Evaluation of tablet-based methods for assessment of contrast sensitivity

Jeffrey B. Mulligan; NASA Ames Research Center; Moffett Field, CA / USA

Abstract

Some astronauts have suffered degradation of vision during long-duration space flight, suffering from a condition that has come to be known as Spaceflight Associated Neuro-ocular Syndrome (SANS). While related morphological changes can be observed with imaging technologies such as optical coherence tomography (OCT), it may be useful to have a rapid method for functional vision assessment. In this paper, we compare three tablet-based methods for rapid assessment of contrast sensitivity. First, a relatively novel method developed expressly for touch screens, in which the subject "swipes" a frequency/contrast sweep grating to indicate the boundary between visible and invisible patterns; second, a method-of-adjustment task in which the subject adjusts the contrast of a grating patch up and down to bracket the visual threshold; and third, a traditional temporal two-alternative forced choice (2AFC) task, in which the subject is presented with a near-threshold stimulus in one of two intervals, and must report the interval containing the stimulus. The swipe method shows variability comparable to the 2AFC method, and shows good agreement in estimates of the spatial frequency of peak sensitivity. The absolute sensitivity estimated with the swipe method is higher than that of the other methods, perhaps because subjects are biased to trace outside of the visible pattern region, or perhaps due to stimulus differences.

Introduction

It has come to light in recent years that some astronauts suffer structural changes in the eye and degradation of vision following prolonged exposure to micro-gravity. This condition has come to be known *Spaceflight Associated Neuro-ocular Syndrome* (SANS) [1, 2, 3, 4, 5, 6]. It is generally thought that in the absence of terrestrial gravity, mechanisms that normally prevent blood and other fluids from accumulating in the feet send an excess of fluids to the head, elevating intracranial pressure. This in turn presses on the back of the eye, shortening the eye length inducing hyperopia (far-sightedness), and introducing other problems.

NASA's Human Research Program has determined that the risk of SANS is acceptable for near-earth missions up to one year in duration (including lunar habitation) , but that mitigation will be required for longer missions to other planets [7]. While the International Space Station (ISS) is now equipped with an instrument to measure the eye using Optical Coherence Tomography (OCT), it is used infrequently, and little is known about the fine-scale time course of the progression of SANS. It therefore seemed useful to develop a tool to perform a rapid functional assessment of vision, which could be used by crew members on a daily basis without substantially interfering with their other duties.

Various methods have been developed over the years to assess human visual function. Most familiar is perhaps the Snellen letter eye chart, first introduced over 150 years ago. An improved version in use today is the Bailey-Lovie LogMAR chart ([8, 9]), which uses the Sloan font, and presents the same number of letters on each line. These charts, as well as others, measure visual acuity by determining the smallest high-contrast object that can be reliably identified.

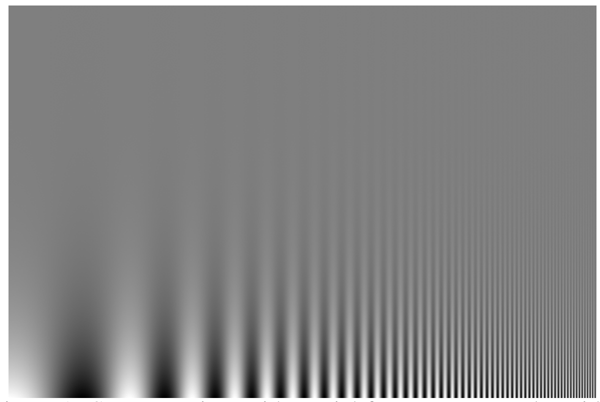

Figure 1: Sweep grating, with spatial frequency swept logarithmically in the horizontal dimension, and contrast swept logarithmically in the vertical dimension.

