



AFRL-SA-WP-TR-2018-0019

Operational Based Vision Assessment: Automated Vision Test Research and Reliability



Marc Winterbottom, James Gaska

*U.S. Air Force School of Aerospace Medicine, Aeromedical Research
Department*

Elizabeth Shoda, Eleanor O'Keefe, and Alex Van Atta

KBRwyle Laboratories, Beavercreek, OH



October 2018

**Final Report
for January 2015 to December 2017**

**DISTRIBUTION STATEMENT A. Approved
for public release. Distribution is unlimited.**

**Air Force Research Laboratory
711th Human Performance Wing
U.S. Air Force School of Aerospace Medicine
Aeromedical Research Department
2510 Fifth St., Bldg. 840
Wright-Patterson AFB, OH 45433-7913**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-SA-WP-TR-2018-0019 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNATURE//

DR. JAMES McEACHEN
CRCL, Human Performance

//SIGNATURE//

DR. RICHARD A. HERSACK
Chair, Aeromedical Research Department

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 4 Oct 2018		2. REPORT TYPE Final Technical Report		3. DATES COVERED (From – To) January 2015 – December 2017	
4. TITLE AND SUBTITLE Operational Based Vision Assessment: Automated Vision Test Research and Validation				5a. CONTRACT NUMBER FA8650-12-D-6280	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) James Gaska, Marc Winterbottom, Elizabeth Shoda, Eleanor O’Keefe				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) USAF School of Aerospace Medicine KBRwyle Aeromedical Research Dept/FHOH Beaver creek, OH 2510 Fifth St., Bldg. 840 Wright-Patterson AFB, OH 45433-7913				8. PERFORMING ORGANIZATION REPORT NUMBER AFRL-SA-WP-TR-2018-0019	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSORING/MONITOR’S ACRONYM(S)	
				11. SPONSOR/MONITOR’S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.					
13. SUPPLEMENTARY NOTES Cleared, 88PA, Case # 2018-5618, 7 Nov 2018. Report contains color.					
14. ABSTRACT The work detailed in this report was conducted by the Operational Based Vision Assessment Laboratory, Aeromedical Research Department, Human Performance Branch, U.S. Air Force School of Aerospace Medicine, Wright-Patterson AFB, OH, with support from KBRwyle. The report describes research conducted to develop normative data for several new automated vision tests and to examine the test-retest reliability of each of the tests. The automated vision tests described in this report include a color vision test, near and far stereo acuity tests, luminance contrast and acuity tests, a fusion range test, and a motion sensitivity test. Research and development of these tests is being pursued by the U.S. Air Force School of Aerospace Medicine Operational Based Vision Assessment Laboratory to modernize U.S. Air Force vision screening and establish quantitative relationships between vision test results and operationally relevant performance. This report briefly describes each test and summarizes the normative data collected to establish descriptive statistics for each test and test reliability.					
15. SUBJECT TERMS Vision screening, color vision, vision testing, aircrew selection, U.S. Air Force color vision standards, cone contrast test, CCT, stereo acuity, contrast sensitivity, visual acuity, motion perception					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 46	19a. NAME OF RESPONSIBLE PERSON Marc Winterbottom, PhD
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

This page intentionally left blank.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iii
LIST OF TABLES	v
1.0 SUMMARY	1
2.0 PURPOSE/BACKGROUND.....	1
3.0 AVT DATA	1
3.1 Landolt C Contrast Sensitivity Test (LCST)	1
3.1.1 LCST Normative Data Collection	2
3.1.2 Effect of Viewing Distance on Acuity.....	4
3.1.3 Contrast/Acuity Test Comparison.....	5
3.1.4 Discussion	8
3.2 OBVA Cone Contrast Test (OCCT).....	8
3.2.1 Methods.....	9
3.2.2 Results.....	9
3.3 Dual Ring Stereo Acuity Test.....	9
3.3.1 Methods.....	11
3.3.2 Results.....	11
3.4 Stereo Test Comparison.....	14
3.4.1 Methods.....	14
3.4.2 Results.....	15
3.4.3 Discussion	22
3.5 Fusion Range Test.....	22
3.5.1 Methods.....	23
3.5.2 Results.....	23
3.6 OBVA Motion Coherence Test	26
3.6.1 Methods.....	26
3.6.2 Results.....	27
3.6.3 Discussion	29

TABLE OF CONTENTS (concluded)

	Page
3.7 OBVA Motion Coherence Test 2.0	30
3.7.1 Methods.....	30
3.7.2 Results.....	30
3.7.3 Discussion	32
3.8 AVT Descriptive Statistics	32
4.0 GENERAL DISCUSSION	33
5.0 REFERENCES	34
LIST OF ABBREVIATIONS AND ACRONYMS	36

LIST OF FIGURES

	Page
Figure 1. Landolt C contrast sensitivity test.....	2
Figure 2. Distribution of LCST visual acuity thresholds	2
Figure 3. The distribution of scores on the LCST for each of the three letter sizes.	3
Figure 4. LCST test-retest results for visual acuity and each of the three contrast letter sizes....	3
Figure 5. Differences between test-retest for visual acuity and each of the three contrast letter sizes.....	4
Figure 6. Average LCST visual acuity and SE at each of the four viewing distances.....	5
Figure 7. The relationship between the OBVA LCST and the AST area under the curve metrics (left) and the LCST and AST acuity metrics (right)	6
Figure 8. The relationship between each of the LCST contrast and acuity letter size thresholds and the SV contrast and acuity chart test scores.....	6
Figure 9. The relationship between the AST contrast test for several spatial frequencies and the SV contrast score.....	7
Figure 10. OCCT L, M, and S Landolt C optotypes	8
Figure 11. The distribution of OCCT scores for L, M, and S cones	9
Figure 12. Differences between test-retest for each of the three cones, L (top left), M (top right), and S (bottom).....	10
Figure 13. Dual ring stereo acuity test stimulus.....	10
Figure 14. Logitech game controller used to enter responses for vision tests.....	11
Figure 15. The distribution of scores for the near and far dual ring stereo acuity tests	11
Figure 16. Scatterplots showing relationship between test-retest stereo acuity results at near (left) and far (right)	12
Figure 17. Scatterplot showing the relationship between stereo acuity test scores at near and far viewing distances.....	12
Figure 18. Differences between pairs of dual ring stereo acuity tests for near test 1 and test 2 (top left), far test 1 and test 2 (top right), and near 1 – far 1 (bottom)	13
Figure 19. Effect of practice on near stereo acuity test scores	13
Figure 20. Scatterplot showing the relationship between OBVA far stereo acuity test results and AFVT/AOV stereo acuity test results	14
Figure 21. Stereo search test (right and left eye images)	15
Figure 22. Scatterplots showing relationship between test-retest stereo acuity results for dual ring 2-second presentation time (top left), dual ring 4-second presentation time (top right), dual ring 8-second presentation time (bottom left), and single ring (bottom right).....	16

LIST OF FIGURES (continued)

	Page
Figure 23. Differences between test-retest stereo acuity results for dual ring 2-second presentation time (top left), dual ring 4-second presentation time (top right), dual ring 8-second presentation time (bottom left), and single ring (bottom right) ...	17
Figure 24. Scatterplots showing relationship between test-retest stereo acuity results for the SST with no monocular cue masking dual ring (left), with monocular cue masking (right).....	17
Figure 25. Scatterplot showing the relationship between the SST with and without monocular cue masking enabled	18
Figure 26. Test-retest differences for the SST	19
Figure 27. Scatterplots showing the correlations between SST with monocular cue masking (left column) and SST with no monocular cue masking (right column) with each of the dual ring tests (2-second stimulus duration top row, 4-second middle row, and 8-second bottom row)	20
Figure 28. Scatterplots showing the differences between the SST with monocular cue masking (left column) and SST with no monocular cue masking (right column) with each of the dual ring tests (2-second stimulus duration top row, 4-second middle row, and 8-second bottom row)	21
Figure 29. Horizontal fusion range stimuli (left and right eye images)	22
Figure 30. Distribution of horizontal fusion range scores in log arcmin	23
Figure 31. Scatterplot showing repeated test results for the horizontal fusion range test.....	24
Figure 32. Distribution of vertical fusion range scores in log arcmin for 93 subjects	24
Figure 33. Scatterplot showing repeated test results for the vertical fusion range test	25
Figure 34. Differences between test-retest for horizontal (left) and vertical fusion range (right)	25
Figure 35. Motion sensitivity test stimuli.....	26
Figure 36. Histograms for rotational (left) and radial (right) motion for 55 subjects	27
Figure 37. Scatterplots showing the relationship between repetitions of the motion perception test for clockwise (top left), counterclockwise (top right), expansion (bottom right), and contraction (bottom left).....	27
Figure 38. The relationship between clockwise rotation and counterclockwise rotation motion thresholds (left) and between expansion and contraction thresholds (right) ..	28
Figure 39. The relationship between clockwise rotation thresholds and contraction motion thresholds (left) and between expansion thresholds (right)	28
Figure 40. The differences between test-retest for clockwise (top left), counterclockwise (top right), expansion (bottom left), and contraction (bottom right)	29

LIST OF FIGURES (concluded)

	Page
Figure 41. Histograms for radial motion 2.0 thresholds (black bars, n = 30) and rotational motion 2.0 thresholds (white bars, n = 32)	30
Figure 42. Test-retest reliability for rotation (top left) and radial motion (top right) and the relationship between rotational and radial thresholds (bottom).....	31
Figure 43. Motion 1.0 test-retest for average rotational (left) and radial motion (right)	31
Figure 44. Differences between test-retest for rotation (left) and radial (right) motion 2.0 thresholds	32

LIST OF TABLES

	Page
Table 1. The Correlations Between Each AST Spatial Frequency Contrast Threshold and SV Contrast Scores	8
Table 2. Mean, Standard Deviation, and Test-Retest Correlations for Motion 1.0 and Motion 2.0	32
Table 3. Descriptive Statistics for Each of the AVTs	33

This page intentionally left blank.

1.0 SUMMARY

The work detailed in this report was conducted by the Operational Based Vision Assessment Laboratory, Aeromedical Research Department, Human Performance Branch, U.S. Air Force School of Aerospace Medicine, Wright-Patterson Air Force Base, OH, with support from KBRwyle. The report describes research conducted to support the development of new automated vision tests (AVTs) designed to more precisely measure color vision, visual acuity, contrast sensitivity, stereo acuity, ocular alignment (fusion range), and motion perception. This report summarizes research toward developing normative data for the new AVTs and to evaluate the test-retest reliability of each test in preparation for commercializing and deploying the AVT.

2.0 PURPOSE/BACKGROUND

The U.S. Air Force School of Aerospace Medicine (USAFSAM) Operational Based Vision Assessment (OBVA) Laboratory has developed a set of computer-based, automated vision tests, or AVT. Development of the AVT was required to support threshold-level vision testing capability needed to investigate the relationship between ocular health and operationally relevant performance. The design and operation of the AVT have been described in previous reports [1,2]. Existing standard tests generally do not support the level of accuracy and repeatability to support correlation analysis. Furthermore, more precise and repeatable vision tests are required to support U.S. Air Force Surgeon General objectives for enhanced health diagnostics and operational medicine. The AVT research grade tests also support interservice, international, industry, and academic partnerships. The AVT software has been provided to key partners to support ongoing research collaboration. This report summarizes AVT data collected to date and provides descriptive statistics for each test as well as test-retest reliability.

3.0 AVT DATA

3.1 Landolt C Contrast Sensitivity Test (LCST)

In the LCST, a Landolt C is presented on the display with the gap in the C at four possible positions: left, right, top, or bottom. The participant's task is to identify the gap location using the keyboard arrows to respond. Across trials, the Landolt C may appear gray (contrast test) or bright white/black (visual acuity test). The Landolt C can be presented in either positive (i.e., white/light colored letter against darker gray background) or negative contrast (i.e., black/dark letter against a light gray background). Figure 1 shows an example of the size reticle and Landolt C at a high and a low contrast (negative contrast). The size of the letter is varied across trials for the contrast and acuity Landolt C tests. For contrast sensitivity testing, the Landolt C set has three sizes: 50, 12.5, and 6.25 arcmin (16.7-, 2.5-, 1.25-arcmin gap size). The contrast of the Landolt C is varied according to the psi adaptive procedure [3] for the contrast test. The size is varied according to the adaptive procedure for the acuity test. The participant's accuracy in identifying the Landolt C orientation as contrast/size varies is recorded by the application. Speakers or headphones are used to enable auditory feedback. Correct responses are indicated using a positive/pleasant sound ("ding"), while incorrect responses are indicated using a negative/unpleasant sound (buzzer).



Figure 1. Landolt C contrast sensitivity test. An example of the size reticle (left). Landolt C with the gap positioned on the left and relatively high contrast (center). Landolt C with upward orientation and low contrast (right).

3.1.1 LCST Normative Data Collection. The LCST was administered to 66 subjects at the USAFSAM OBVA Laboratory. All subjects provided informed consent.

Figure 2 shows the distribution of LCST acuity thresholds. The top horizontal axis shows Snellen fractions and the bottom horizontal axis shows logMAR values. As shown, acuity varies only over a fairly narrow range between 20/10 and 20/20 (0.3 logMAR range). Figure 3 shows the distribution of contrast sensitivity scores on the LCST for each of the three letter sizes. As shown, contrast sensitivity varies by nearly a log unit for the two smaller letter sizes and approximately 0.5 log units for the largest letter size. Figure 4 shows the LCST test-retest results for acuity ($r = 0.72$, $p < 0.001$), 1.25-arcmin gap ($r = 0.81$, $p < 0.001$), 2.5-arcmin gap ($r = 0.84$, $p < 0.001$), and 16.67-arcmin gap ($r = 0.71$, $p < 0.001$) letter sizes. The LCST clearly provides highly repeatable results.

