrast sensitivity tests 
22. Pesudovs K, Hazel CA, Doran RM, Elliott DB. The usefulness of Vistech and FACT contrast 
sensitivity charts for cataract and refractive surgery outcomes research. *Br J Ophthalmol* 
2004;88:11-16.
26. Kelly DH, Savoie RE. A study of sine-wave contrast sensitivity by two psychophysical 
27. Harvey LO, Jr. Efficient estimation of sensory thresholds with ML-PEST. *Spat Vis* 
1997;11:121-128.
28. Hou F, Lesmes L, Bex P, Dorr M, Lu Z-L. Using 10AFC to further improve the efficiency of 
quick CSF. *Journal of Vision* 2015;In press.
second-order sensitivity using the quick Contrast Sensitivity Function (qCSF). *Journal of Vision* 
2014;14:1428.
31. Hou F, Huang C-B, Lesmes L, et al. qCSF in clinical application: efficient characterization 
and classification of contrast sensitivity functions in amblyopia. *Invest Ophthalmol Vis Sci* 
2010;51:5365-5377.
32. Lesmes L, Wallis J, Lu Z-L, Jackson ML, Bex PJ. Clinical Application Of A Novel Contrast 
Sensitivity Test To A Low Vision Population: The Quick CSF Method. *ARVO Meeting Abstracts* 
2012;53:4358.
33. Lesmes L, Wallis J, Jackson ML, Bex P. The reliability of the quick CSF method for 
contrast sensitivity assessment in low vision. *Investigative Ophthalmology and Visual Science* 
2013;54:2762.
34. Ramulu P, Dave P, Friedman D. Precision of contrast sensitivity testing in glaucoma. 
*ARVO Meeting Abstracts* 2015;2225-B0078.
Characterization and Classification of Contrast Sensitivity Functions in Aging. *ARVO Meeting 
Abstracts* 2014;762.
36. Gepshtein S, Lesmes LA, Albright TD. Sensory adaptation as optimal resource allocation. 


47. Tyler CW. Colour bit-stealing to enhance the luminance resolution of digital displays on a single pixel basis. *Spat Vis* 1997;10:369-377.