A somewhat more informative assessment of vision can be obtained by measuring *contrast sensitivity*. This refers to the lowest contrast at which a pattern of a given size can be seen. Large things typically can be seen at relatively low contrasts, while small things require high contrast. The *Contrast Sensitivity Function* (CSF) plots the threshold sensitivity as a function of spatial frequency. Figure 1 displays an image of a sinusoidal luminance grating, with the spatial frequency increasing from left-to-right and the contrast increasing from top-to-bottom. An image like this was first created for use in optical testing [10], but was introduced to vision research by the earlier pioneers of the CSF, John Robson and Fergus Campbell, who presented it at a meeting of the Optical Society in 1964 [11]. This image allows an observer to "see" their own CSF, as the inverted U-shaped boundary of the region containing visible stripes.

The fact that the sweep grating affords instant visualization of contrast sensitivity was the motivation to develop a touch screen based method for rapid measurement [12]. In this paper, that method is compared with two more traditional psychophysical methods.

Methods

All measurements were made using an iPad Pro 12.9" tablet computer (Apple Computer, model MP6G2LL/A), running a custom application developed using the QuIP interpreter [13]. Accurate rendition of low contrast stimuli was accomplished by half-toning the quantization errors, and applying the halftone pattern the the least-significant bit of the image data, using an iterative half-toning algorithm [14]. The process incorporated gamma correction, with calibration obtained using a psychophysical matching procedure [15], although the device used for these experiments had a relatively linear response. An earlier paper describes the methods in greater detail [12]. All images were computed at a resolution of 768x1024 (the native resolution of older iPad devices, but half the full device resolution of the iPad Pro).

The experimental protocol was approved by the NASA Human Institutional Review Board. Data were collected from nine volunteer subjects, all of whom were briefed on the purpose of the experiment, and provided written informed consent.

The subjects were seated at a desk in the lab, with the iPad supported by a stand. They were instructed to adopt a comfortable seating posture that allowed easy interaction with the touch screen, and the viewing distance was measured with a tape measure. Viewing distances were fairly consistent, ranging from 18 to 20 inches. The viewing distance can be used to convert the frequencies (here shown in cycles per sample) to visual frequencies expressed in cycles per degree; for a viewing distance of 18.5 inches, the numbers work out particularly nicely, with the highest frequency of 0.25 cycles/sample (period = four pixels) corresponding to eight cycles per degree, and the middle frequency of 0.03125 cycles/sample (period = 32 pixels) corresponding to one cycle per degree. Note that the hyperopic shifts associated with SANS are generally in excess of 0.75 diopters [6]; a blue of one diopter results in a blur circle diameter of approximately 0.2 degrees [16], producing the first zero in the optical modulation transfer function at around five cycles per degree. Thus the test's highest frequency of eight cycles per degree is sufficiently high to detect a one diopter blur, but only just. Thus using the full resolution of the iPad Pro (or using a device with even higher resolution) would be important for operational use.

The swipe method

The implementation of the swipe method was generally as previously described [12], with a few exceptions noted here. Each subject swiped a total of 15 different sweep images, which were varied in two ways: three different frequency ranges were presented, and five different "skews" controlled the relation between contrast and screen position. After completing a swipe, the subject was then presented with a visualization of their action as a series of small blue dots superimposed on the sweep grating, along with three buttons labeled "accept," "redo," and "abort."

Figure 2 shows the results for a typical subject (in black). Each swipe has been fit with a parabola (in blue), and the entire data set has been fit with a single parabola, shown with the heavy red curve.

The method of adjustment

The *method of adjustment* allows a subject to manually control the strength of a stimulus to quickly find a threshold or match. In the present case, the subject swiped up or down to increase or decrease the contrast of a Gabor patch. Seven frequencies were measured, ranging in octave steps from 0.25 cycles/sample (period = four pixels) to 0.0039 cycles/sample (period = 256 pixels). Each grating patch was windowed by a two-dimensional Gaussian contrast window having a standard deviation of 0.781 times the grating period. The smaller sizes were flanked by light cross-hairs, to reduce positional uncertainty. The image with the middle frequency (period = 32 pixels) is shown in Figure 3. This is the largest size for which the cross-hairs were included. Each of the seven spatial frequencies was tested in a separate block, in fixed order from highest to lowest frequency.