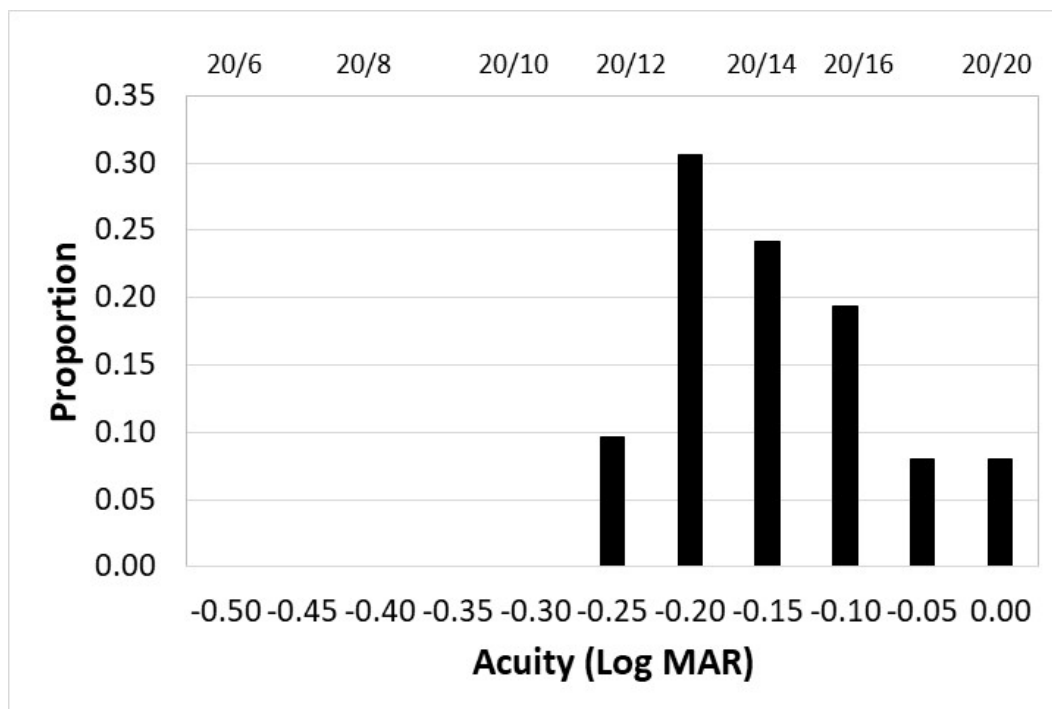


Figure 2. Distribution of LCST visual acuity thresholds.

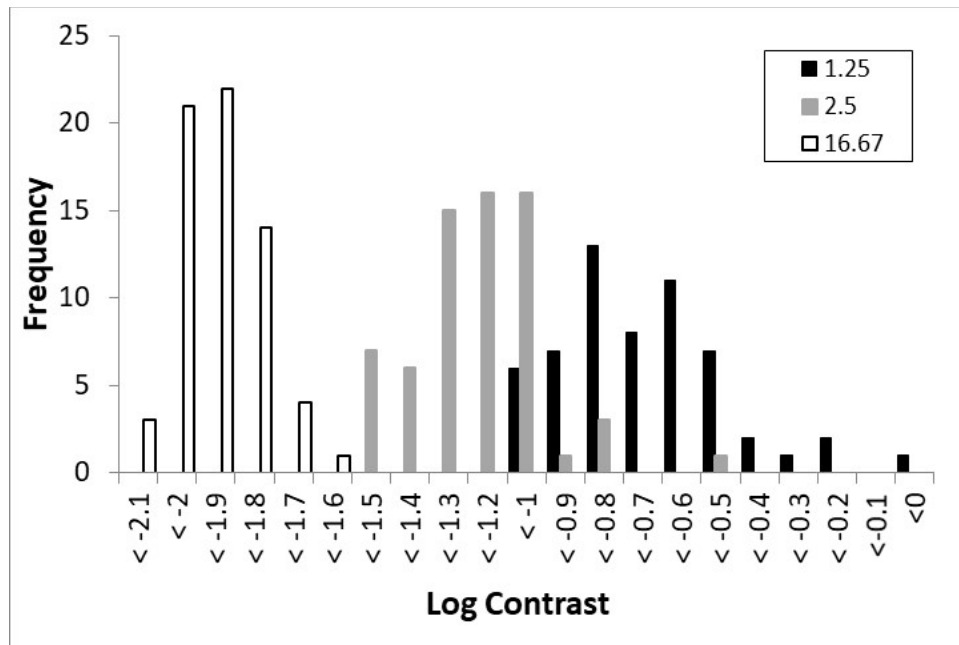


Figure 3. The distribution of scores on the LCST for each of the three letter sizes.

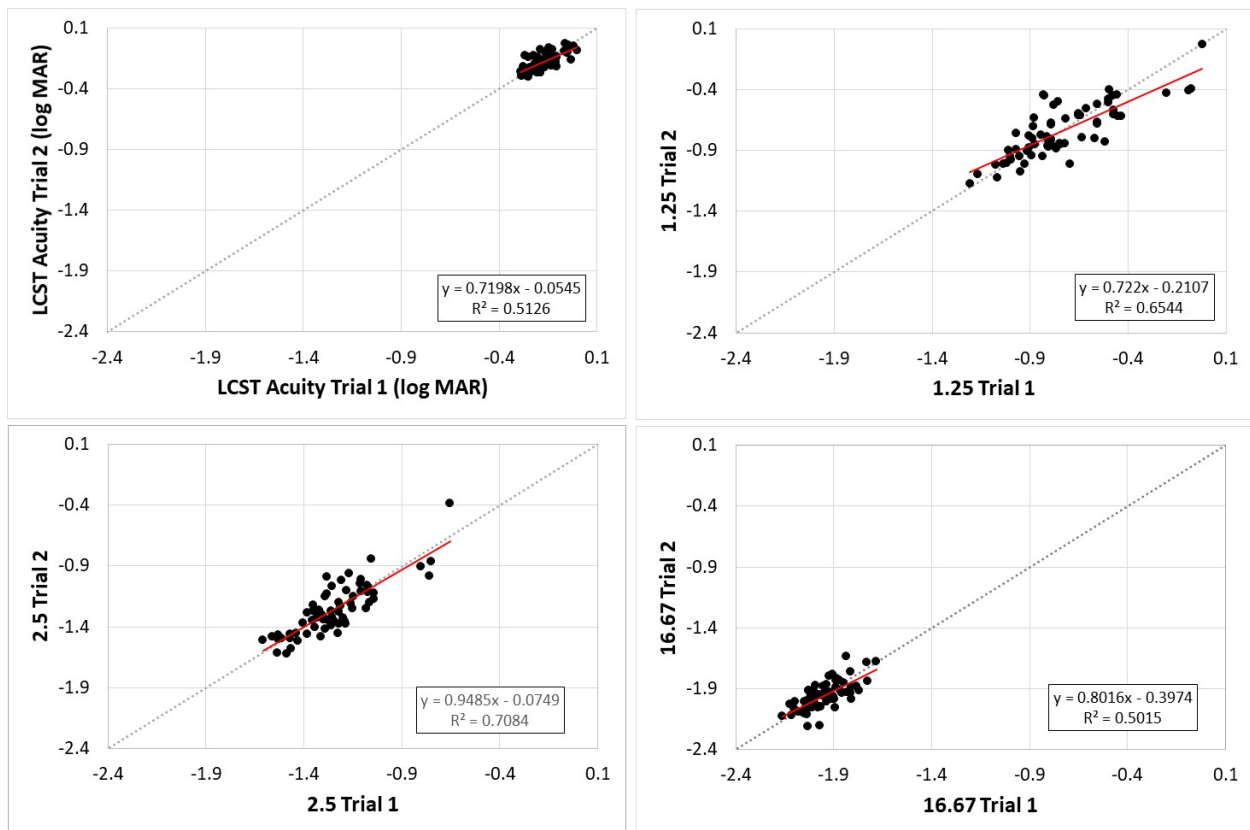


Figure 4. LCST test-retest results for visual acuity and each of the three contrast letter sizes. Top left: visual acuity (logMAR). Top right: 1.25-arcmin gap (log contrast). Bottom left: 2.5-arcmin gap (log contrast). Bottom right: 16.67-arcmin gap (log contrast).

Figure 5 shows the differences between test-retest for acuity (top left), 1.25 gap size (top right), 2.5 gap size (bottom left), and 16.67 gap size (bottom right). The mean difference and standard deviation (SD) of the differences are also shown for each pair of LCST tests. As shown in Figures 4 and 5, test-retest reliability for the LCST is very good. However, the slope of the fits in Figure 4 deviate somewhat from a desired value of 1, or from a desired value of 0 in Figure 5, indicating that there may be a small learning effect for this test. This will be examined more closely in future research.

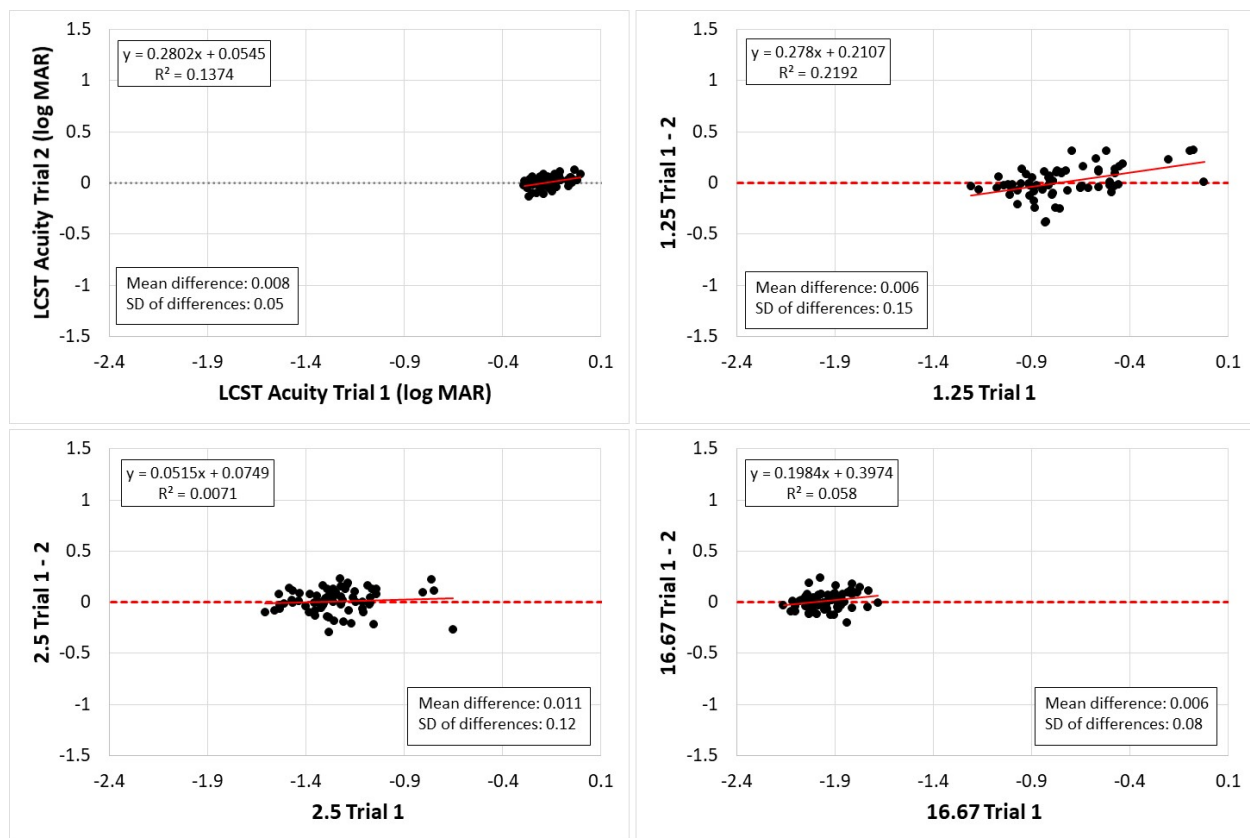


Figure 5. Differences between test-retest for visual acuity and each of the three contrast letter sizes. Top left: visual acuity (logMAR). Top right: 1.25-arcmin gap (log contrast). Bottom left: 2.5-arcmin gap (log contrast). Bottom right: 16.67-arcmin gap (log contrast). The average of the differences and SD of the differences are also shown for each pair of LCST tests.

3.1.2 Effect of Viewing Distance on Acuity. The LCST acuity test was administered to eight subjects with good visual acuity. Each subject completed the LCST acuity twice at each of four viewing distances: 1, 2, 4, and 8 meters. The objective was to determine whether viewing distance affected visual acuity thresholds. This experiment was conducted to determine whether test results could be limited by the resolution of the NEC monitor as the viewing distance decreased and angular pixel size increased.

Figure 6 shows average visual acuity at each of the four viewing distances. As shown, viewing distance had very little effect on estimated thresholds. The standard errors (SEs) are also shown; however, the error bars are smaller than the data markers.

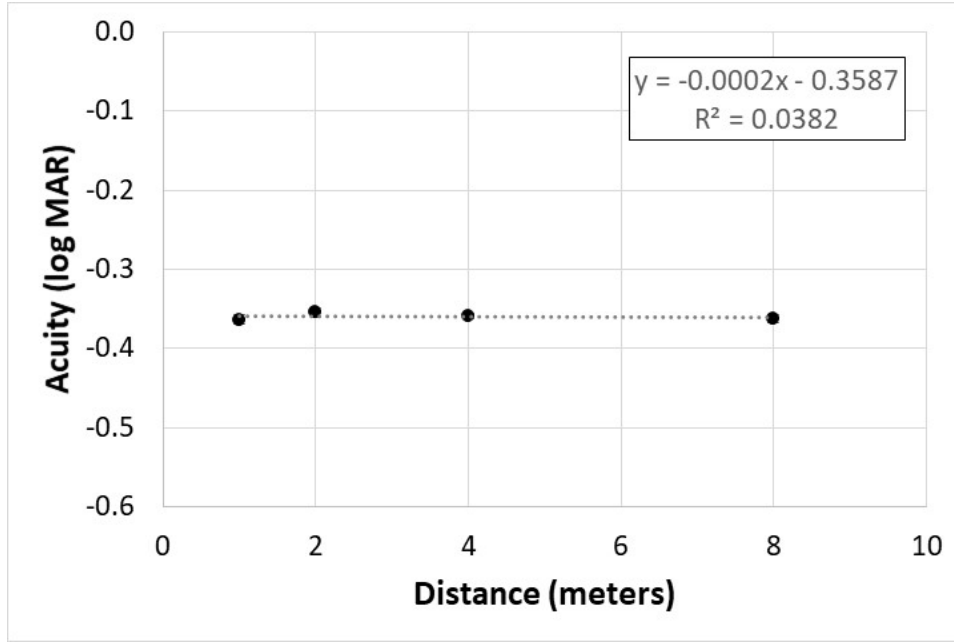


Figure 6. Average LCST visual acuity and SE at each of the four viewing distances.