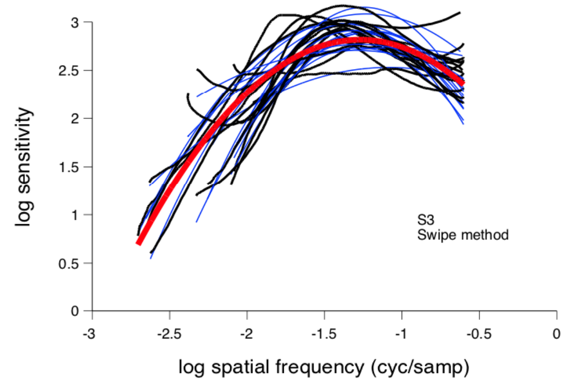

Figure 2: The raw data for 15 individual swipes are shown in black; each swipe is individually fit with a parabola (shown in blue). The heavy red curve shows the fit to all of the data combined together.

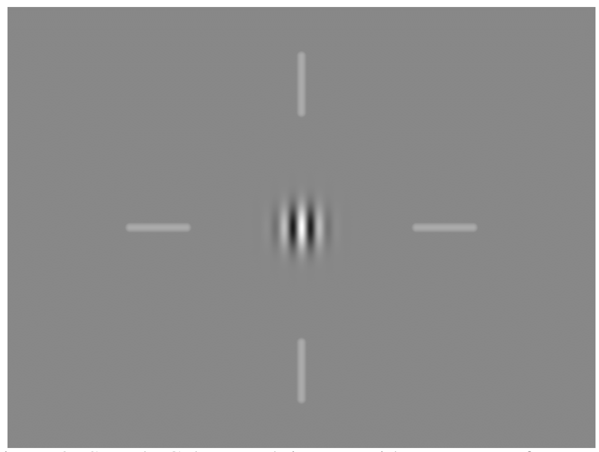

Figure 3: Sample Gabor patch image, with a contrast of one and a carrier period of 32 pixels.

Subjects were initially presented with the full contrast image, and reduced the contrast by swiping downwards on the touch-screen. Swiping up increased the contrast. The size of the contrast changes was controlled by an adaptive procedure designed to minimize the time required to reach the threshold. In a pilot experiment, subjects were allowed to terminate the procedure when they felt they had homed in on the threshold, but eventually it was determined that a fixed number of adjustments would provide more reliable data, and that number was fixed at 50. Effectively, the subjects performed 50 yes/no trials, with the stimulus continuously visible. A typical series of contrasts is presented in Figure 4.

The series of adjustments was transformed to a psychometric function by treating each adjustment as a yes/no trial. For each contrast level that was presented, the number of times that the subject swiped down (to decrease contrast) was divided by the

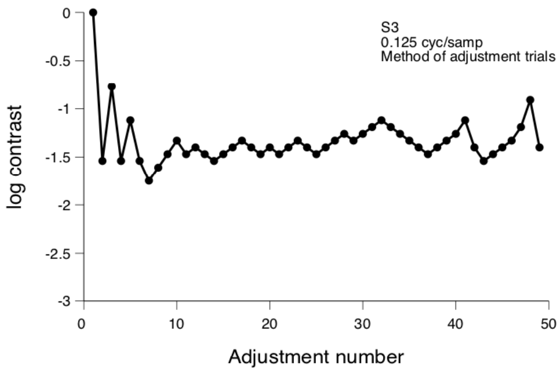

Figure 4: Series of contrasts shown to subject S3 performing the method-of-adjustment task.

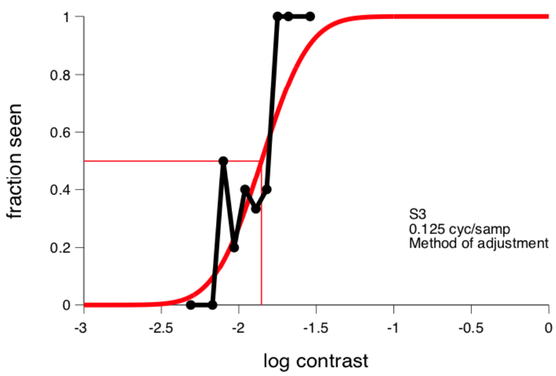

Figure 5: Psychometric function derived from trial sequence show in Figure 4. Symbols connected by black line segments show the raw data, while the red curve shows the fit of a Gaussian ogive. Threshold is estimated as the contrast for which the fit curve predicts 50% seen.

total number of presentations to obtain the fraction seen. Figure 5 shows these values plotted (in black) for the set of adjustments shown in Figure 4. These data were fit with a Gaussian ogive, with the parameters chosen to maximize the likelihood of obtaining the observed values. The fit is shown in red in Figure 5. The contrast for which the fit obtained a value of 0.5 was taken as the threshold.