3.1.3 Contrast/Acuity Test Comparison. The LCST was compared to two other contrast/acuity vision tests – a contrast/acuity test developed by Adaptive Sensory Technology (AST) modified to use a Landolt C optotype and the Precision Vision Rabin Super Vision (SV) contrast and acuity charts. Twenty-eight subjects participated in this experiment. It is important to note that the Precision Vision chart contrast values are Michelson contrast, which is usually applied for periodic patterns (e.g., sinewave gratings).

$$\text{Michelson Contrast} = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}$$

The SV chart contrast values were confirmed based on photometric measurements. All results reported here are based on Weber contrast, which is more appropriate for targets (e.g., letters) presented against a uniform background.

$$\text{Weber Contrast} = \frac{L_{\text{target}} - L_{\text{background}}}{L_{\text{background}}}$$

Figure 7 (left) shows the relationship between the OBVA LCST and the AST contrast test. The two tests are compared based on the metric provided by AST, area under the log contrast sensitivity function, and by a similar metric computed for the LCST, log area under the curve (LAUC). LAUC was computed based on the contrast sensitivity values for each of the three letter sizes and the acuity threshold. The AST test also produces a visual acuity metric. Figure 7 (right) shows the relationship between the AST acuity and the LCST visual acuity. As shown, both the overall area under the curve ($r = 0.88$, $p < 0.001$) and visual acuity tests ($r = 0.67$, $p < 0.001$) result in very good agreement.

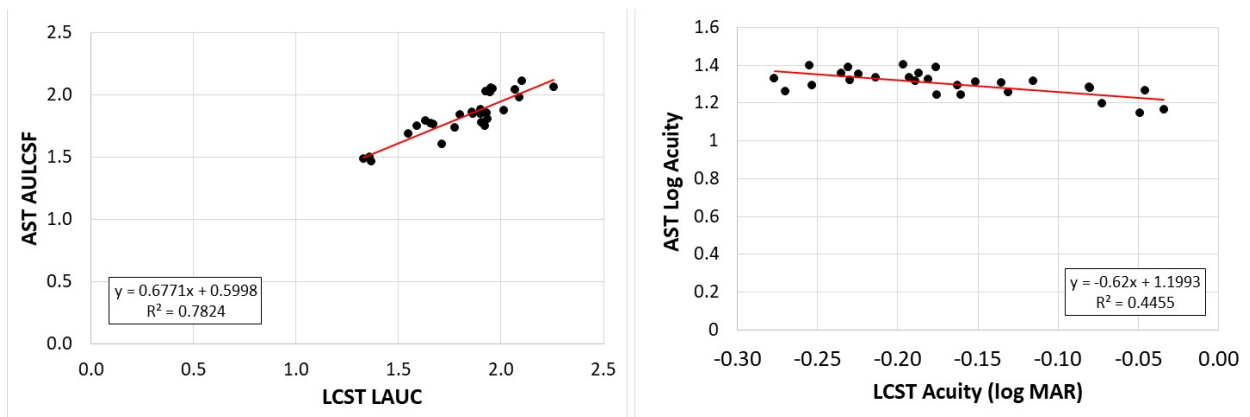


Figure 7. The relationship between the OBVA LCST and the AST area under the curve metrics (left) and the LCST and AST acuity metrics (right). AULCSF = area under the log contrast sensitivity function.

Figure 8 shows the relationship between each of the LCST contrast and acuity letter size thresholds and the SV contrast and acuity chart test scores. As shown, the LCST and SV tests agree very well for the 1.25 ($r = 0.69$, $p < 0.001$) and 2.5 ($r = 0.76$, $p < 0.001$) LCST letter sizes and quite well for the 16.67 ($r = 0.54$, $p = 0.003$) and acuity tests ($r = 0.59$, $p = 0.001$), although the agreement is not as strong as between the LCST and AST test results.

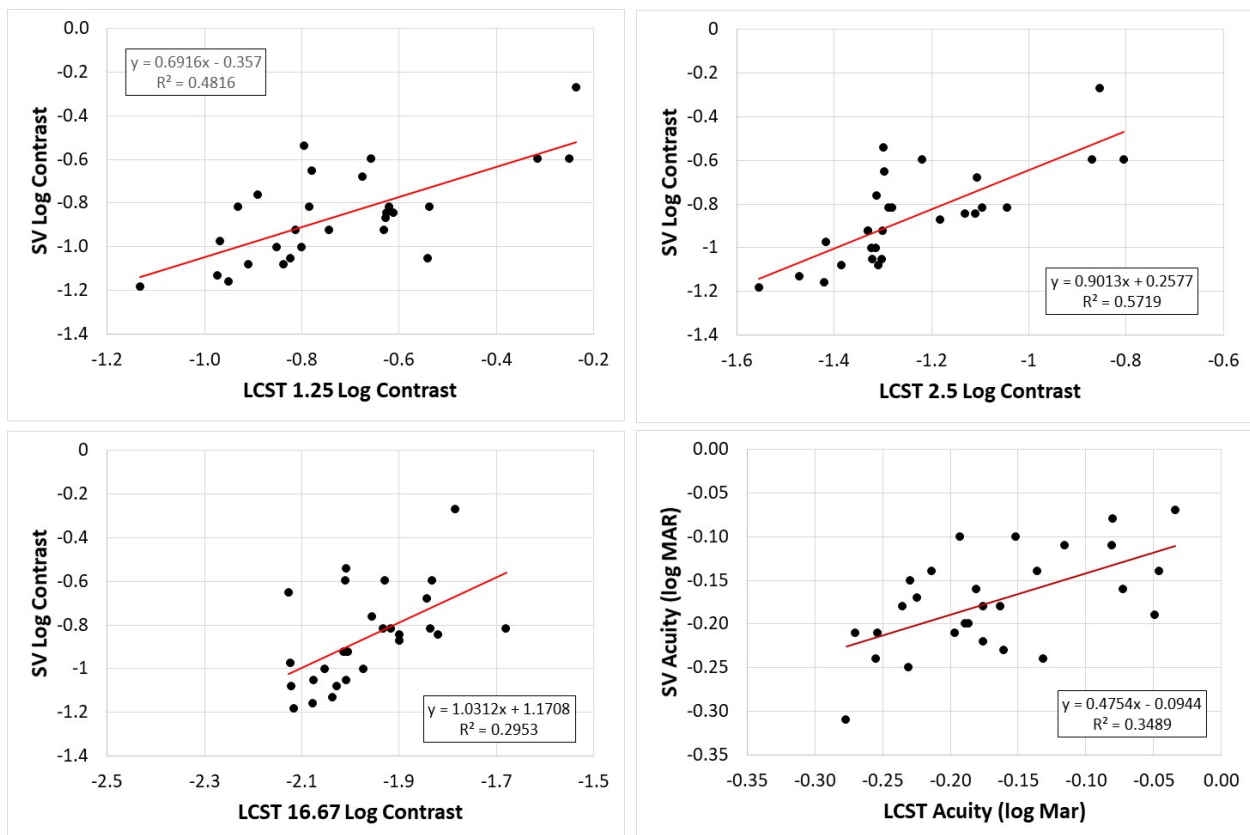


Figure 8. The relationship between each of the LCST contrast and acuity letter size thresholds and the SV contrast and acuity chart test scores.

Figure 9 shows the relationship between the AST contrast test for several spatial frequencies and the SV contrast score. Table 1 summarizes the correlations between each AST spatial frequency contrast threshold and SV contrast scores.

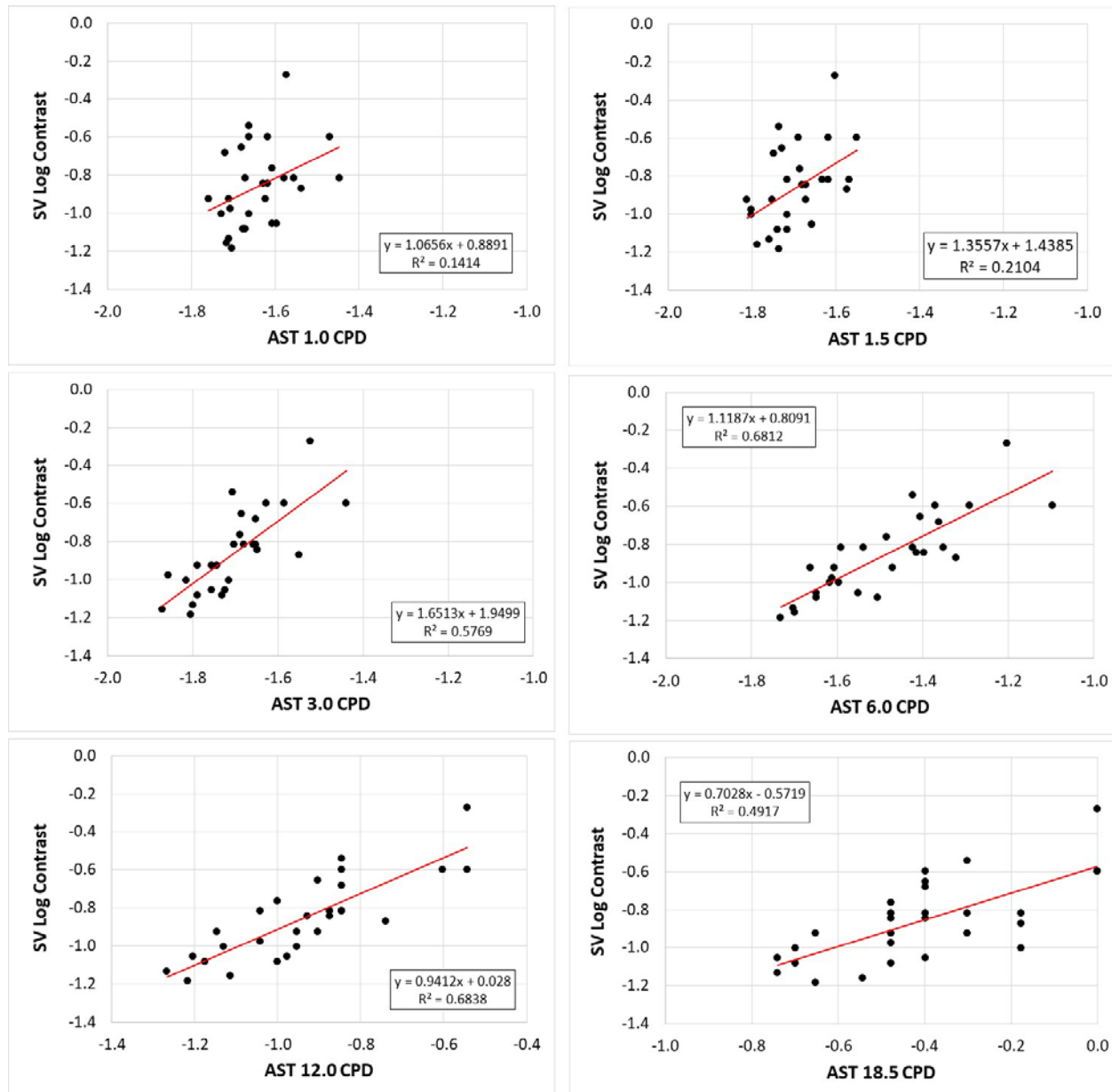


Figure 9. The relationship between the AST contrast test for several spatial frequencies and the SV contrast score.

Table 1. The Correlations between Each AST Spatial Frequency Contrast Threshold and SV Contrast Scores

AST Frequency	r	Significance
1.0	0.38	0.05
1.5	0.46	0.01
3.0	0.76	<<0.001
6.0	0.83	<<0.001
12.0	0.83	<<0.001
18.5	0.70	<<0.001

3.1.4 Discussion. The OBVA LCST provides highly repeatable measures of contrast sensitivity and visual acuity. The newly developed LCST test results agree very well with the AST contrast sensitivity test, a recently introduced commercially available computer-based contrast sensitivity test, as well as traditional chart-based visual acuity and contrast tests. The electronic display used for the development of the LCST provides adequate resolution to accurately measure visual acuity even at relatively near distances (1 meter). Finally, the test duration can very likely be reduced by forgoing the visual acuity test.

3.2 OBVA Cone Contrast Test (OCCT)

The OCCT builds on the success of the Rabin CCT [4] through the use of highly accurate color display calibration, use of Landolt C optotypes to simplify response entry, and adoption of adaptive threshold estimation procedures well described in the research literature [3]. The OCCT is described in more detail in a previous report [1]. Here, we describe data collected using the OCCT to build a normative database. The OCCT is similar to the LCST except that, instead of varying luminance contrast, long (L), medium (M), and short (S) wavelength cone contrast (or red, green, and blue color) is varied according to the adaptive procedure. The colors red, green, and blue (r, g, and b) are selected to isolate the three cones for individuals with normal color vision (i.e., test the function of the red, green, and blue, or long, medium, and short wavelength receptors). The size of the gap in the Landolt C is 16.67 arcmin. Figure 10 illustrates the appearance of the L, M, and S Landolt C optotypes.



Figure 10. OCCT L, M, and S Landolt C optotypes.

3.2.1 Methods. Sixty-nine subjects were administered the OCCT to develop normative data and to examine test-retest reliability. All subjects provided informed consent.