The continuous visibility of the stimuli proved problematic in some respects: adaptation to the high contrast gratings produced negative afterimages when the contrast was quickly reduced. For some subjects, this caused the early settings (when the contrast increment was large) to be unstable. For this reason, the first twenty adjustments were not included in the construction of the psychometric function.

The seven threshold estimates obtained from subject S3 (whose data is shown in Figure 5) are plotted (in black) in Figure 6, along with a parabolic fit (in red).

Two-alternative forced choice (2AFC)

After completing the swipe and method-of-adjustment tasks, subjects performed a temporal two-alternative forced choice (2AFC) task. This task consisted of a number of trials, each consisting of two intervals marked by a 220 Hz tone. The amplitude of the tone was modulated by one cycle of a raised cosine envelope having a total duration of 500 milliseconds. On each trial, the subject heard two presentations of the tone, separated by an interstimulus interval of 500 milliseconds. The visual stimulus was assigned to one of the two intervals using a pseudo-random number generator. The stimuli were the same Gabor patches used in the method-of-adjustment task, but instead of being continuously visible, they were pulsed using a temporal contrast envelope consisting of one cycle of a raised cosine, with a total duration of 500 milliseconds, as was used for the audio tone. After the two intervals, a response screen was displayed with two large buttons used to indicate the visual stimulus was seen in the first or second interval, and three smaller buttons to redo the previous trial, abort the run, or to indicate an incorrect response (finger error) on the previous trial. Subjects were instructed that the redo button was only to be used in the event that they had not been looking at the stimulus, but should not be used to get a second look at a difficult-to-see stimulus.

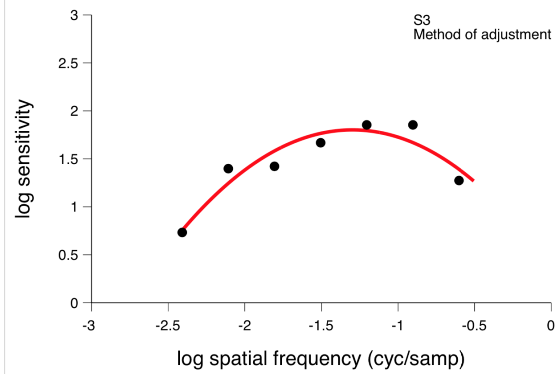

Figure 6: Black symbols connected by line segments show thresholds obtained using the method of adjustment, while the red curve shows the best-fitting parabola.

The sequence of contrasts presented was controlled by a staircase procedure, using a two-to-one rule that decreased the contrast after two successive correct responses, and increased the contrast after a single incorrect response. The starting contrast and initial step size were both initialized to half of the range, with the step size being halved at each downward reversal. The number of trials per block was fixed at 32. A typical series of contrasts is shown in Figure 7.

For each contrast that was presented by the staircase, the fraction correct was tabulated, and a psychometric function was fit to the data, much as was done for the method-of-adjustment data, except that in this case the function was constrained to have a lower asymptote of 0.5 (the chance proportion correct). Figure 8 shows data (in black) for a single frequency, and the best fitting Gaussian ogive (in red). The threshold was estimated at the contrast for which the estimated psychometric function obtained a value of 0.75.

Each subject ran a block of 32 trials at each frequency, from highest to lowest. Thresholds were estimated independently for each frequency. Figure 9 shows the thresholds so obtained (in black), along with a parabolic fit (in red).

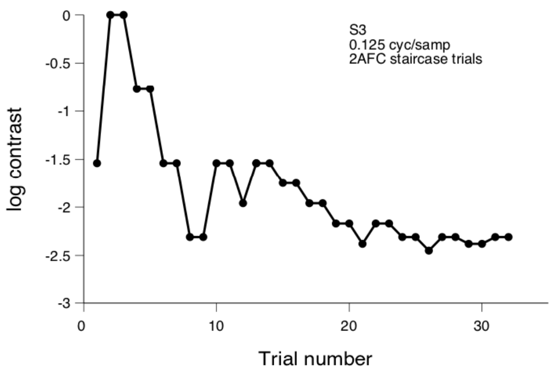

Figure 7: Sequence of trials presented to subject S3 running the forced-choice procedure for a single frequency. The contrast is decreased following two consecutive correct responses, and increased after a single incorrect response.

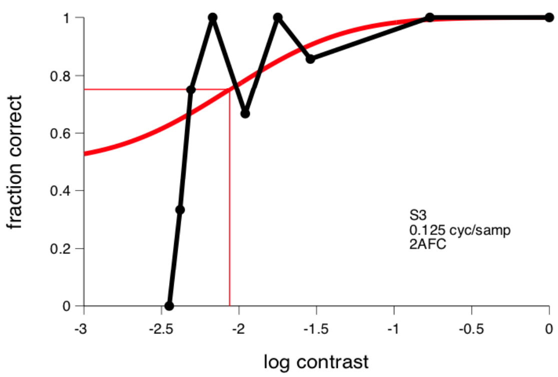

Figure 8: Psychometric function derived from the data shown in Figure 7. Black symbols connected by line segments show the raw data, while the red curve shows the fit of a Gaussian ogive constrained to have an asymptote of 0.5 for low contrasts.

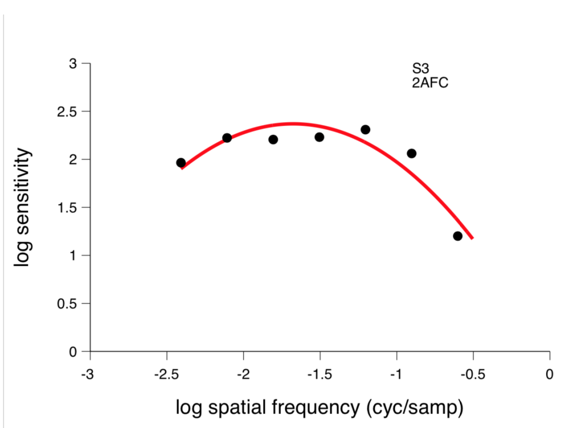

Figure 9: Black symbols connected by line segments show threshold derived as in Figure 8, while the red curve shows the best-fitting parabola.

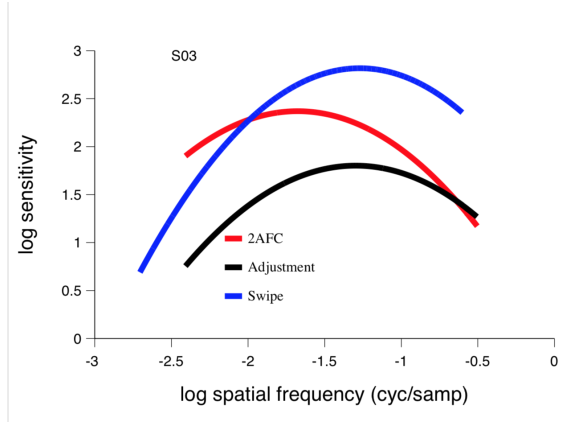

Figure 10: Estimates of the contrast sensitivity function obtained by three methods shown for subject S3.