3.2.2 Results. Figure 11 shows the distribution of OCCT scores for L, M, and S cones.

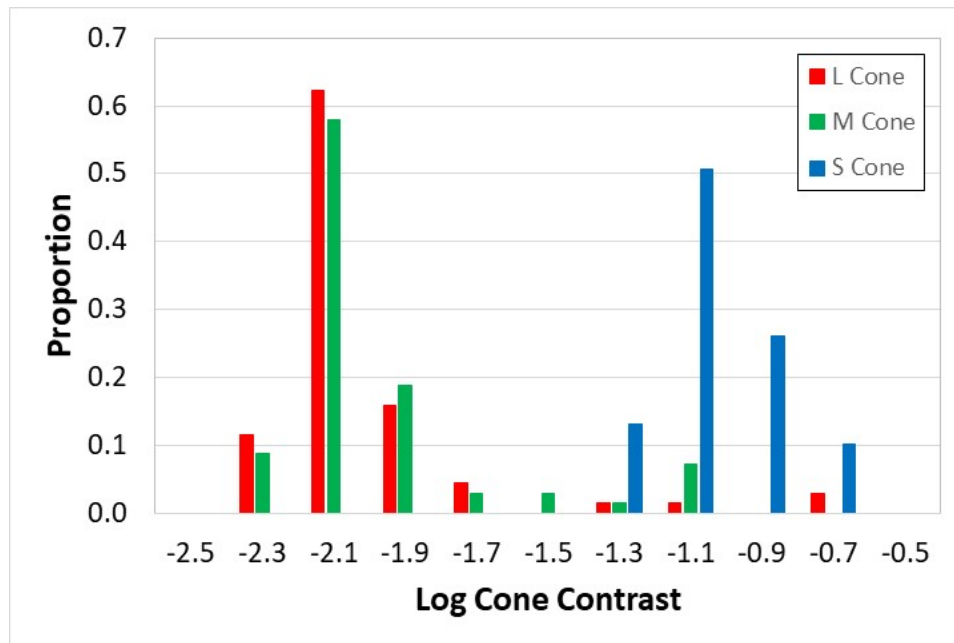


Figure 11. The distribution of OCCT scores for L, M, and S cones.

Figure 12 shows the test-retest reliability for L ($r = 0.93$, $p < 0.001$), M ($r = 0.96$, $p < 0.001$), and S ($r = 0.65$, $p < 0.001$) cones using the OCCT. Based on data collected to date, the repeatability of the L and M cone results is excellent with good reliability with the S cone.

3.3 Dual Ring Stereo Acuity Test

A well-known concern with the use of many standard stereo acuity tests is that some subjects may be able to pass the test by using monocular cues [5-8]. The dual ring stereo acuity test was designed to minimize monocular cues. The test also uses antialiasing/blurring to enable very small (sub-pixel) changes in disparity [9]. The dual ring stereo test also requires observers to discriminate between crossed vs. uncrossed disparity (i.e., decide whether the inner circle appears in front of or behind the reference plane). This is a more difficult task compared to most standard stereo acuity tests available today that require only that the observer detect crossed-only disparity (or maybe only to detect which target differs slightly within a set); therefore, we might expect higher thresholds for this test compared to more commonly used stereo acuity tests. These features were designed into the test to support very accurate stereo acuity threshold estimates required to support research concerning the importance of depth perception in operational tasks. Figure 13 shows the stimulus used for the dual ring stereo acuity test, which is similar in appearance to the Armed Forces Vision Tester (AFVT) circles test, or the American Optical vectograph (AOV) circles test.

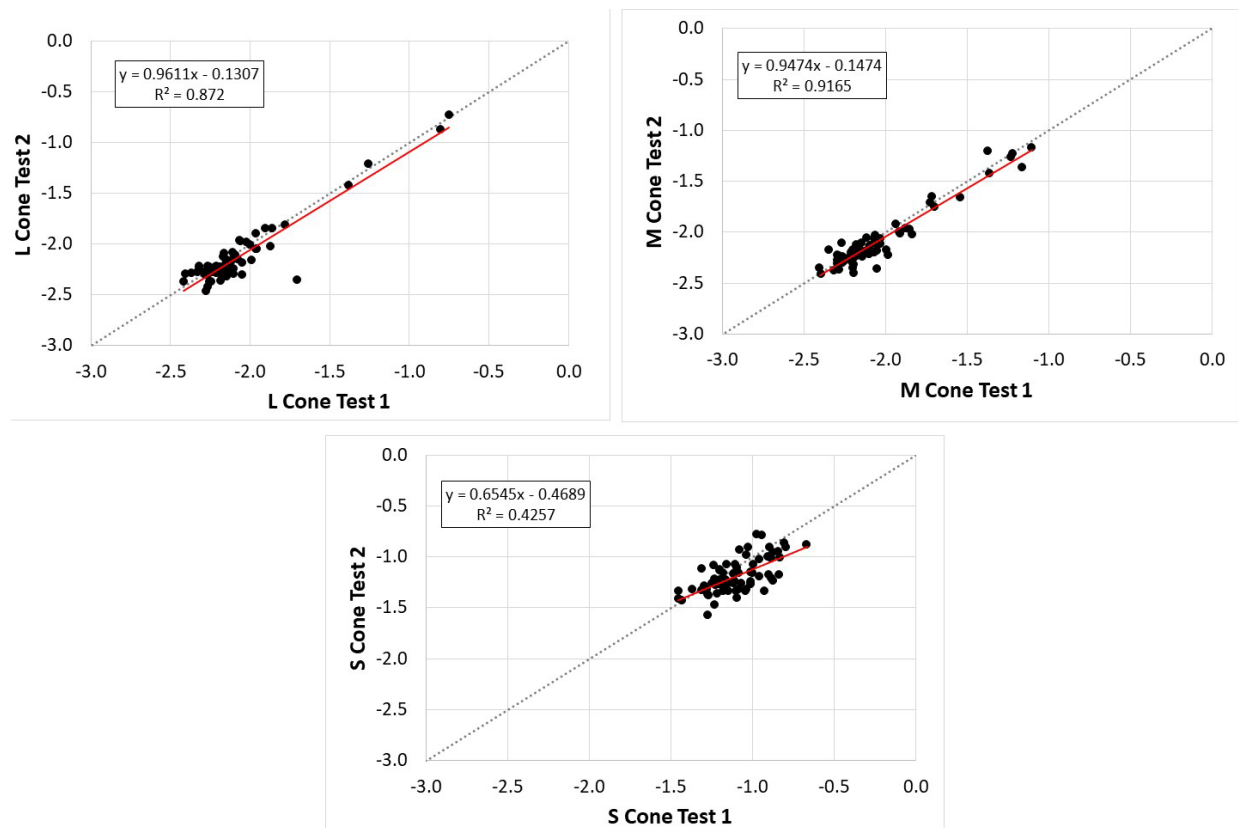


Figure 12. Differences between test-retest for each of the three cones, L (top left), M (top right), and S (bottom). The mean difference and SD of the differences are also shown for each pair of OCCT tests.

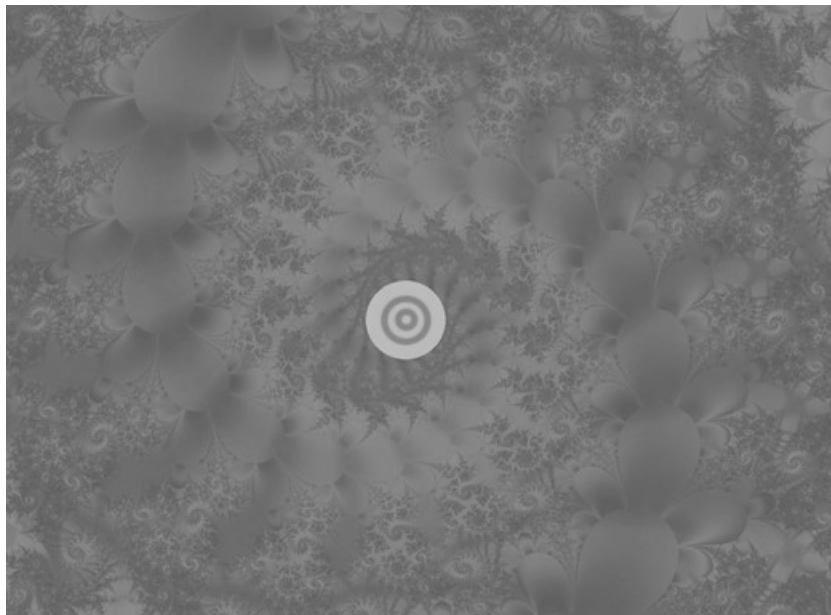


Figure 13. Dual ring stereo acuity test stimulus.

3.3.1 Methods. The dual ring stereo acuity test requires only a simple in front of or behind response, which was entered using a standard game pad controller (Figure 14). The stimuli were generated using a Dell Precision T7610 with Nvidia GeForce GTX 680 graphics card and displayed on an Asus VG278HE 3D monitor with 1920 x 1080 pixels that was compatible with Nvidia 3D Vision2 using active shutter glasses. The test was administered at two distances – 1 meter (near) and 4 meters (far). The software scales the size of the ring stimuli according to viewing distance. The disparity of the rings is varied according to the psi adaptive procedure [3] for the dual ring stereo acuity test.



Figure 14. Logitech game controller used to enter responses for vision tests.

3.3.2 Results. Figure 15 shows the distribution of scores for the near and far dual ring stereo acuity tests. As shown, stereo acuity scores are widely distributed and do not appear to have a normal distribution.

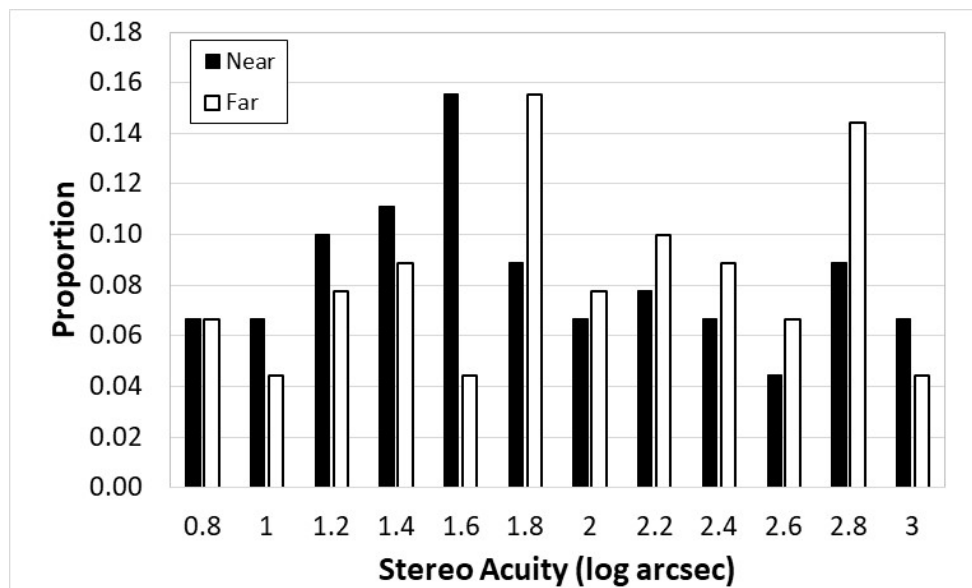


Figure 15. The distribution of scores for the near and far dual ring stereo acuity tests.

Figure 16 shows test-retest scores for the dual ring stereo acuity test at near (left) and at far (right) distances for 90 subjects. The correlation between each instance of the stereo test was high ($r = 0.83$, $r = 0.81$, respectively, p 's < 0.001). However, as shown, the stereo acuity test result for a few observers varies substantially across repeated tests for both near and far viewing distances. Preliminary analysis suggests that subjects may benefit from practice for near stereo acuity [$t(89) = 3.06$, $p = 0.003$], but not for far [$t(89) = 1.91$, $p = 0.06$]. However, the average difference is small (0.12 log units).

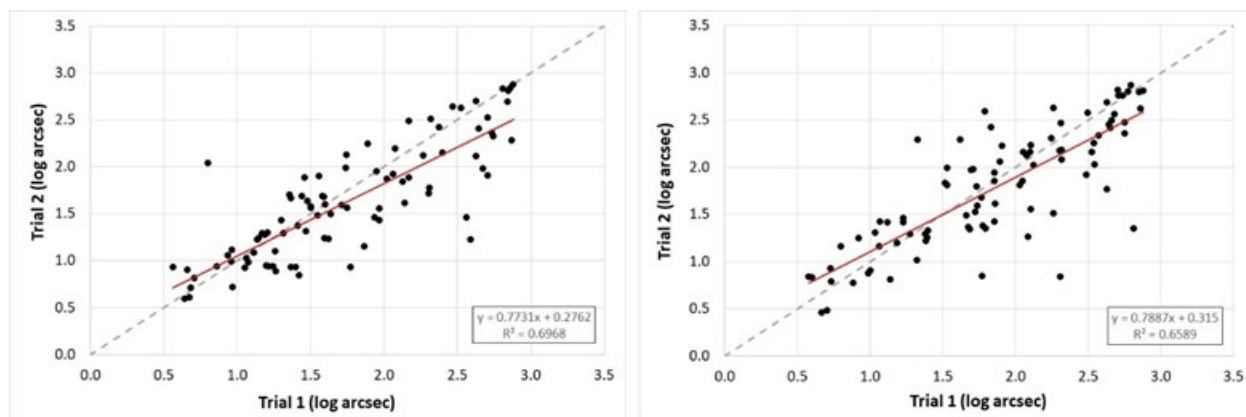


Figure 16. Scatterplots showing relationship between test-retest stereo acuity results at near (left) and far (right).

Figure 17 shows the relationship between near and far stereo acuity test results ($n = 89$). The correlation between near and far test scores is high ($r = 0.65$, $p < 0.001$). However, there is more variability across near and far stereo acuity compared to variability across repeated administrations of the stereo test at the same distance.

Figure 18 shows the differences between pairs of dual ring stereo acuity tests for near test 1 and test 2 (top left), far test 1 and test 2 (top right), and near 1 – far 1 (bottom). The average of the differences and SD of the differences are also shown for each pair of stereo tests.