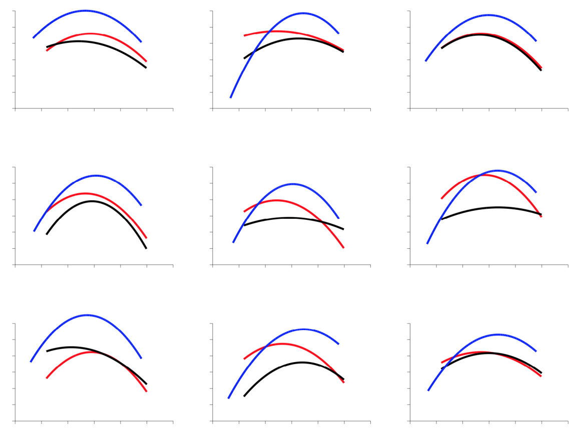

Figure 11: Similar to Figure 10, with all nine subjects shown.

Results

In this section we compare the results obtained from the three methods with one another. Figure 10 shows the CSF estimates obtained from the swipe method (in blue), the method-of-adjustment (in black), and 2AFC (in red). The curves are the same as those shown above in Figures 2, 6 and 9. For this subject, there is not particularly good agreement between any of the methods, with differences in estimated sensitivities exceeding one log unit. Figure 11 shows the data for all nine subjects, plotted just as in Figure 10. It can be seen that a number of the subjects show reasonable agreement between the method-of-adjustment (in black) and 2AFC (in red).

The parabolic fits can be uniquely characterized by three parameters. Two of these can be chosen to be the location of the apex of the parabola: the vertical position being the peak sensitivity, and the horizontal position being the spatial frequency of greatest sensitivity. Figure 12 shows a scatter plot of the frequency of peak sensitivity estimated from the method-of-adjustment versus that estimated from the swipe method. It can be seen that there is a strong correlation, but that the swipe method appears to be biased toward higher frequencies.

Figure 13 plots the frequency of peak sensitivity estimated from the 2AFC method versus that estimated from the swipe method. Although we see the same bias towards higher frequen-

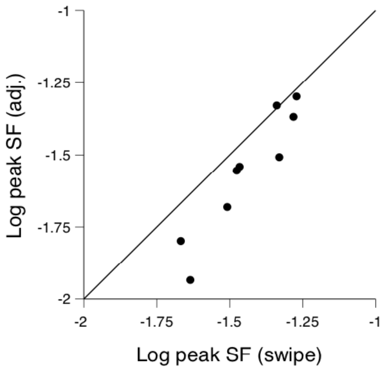

Figure 12: Scatter plot showing the log of the spatial frequency with maximal sensitivity estimated with the method-of-adjustment plotted against those obtained with the swipe method.

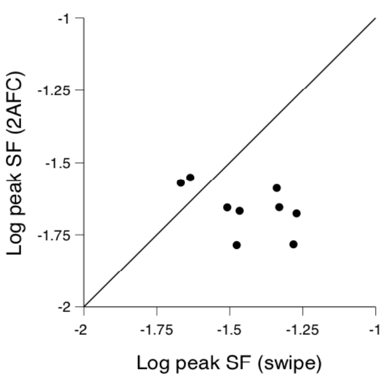

Figure 13: Similar to Figure 12, but here the spatial frequency with peak sensitivity obtained with the forced-choice method is plotted against that obtained with the swipe method.

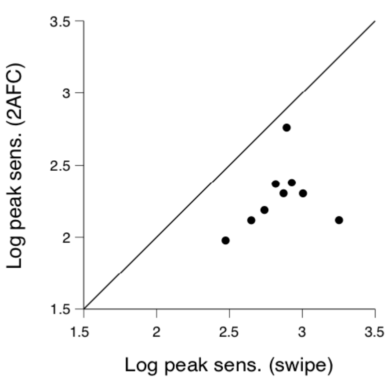

Figure 14: Scatter plot showing estimates of the peak sensitivity obtained with the forced-choice method plotted against estimates from the swipe method.

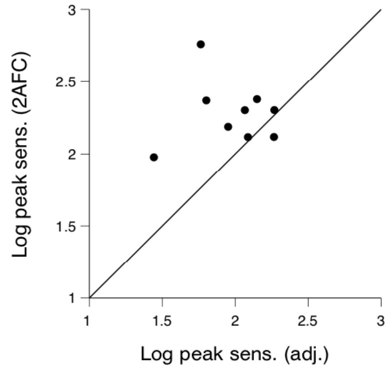

Figure 15: Similar to Figure 14, but here the peak sensitivity obtained with the forced-choice method is plotted against that obtained with the method-of-adjustment.

cies for the swipe method as seen in Figure 12 for seven of the nine subjects, here the data appear to be uncorrelated.