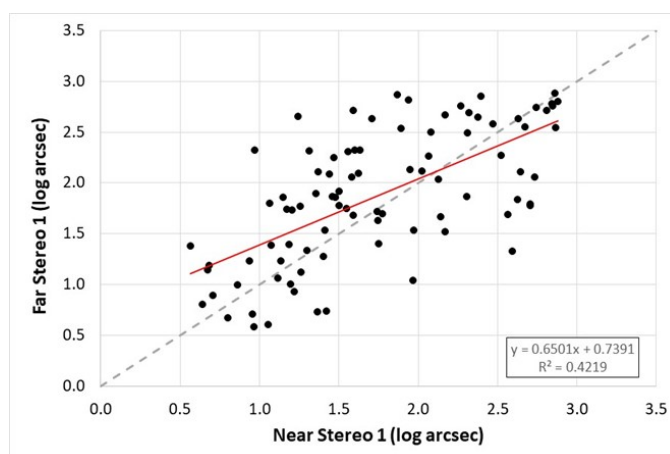


Figure 17. Scatterplot showing the relationship between stereo acuity test scores at near and far viewing distances.

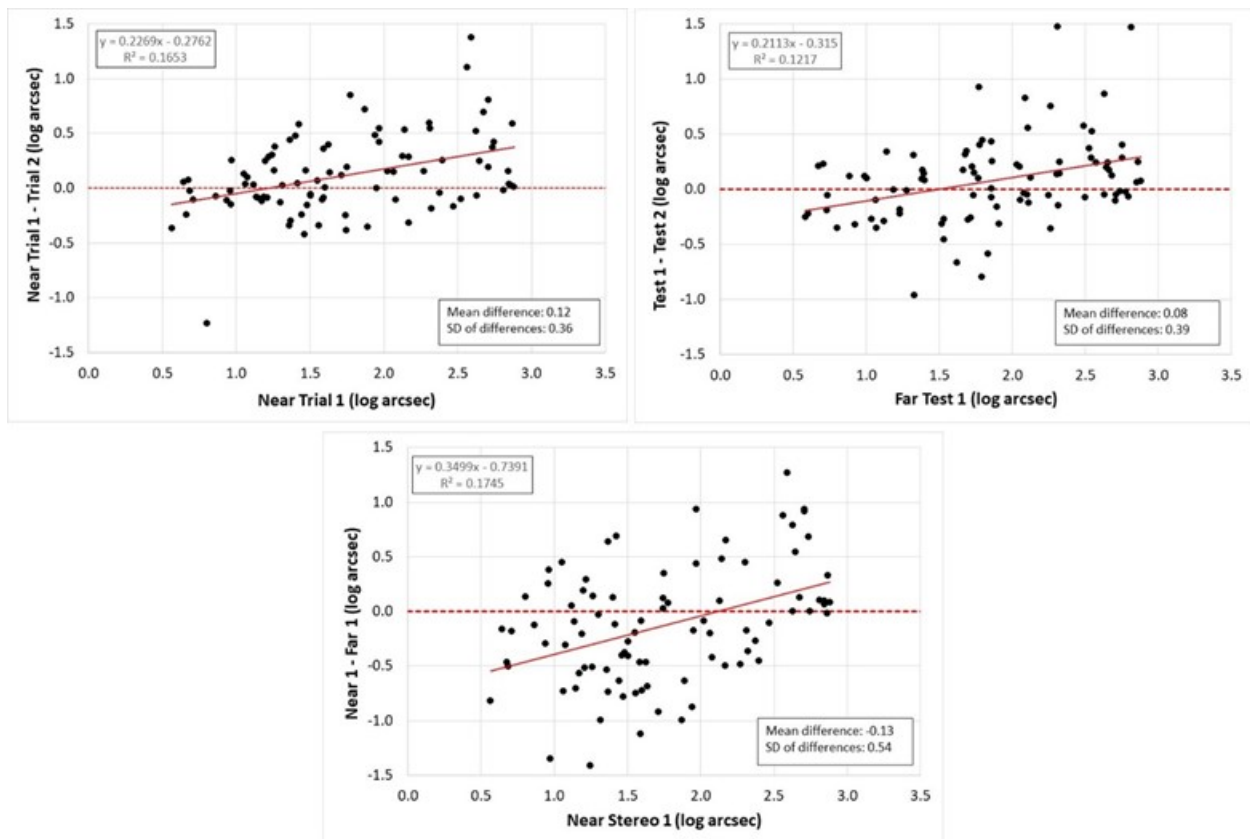


Figure 18. Differences between pairs of dual ring stereo acuity tests for near test 1 and test 2 (top left), far test 1 and test 2 (top right), and near 1 – far 1 (bottom). The average of the differences and SD of the differences are also shown for each pair of stereo tests.

Thirty subjects were administered the near stereo acuity test six times over the course of several days. As shown in Figure 19, subjects' stereo acuity appears to improve, on average, with practice. Stereo acuity improved from approximately 1.75 log arcsec (56 arcsec) to approximately 1.3 log arcsec (20 arcsec).

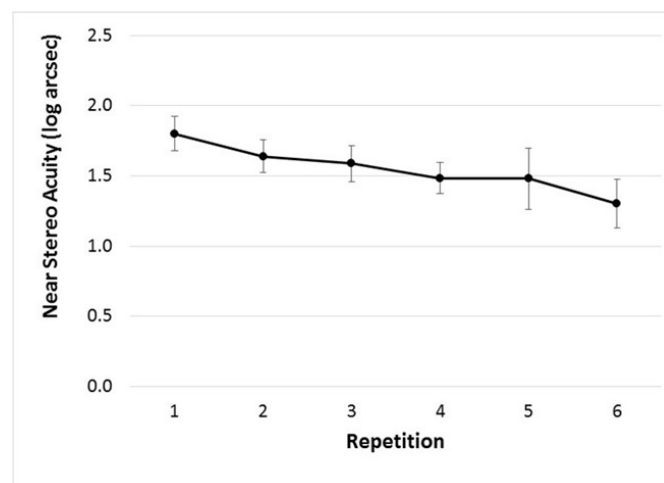


Figure 19. Effect of practice on near stereo acuity test scores.

Figure 20 shows the relationship between OBVA stereo acuity test results (far) and the AFVT stereo acuity test. Although the correlation between test scores is highly significant ($r = 0.62$, $p < 0.001$), there are clearly substantial differences between the outcomes of each test. The AFVT/AOV suffers from substantial quantization errors and floor effects. Further, subjects obtaining a passing score (< 25 arcsec) on the AFVT obtain widely varying scores on the OBVA stereo acuity test.

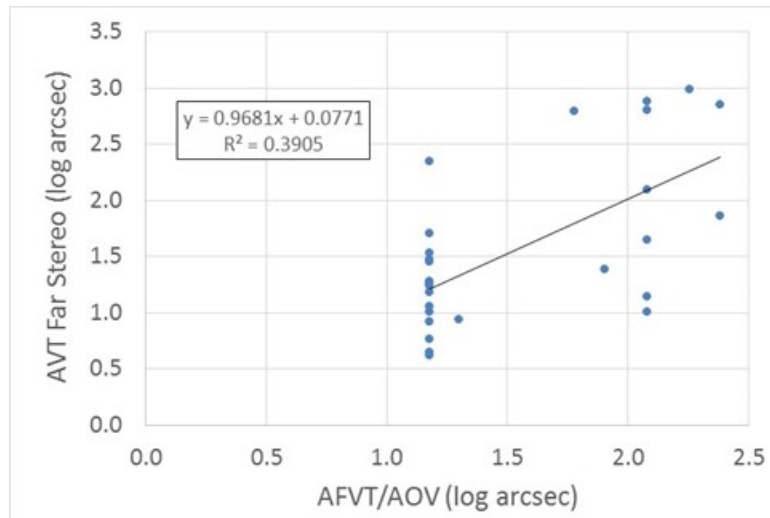


Figure 20. Scatterplot showing the relationship between OBVA far stereo acuity test results and AFVT/AOV stereo acuity test results.

3.4 Stereo Test Comparison

Because some subjects exhibit a significant amount of variability across repeated dual ring stereo acuity tests, we set out to identify potential factors that could contribute to that variability. These factors included stimulus presentation time, ring stimulus features, and type of response choice.

3.4.1 Methods. Three different versions of the stereo test were used for this experiment: 1) the dual ring stereo test described in section 3.3; 2) a single ring version of the stereo test; and 3) a third test designed to be more similar in appearance to the Titmus booklet test, which we will refer to as the stereo search test (SST). Additionally, three different stimulus presentation times were used for the standard dual ring stereo tests: 2, 4, and 8 seconds. Finally, monocular cue masking (random horizontal shifts in position of the rings) was either enabled or disabled for the SST. For all three tests, the stimuli were generated using a Dell Precision T7610 with Nvidia GeForce GTX 680 graphics card and displayed on an Asus VG278HE 3D monitor with 1920 x 1080 pixels that was compatible with Nvidia 3D Vision2 using active shutter glasses. The tests were all administered at a 1-meter viewing distance. For the dual ring and single ring tests, the response was the same as previously described in section 3.3, determine whether the inner ring was in front of or behind the outer ring and the reference plane, or, for the single ring test, determine whether the single ring was in front of or behind the reference plane. Responses were entered using the game pad. The disparity was varied according to the psi adaptive procedure. For the single ring test, only a 2-second stimulus presentation was used. For the SST, four circles were presented within a reference plane as shown in Figure 21. Participants were asked to

determine which of the four circles was popped out in depth (i.e., crossed disparity). One of the other four circles was displayed with an equal level of uncrossed disparity, and the two remaining circles were always presented at the same disparity as the reference plane. Disparity of the target circle was varied according to the psi adaptive procedure. Thus, the response decision for this test differed from the dual ring and single ring tests and required inspection of all four potential targets. Participants entered their responses (left, right, up, or down) using the game pad.

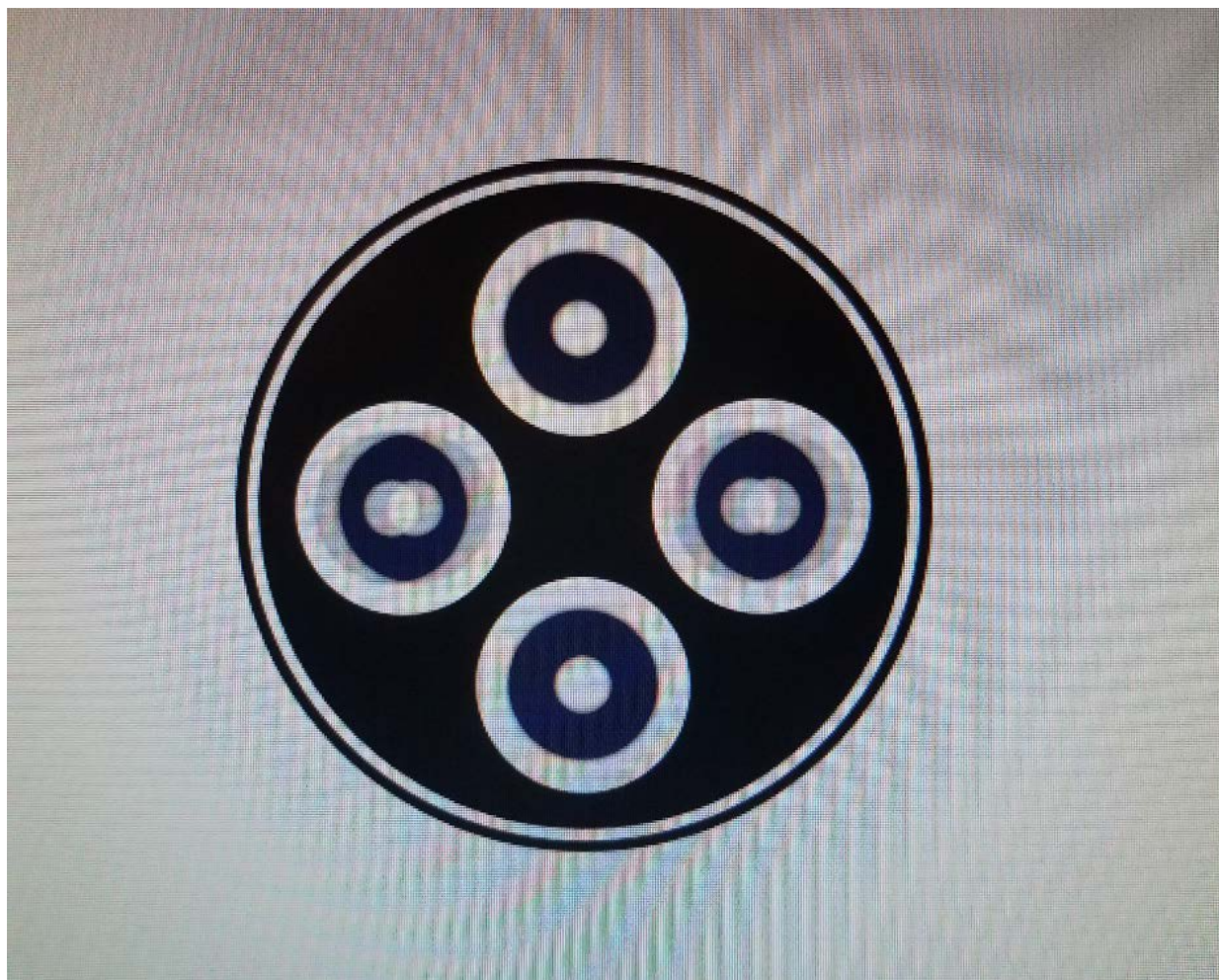


Figure 21. Stereo search test (right and left eye images).

3.4.2 Results. Figure 22 shows test-retest scores for both the dual ring (2-, 4-, and 8-second stimulus presentation) and single ring stereo acuity test for 30 subjects. All dual ring tests showed good test-retest reliability: 2 seconds ($r = 0.8$, $p < 0.001$), 4 seconds ($r = 0.84$, $p < 0.001$), and 8 seconds ($r = 0.84$, $p < 0.001$). However, the single ring stereo acuity had lower test-retest reliability ($r = 0.69$, $p < 0.001$). Note that the data for one subject were excluded for the dual ring 4-second and 8-second conditions due to widely varying results across test repetitions, very shallow slope estimates, and high SE estimates.