Figure 14 compares absolute peak sensitivity estimated from the 2AFC method plotted against that estimated by the swipe method. For seven of the nine subjects, the value obtained with the swipe method is approximately 0.5 log units greater than that obtained with the 2AFC method. Figure 15 plots the same peak sensitivities from the 2AFC method against those obtained from the method-of-adjustment. Here we see that the forced-choice method estimates a peak sensitivity that is greater than or equal to that obtained from the method-of-adjustment for eight of the nine subjects. However, there is little or no correlation between the two sets of estimates.

Discussion

The lack of strong correlations between any two of the methods makes it somewhat difficult to rank them in terms of reliability. Several possibilities exist that could explain this result: first, because the different methods employ different stimulus presentations, it is possible that they are measuring independent capabilities that vary between subjects due to individual differences. For example, in the 2AFC method, the stimulus is pulsed, while in the adjustment and swipe methods the stimulus is steady. It is well-known that as temporal frequency is increased, the spatio-temporal CSF changes from band-pass to low-pass, with temporal modulation improving the visibility of low frequencies [17].

Alternatively, it is possible that the data are simply noisy and need more observations in order to arrive at a more precise estimate. Additional work is needed to develop statistical methods for assessing error bounds on the CSF estimates.

The approach described above of first estimating a threshold for each spatial frequency, and then fitting a CSF to the thresholds may not be optimal. A better approach might be to fit all of the psychometric functions for a single subject simultaneously, assuming a common slope for all psychometric functions and thresholds determined by the CSF parameters. This would let those frequencies for which the psychometric function has a relatively poor fit to have a smaller influence on the overall CSF estimate.

The swipe method estimates sensitivities that are consistently higher that those estimated with the other methods. One explanation could be that subjects are biased to swipe slightly outside of the visible region. That could be tested by zooming the contrast dimension. It is also likely that stimulus differences play a role; the Gabor patches presented approximately three cycles of each grating, while the sweep grating presented many more (although not all at the same frequency). It is known that sensitivity improves with additional cycles up to about ten cycles [18].

The swipe method was developed with the goal of reducing the amount of time needed to estimate the CSF. After adding the confirmation screen (with the possibility of a redo), and 15 variants, the time used by subjects to complete the task was 4-5 minutes, much more than the couple of seconds required by an experienced observer to perform a single swipe. The method-of-adjustment task similarly required 4-5 minutes for most observers, while the 2AFC method took 10-12 minutes. It should be noted that other more complicated algorithms for trial placement have been developed that may allow greater efficiency for 2AFC that what has been presented here [19, 20, 21].

Summary

Results have been presented for three methods for estimating the contrast sensitivity function (CSF). The swipe method provides a crude estimate in a matter of seconds, and while variability in the sensitivity estimate provided by a single swipe is on the order of 0.5 log units (see Figure 2), this may still be adequate for longitudinal monitoring to detect the onset of Spaceflight-associated neuro-ocular syndrome (SANS).

Acknowledgments

This work was supported by the NASA Engineering Safety Center (NESC). Thanks to Giovanna Guevara-Flores for assistance with preliminary data collection, and Kenji Kato for assistance with app development. Thanks to Tina Beard, Scott Daly, and Dov Adelstein for comments on the manuscript.

References

[1] G. Taibbi, R. L. Cromwell, K. G. Kapoor, B. F. Godley, and G. Vizzeri, "The effect of microgravity on ocular structures and visual function: a review," *Surv. Ophthalmol.* **58**(2), pp. 155–163, 2013.

[2] T. H. Mader, C. R. Gibson, N. R. Miller, P. S. Subramanian, N. B. Patel, and L. A. G., "An overview of spaceflight-associated neuro-ocular syndrome (SANS)," *Neurology India* **67**(8), pp. 206–211, 2019.