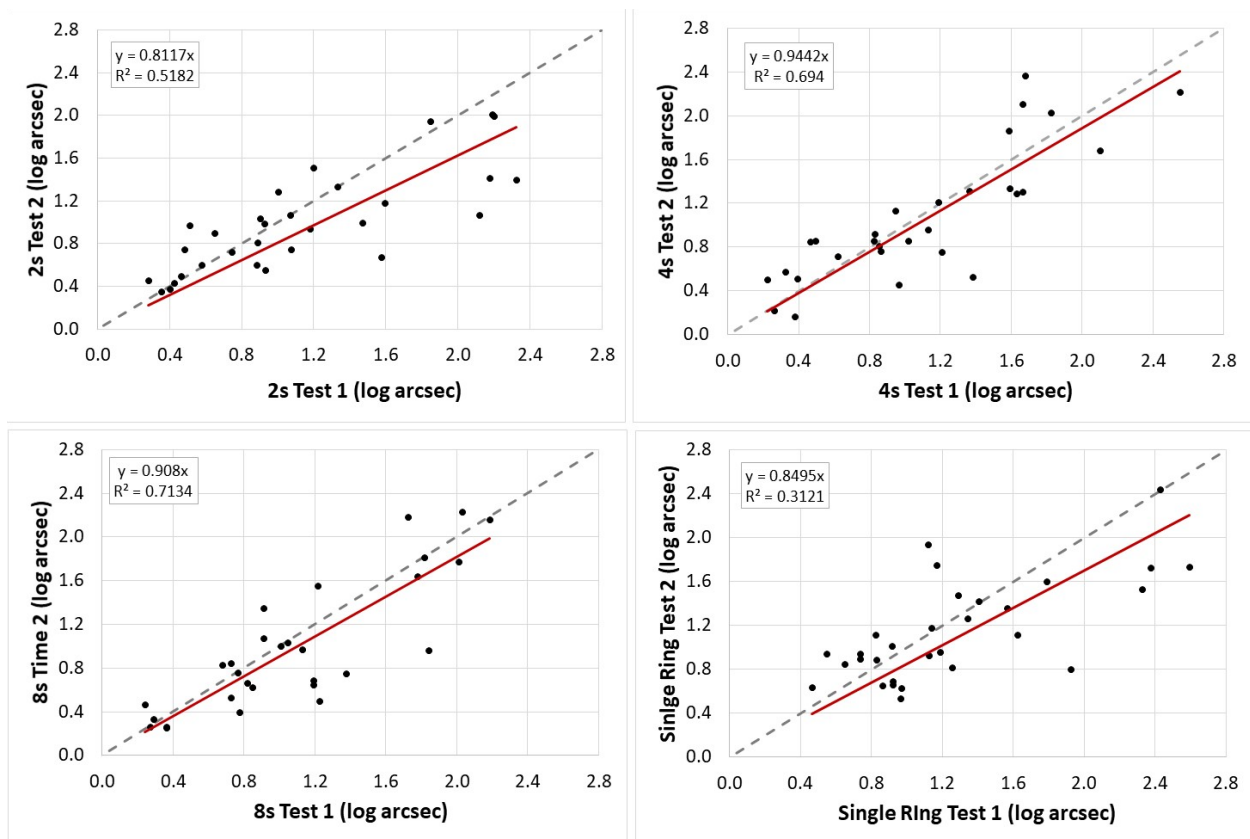


Figure 22. Scatterplots showing relationship between test-retest stereo acuity results for dual ring 2-second presentation time (top left), dual ring 4-second presentation time (top right), dual ring 8-second presentation time (bottom left), and single ring (bottom right).

Figure 23 shows the differences between pairs of dual ring stereo acuity tests for the dual ring (2-, 4-, and 8-second stimulus presentation) and single ring stereo acuity test. The average of the differences and SD of the differences are also shown for each pair of stereo tests. The dual ring test with the 4-second stimulus presentation time shows the least amount of learning.

Figure 24 shows test-retest scores for the SST with and without monocular cue masking for 20 subjects. Both tests show high test-retest reliability: without masking ($r = 0.95$, $p < 0.001$) and with masking ($r = 0.94$, $p < 0.001$).

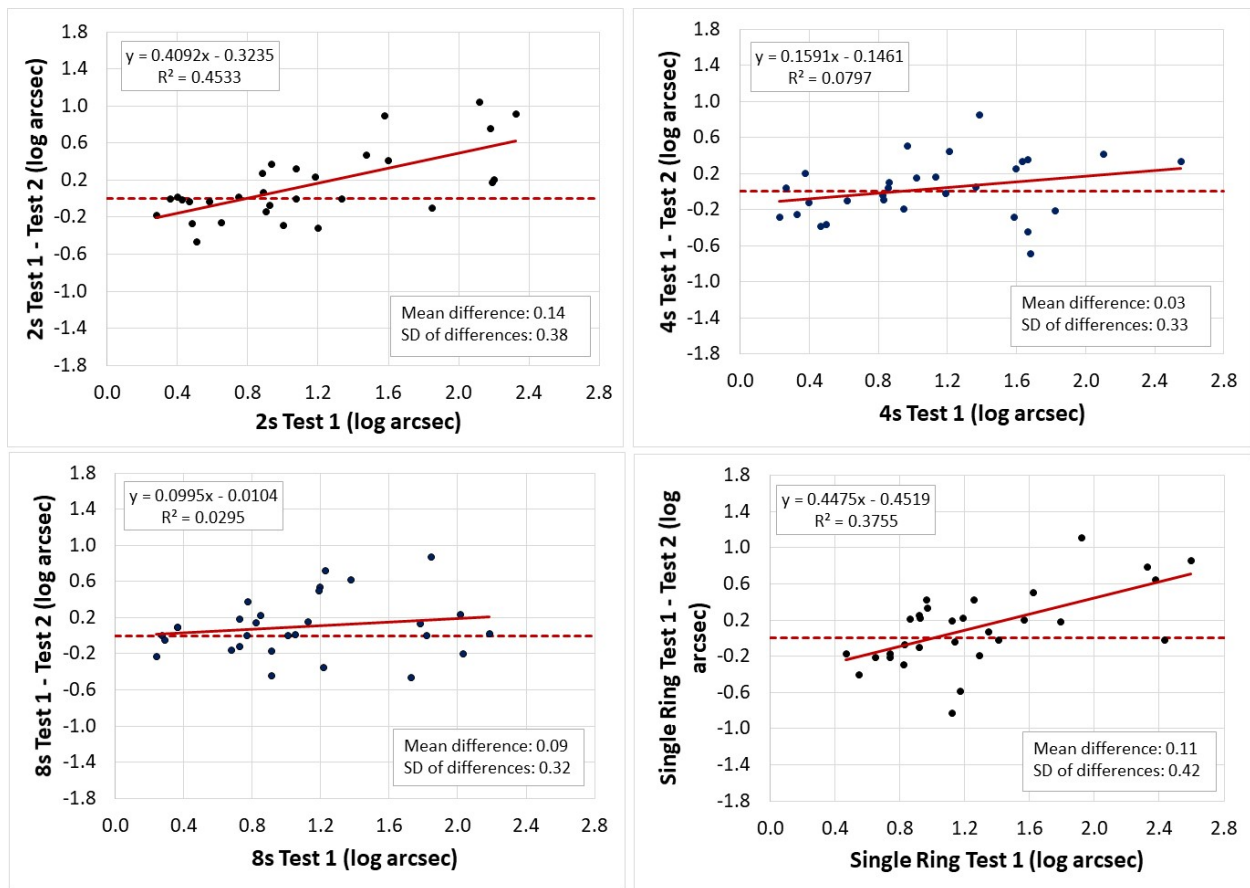


Figure 23. Differences between test-retest stereo acuity results for dual ring 2-second presentation time (top left), dual ring 4-second presentation time (top right), dual ring 8-second presentation time (bottom left), and single ring (bottom right).

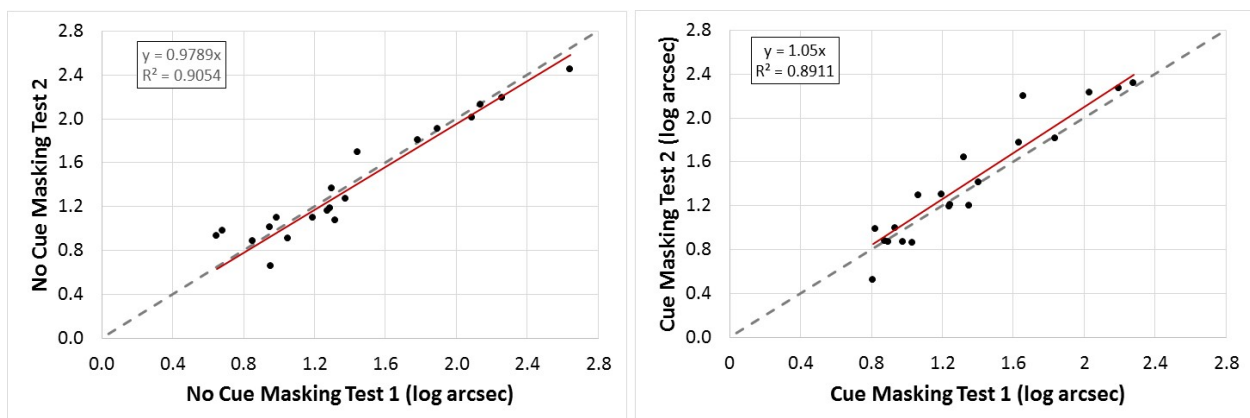


Figure 24. Scatterplots showing relationship between test-retest stereo acuity results for the SST with no monocular cue masking dual ring (left), with monocular cue masking (right).

Figure 25 shows the relationship between the SST with and without monocular cue masking enabled. The correlation between these two tests is high ($r = 0.92$, $p < 0.001$). In fact, the variability across the SST with and without monocular cue masking is about the same compared to variability across repeated administrations of each test with the same masking settings.

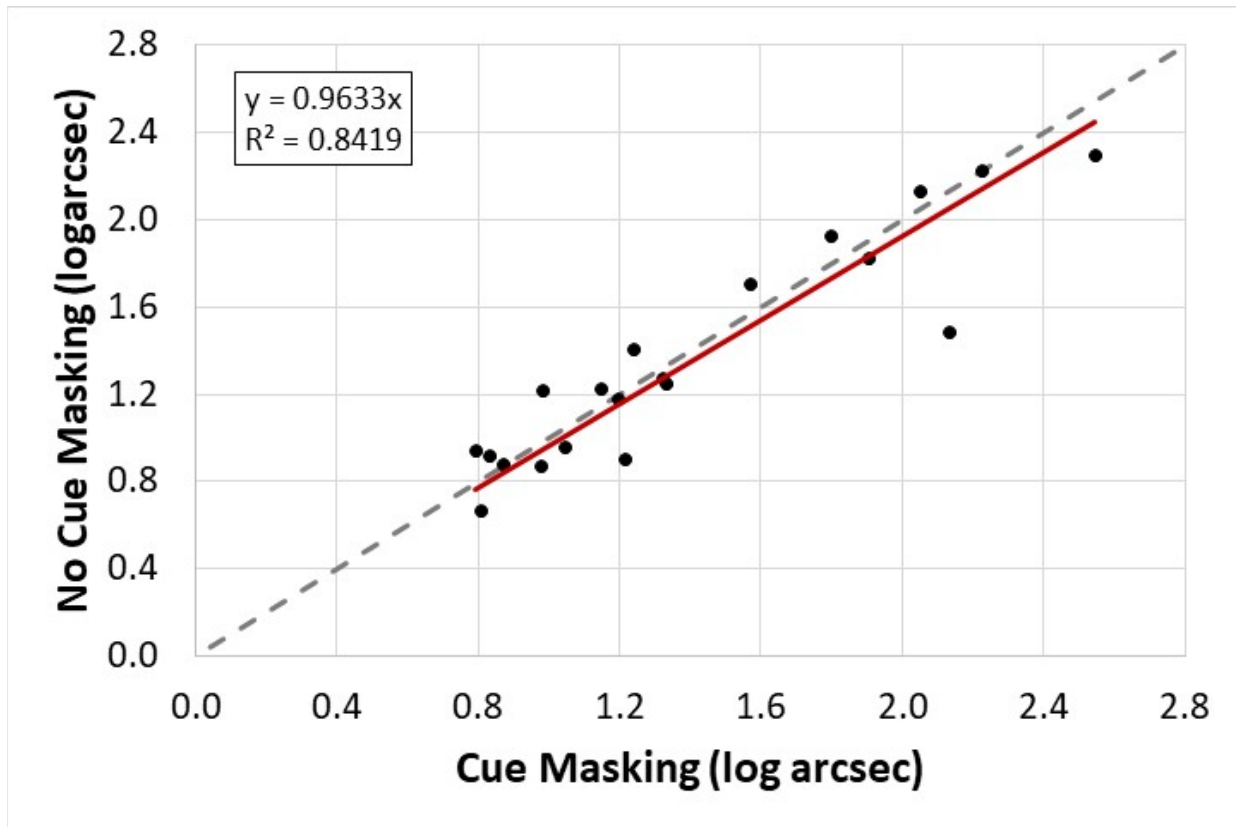


Figure 25. Scatterplot showing the relationship between the SST with and without monocular cue masking enabled.

Figure 26 shows the differences between pairs of SSTs for the no monocular cue masking tests 1 and 2 (top left), with monocular cue masking tests 1 and 2 (top right), and without masking – with masking (bottom). The average of the differences and SD of the differences are also shown for each pair of stereo tests.

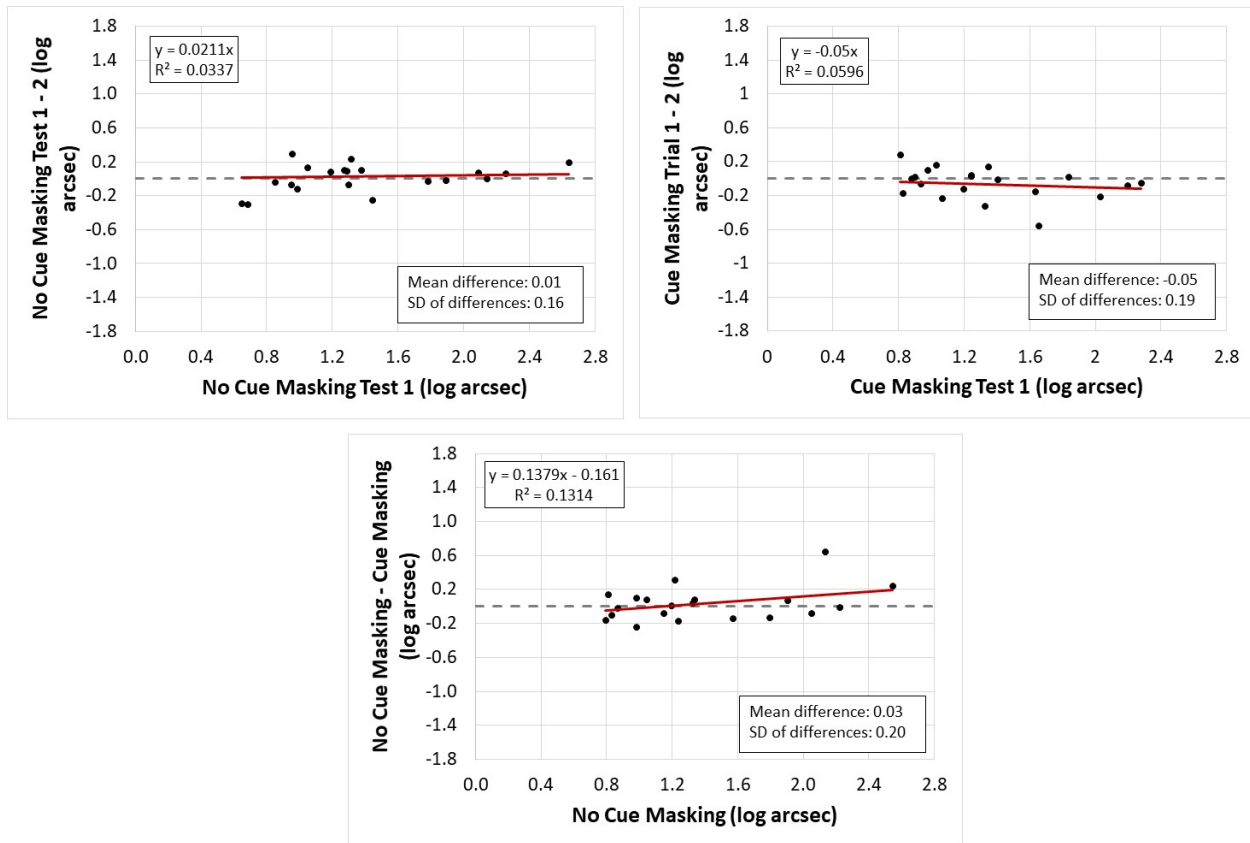


Figure 26. Test-retest differences for the SST. Top left: no monocular cue masking test 1 and test 2; top right: with monocular cue masking test 1 and test 2; bottom: without masking and with masking. The average of the differences and SD of the differences are also shown for each pair of stereo tests.

Figure 27 shows the correlations between the dual ring tests (2-, 4-, and 8-second stimulus presentation) and the SST (with and without monocular cue masking) for 10 subjects. All dual ring tests showed good correlations: 2-second and SST with cue masking ($r = 0.98$, $p < 0.001$), 2-second and SST with no cue masking ($r = 0.82$, $p < 0.001$), 4-second and SST with cue masking ($r = 0.88$, $p < 0.001$), 4-second and SST with no cue masking ($r = 0.96$, $p < 0.001$), 8-second and SST with cue masking ($r = 0.9$, $p < 0.001$), and 8-second and SST with no cue masking ($r = 0.96$, $p < 0.001$). The slope of the linear fit is clearly greater than 1 across all the conditions, indicating that subjects with good stereo acuity on the SST tend to perform very well on the dual ring task while subjects with poorer stereo acuity on the SST tend to perform even more poorly on the dual ring test.

Figure 28 shows the differences between the dual ring tests (2-, 4-, and 8-second stimulus presentation) and the SST (with and without monocular cue masking). The average of the differences and SD of the differences are also shown for each pair of stereo tests.

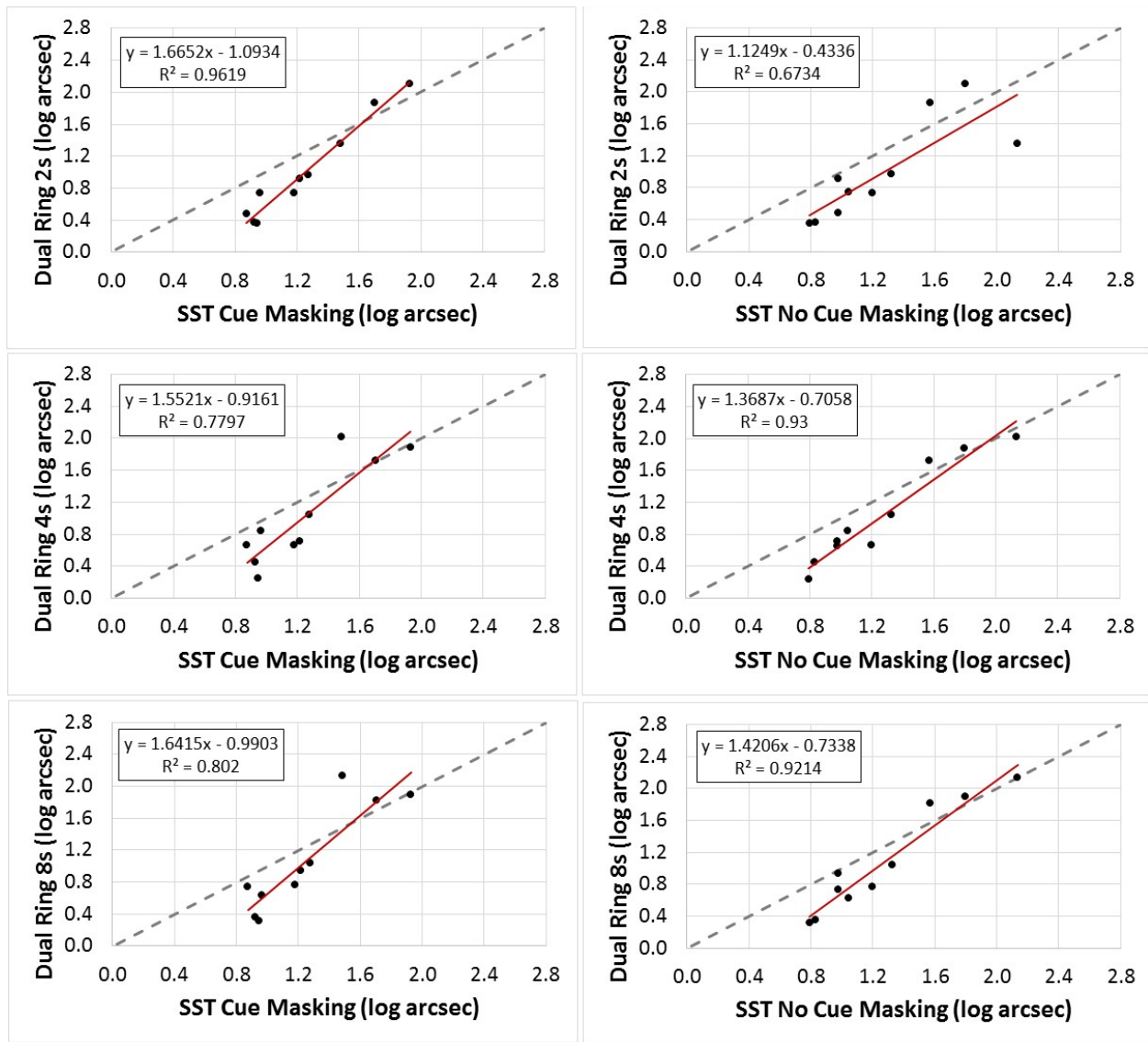


Figure 27. Scatterplots showing the correlations between SST with monocular cue masking (left column) and SST with no monocular cue masking (right column) with each of the dual ring tests (2-second stimulus duration top row, 4-second middle row, and 8-second bottom row).

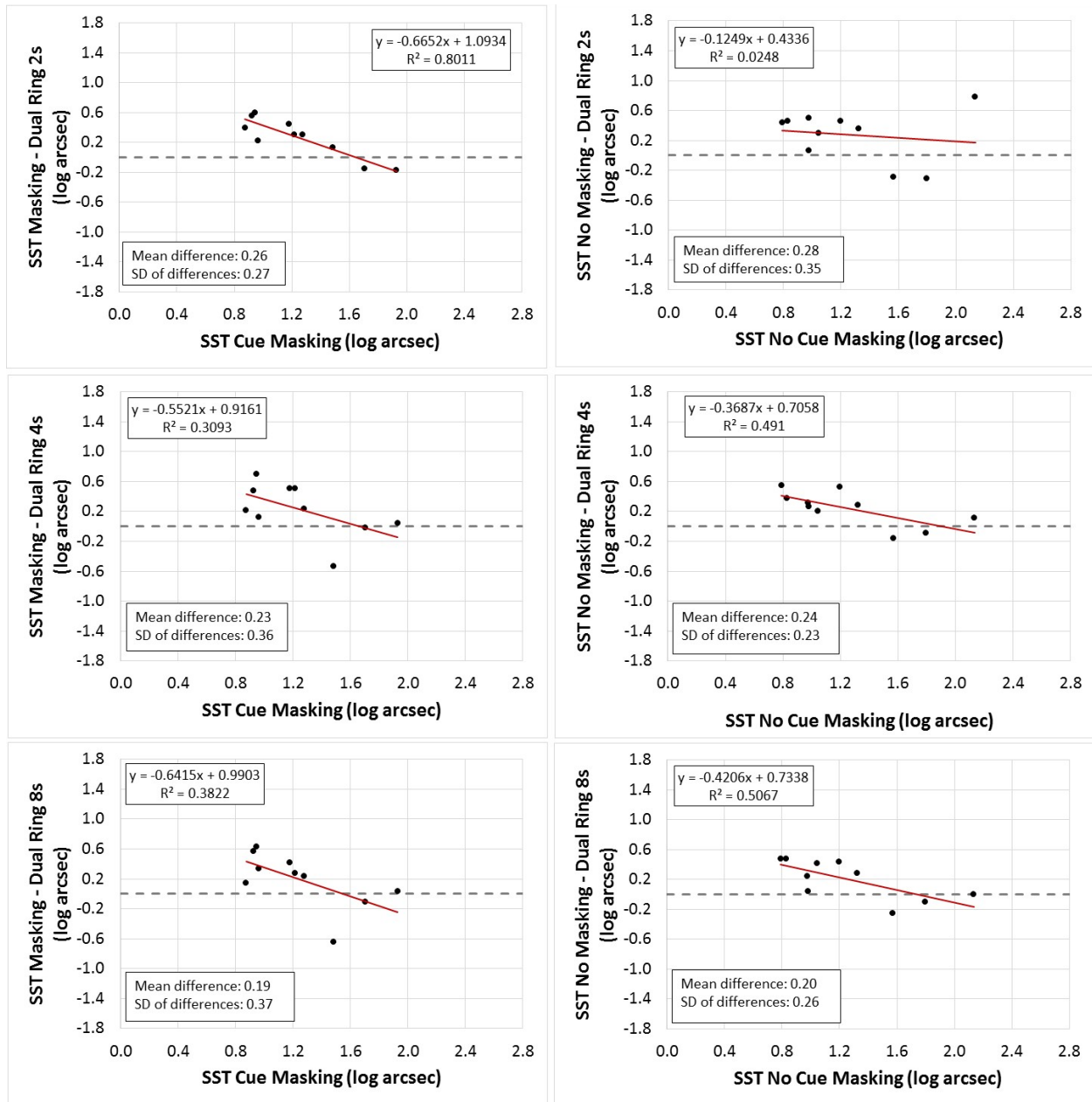


Figure 28. Scatterplots showing the differences between the SST with monocular cue masking (left column) and SST with no monocular cue masking (right column) with each of the dual ring tests (2-second stimulus duration top row, 4-second middle row, and 8-second bottom row).

The psi procedure provides an estimate of the threshold error (alpha SE) and an estimate of the slope (beta) of the psychometric function in addition to the threshold (alpha) estimate. The average alpha SE for the SST was approximately 0.16 and the slope (beta) was approximately 0.47. For comparison, the alpha SE for the dual ring (near) was 0.32 and the slope (beta) was approximately 0.29.

3.4.3 Discussion. Although the dual ring and SST are highly correlated as shown in Figures 27 and 28, the test-retest reliability of the SST is higher and may be less subject to learning effects, particularly for subjects with relatively poor stereo acuity. The SST with/without monocular cue masking test-retest differences is smaller than the differences between test repetitions for the dual ring test. The SST with cue masking is most similar to the dual ring test with a stimulus presentation time of 2 seconds. Comparing these two stereo tests reveals that there is a mean difference of -0.05 for the SST vs. 0.14 for the dual ring test. Further, the slope of the fit for the SST is near zero (-0.13), while the slope is clearly non-zero for the dual ring test (0.4). Finally, the SD of test 1 vs. test 2 scores is clearly smaller for the SST test (0.16 for SST with cue masking vs. 0.38 for the 2-second dual ring test). Further, the estimated threshold error (alpha SE) is much lower for the SST test compared to the dual ring test, and the estimated slope of the psychometric function was not as shallow.

Future research involving operational task performance will include both the dual ring test and SST to determine which test may be more predictive of operational performance requiring depth perception (e.g., simulated remote vision system air refueling). However, the results of this stereo test comparison indicate that the SST may provide better test-retest reliability and so may be preferable to the dual ring stereo test.