[3] C. Aleci, "From international ophthalmology to space ophthalmology: the threats to vision on the way to moon and mars colonization," *Int. Ophthalmol.* , 2019.

[4] L. A. Galdamez, T. J. Brunstetter, A. G. Lee, and W. J. Tarver, "Origins of cerebral edema: implications for spaceflight-associated neuro-ocular syndrome," *J. Neuroophthalmol.* **40**(1), pp. 84–91, 2019.

[5] P. Wojcik, A. Kini, B. Al Othman, L. A. Galdamez, and A. G. Lee, "Spaceflight associated neuro-ocular syndrome," *Curr. Opin. Neurol.* **33**(1), pp. 62–67, 2020.

[6] A. G. Lee, T. H. Mader, C. R. Gibson, W. Tarver, P. Rabiei, R. F. Riascos, L. A. Galdamez, and T. Braunstetter, "Space flight associated neuro-ocular syndrome (SANS) and the neuro-ophthalmologic effects of microgravity: a review and an update," *npj Microgravity* **6**, p. 7, 2020.

[7] National Aeronautics and Space Agency. `https://humanresearchroadmap.nasa.gov/risks/risk.aspx?i=105`. Accessed 2019-12-19.

[8] I. L. Bailey and J. E. Lovie, "New design principles for visual acuity letter charts," *Am. J. Optom. Physiol. Opt.* **53**(11), pp. 740–745, 1976.

[9] I. L. Bailey and J. E. Lovie-Kitchin, "Visual acuity testing. from the laboratory to the clinic," *Vision Research* **90**, pp. 2–9, 2013.

[10] F. E. Washer and F. W. Roseberry, "New resolving power test chart," *J. Opt. Soc. Am.* **41**(9), pp. 597–600, 1951.

[11] J. G. Robson. personal communcation, 2015.

[12] J. B. Mulligan, "A method for rapid measurement of contrast sensitivity on mobile touch-screens," in *Proc. Human Vision and Electronic Imaging*, 2016.

[13] J. B. Mulligan. `https://github.com/nasa/QuIP`, 2016. Accessed 2019-12-19.

[14] J. B. Mulligan and A. J. Ahumada, Jr., "Principled halftoning based on human vision models," in *Proc. SPIE 1666, Human Vision, Visual Processing, and Digital Display III*, 1992.

[15] J. B. Mulligan, "Presentation of calibrated images over the web," in *Proc. SPIE 7240, Human Vision and Electronic Imaging XIV*, 2009.

[16] H. Strasburger, M. Bach, and S. P. Heinrich, "Blur unblurred - a mini tutorial," *i-Perception* **9**(2), 2018.

[17] D. H. Kelly, "Spatiotemporal variation of chromatic and achromatic contrast thresholds," *J. Opt. Soc. Am.* **73**, pp. 742–750, Jun 1983.

[18] E. Peli, L. E. Arend, G. M. Young, and R. B. Goldstein, "Constrast sensitivity to patch stimuli: effects of spatial bandwidth and temporal presentation," *Spat. Vis.* **7**(1), pp. 1–14, 1993.

[19] L. A. Lesmes, Z. L. Lu, J. Baek, and T. D. Albright, "Bayesian adaptive estimation of the constrast sensitivity function: The quick csf method," *J. Vis.* **10**(3), pp. 17.1–17.21, 2010.

[20] L. A. Lesmes and Z. L. Lu, "Methods and devices for rapid measurement of visual sensitivity." U.S. Patent 7938538, 2011.

[21] A. B. Watson, "Quest+: A general multidimensional Bayesian adaptive psychometric method," *J. Vis.* **17**(3), pp. 10–10, 2017.

Author Biography

Jeffrey B. Mulligan received the A.B. degree in physics from the Harvard University (1980), and the M.A. and Ph.D. degrees in psychology from the University of California at San Diego (1982, 1986). Since then he has worked as a computer engineer at the NASA Ames Research Center, in Moffett Field, CA. His work has focused on visual perception and eye movements, and applications to the design of aerospace information displays. He is a member of the Vision Sciences Society, the Optical Society, the Society for Information Display, and the Association for Computing Machinery.