3.5 Fusion Range Test

The fusion range test estimates the ability of individuals to maintain a single fused image in the presence of either horizontal or vertical deviations of the target stimulus in the left/right eye images. Figure 29 shows the horizontal fusion range stimuli (left and right eye images). To maintain a fused image, subjects must cross/uncross their eyes while maintaining a point of accommodation at the display distance. For vertical fusion range, the left/right eye images are displaced in the vertical direction, requiring the eyes to separate in the vertical direction to maintain a single fused image.



Figure 29. Horizontal fusion range stimuli (left and right eye images).

3.5.1 Methods. The stimuli were generated using a Dell Precision T7610 with Nvidia GeForce GTX 680 graphics card and displayed on an Asus VG278HE 3D monitor with 1920 x 1080 pixels that was compatible with Nvidia 3D Vision2 using active shutter glasses. Subjects were instructed to indicate when the circle became doubled (“breaks”) using the green button on the game pad as the circles moved apart. When the direction reversed, subjects indicated when the circles return to a single “fused” image by pressing the green button a second time. This task was repeated two times for each of four directions (horizontal crossed, horizontal uncrossed, vertical left eye up, vertical right eye up).

3.5.2 Results. Figure 30 shows the distribution of horizontal fusion range scores in log arcmin for 93 subjects. Horizontal fusion range is the sum of crossed and uncrossed fusion recovery. As shown, there is a substantial amount of variation in fusion range across individuals – with scores spanning nearly 3 log units.

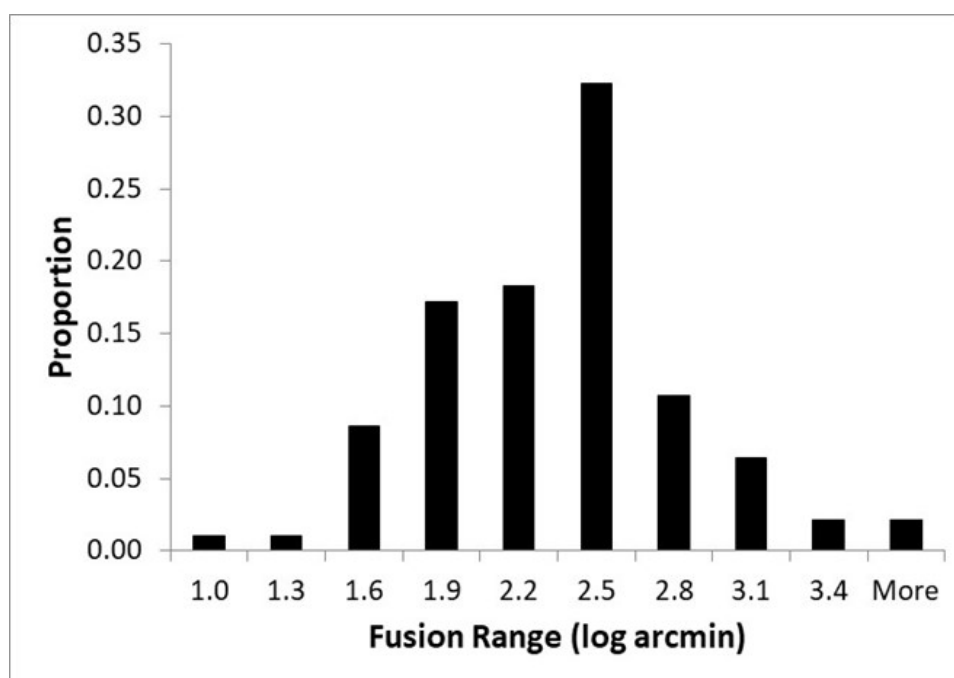


Figure 30. Distribution of horizontal fusion range scores in log arcmin.

Figure 31 shows the relationship between test and retest results for horizontal fusion range for 93 subjects. Although this test relied on the subjective judgment of subjects concerning blur/diplopia, test-retest reliability was quite good ($r = 0.86$, $p < 0.001$).

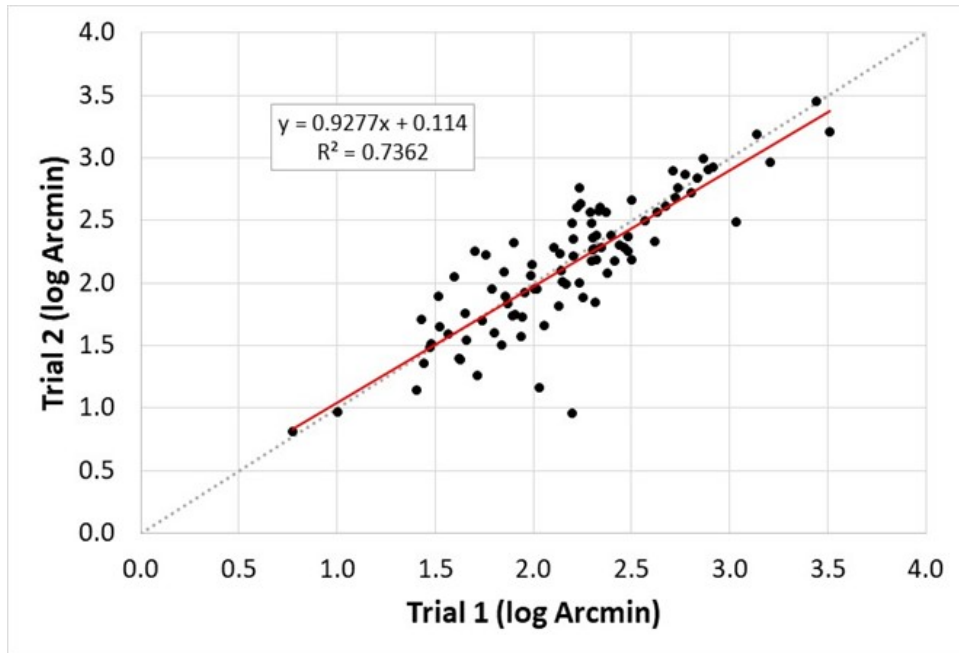


Figure 31. Scatterplot showing repeated test results for the horizontal fusion range test.

Figure 32 shows the distribution of vertical fusion range scores in log arcmin for 87 subjects. As shown, the range of scores for vertical fusion range is much narrower in comparison to horizontal fusion range, spanning only approximately 0.5 log units. However, a few subjects had a vertical fusion range less than log 1.0 arcmin as shown in Figure 33.

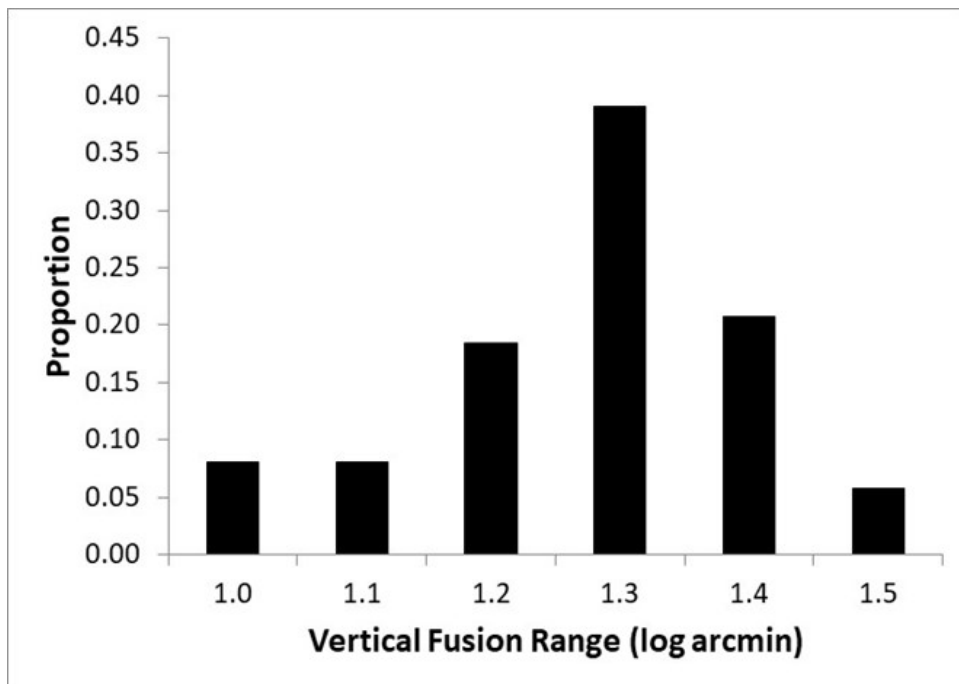


Figure 32. Distribution of vertical fusion range scores in log arcmin for 93 subjects.

Figure 33 shows the relationship between test and retest results for vertical fusion range for 87 subjects. Test-retest reliability for vertical fusion range was also quite good ($r = 0.78$, $p < 0.001$).

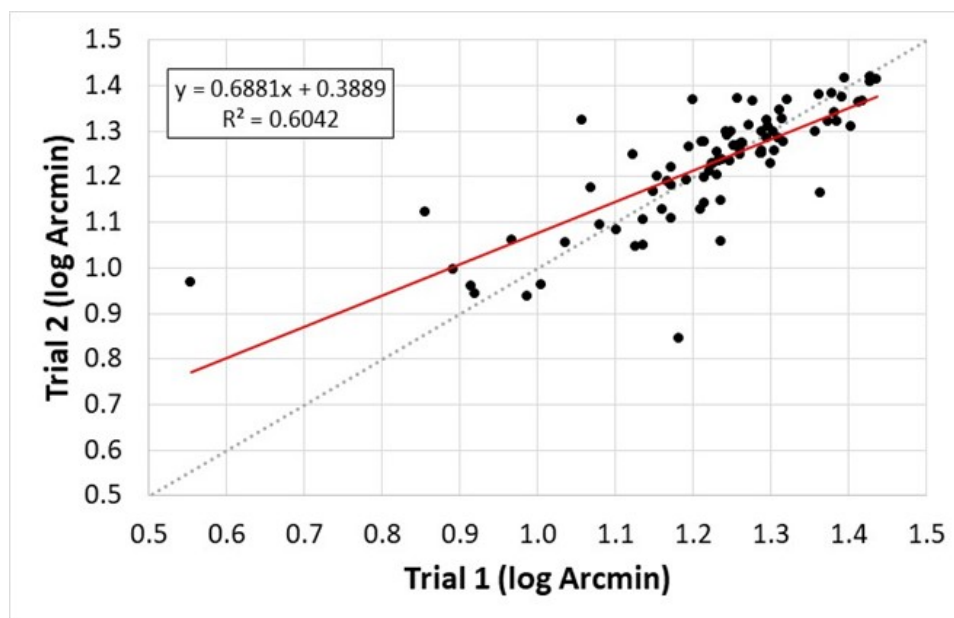


Figure 33. Scatterplot showing repeated test results for the vertical fusion range test.

Figure 34 shows the differences between test-retest for horizontal (left) and vertical fusion range (right). The average of the differences and SD of the differences are also shown for each pair of fusion range tests.

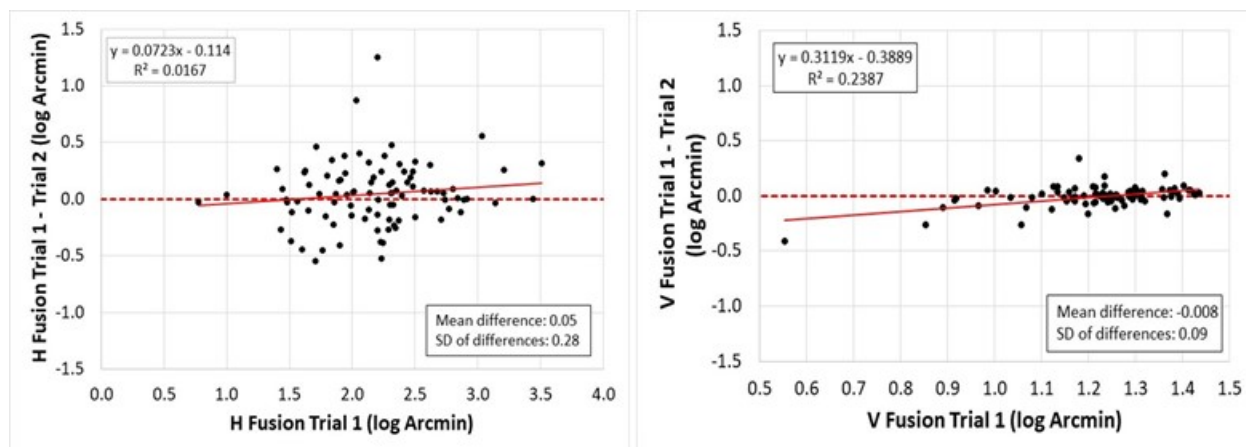


Figure 34. Differences between test-retest for horizontal (left) and vertical fusion range (right). The average of the differences and SD of the differences are also shown for each pair of fusion range tests.

3.6 OBVA Motion Coherence Test

Not only are current vision standards based on static, paper charts dating from the 1940's era, they are limited to standard measures such as visual acuity, phorias, and coarse measures of stereo acuity. A more comprehensive measure of spatial vision, contrast sensitivity, is not part of the existing standard, and motion perception is not tested at all. Computer-based vision tests enable precisely calibrated luminance and color contrast, sophisticated threshold estimation procedures, as well as tests that could use dynamic stimuli. U.S. Air Force pilots and other Airmen perform complex tasks in highly dynamic environments. If it can be shown that individuals vary reliably in sensitivity to motion, a motion test could be a relevant screening test. Furthermore, previous research provides some evidence that sensitivity to motion is predictive of operational performance [10-12]. For this reason, OBVA is also investigating the utility of including a motion perception test for the AVT.

3.6.1 Methods. Subjects viewed a field of dots presented very briefly as shown in Figure 35. A few of the dots moved in one of four directions: 1) clockwise, 2) counterclockwise, 3) expanding (flowing outward toward the edges of the display), or 4) flowing inward toward the center of the display. Subjects were instructed to use the keyboard arrow keys to enter their response (up arrow key for expansion, down arrow key for contraction, right arrow key for clockwise motion, or left arrow key for counterclockwise motion). Auditory feedback was provided. Motion coherence, or the proportion of dots moving in a particular direction, was varied according to the psi adaptive threshold estimation procedure. The motion coherence test was designed based on previous research [13].

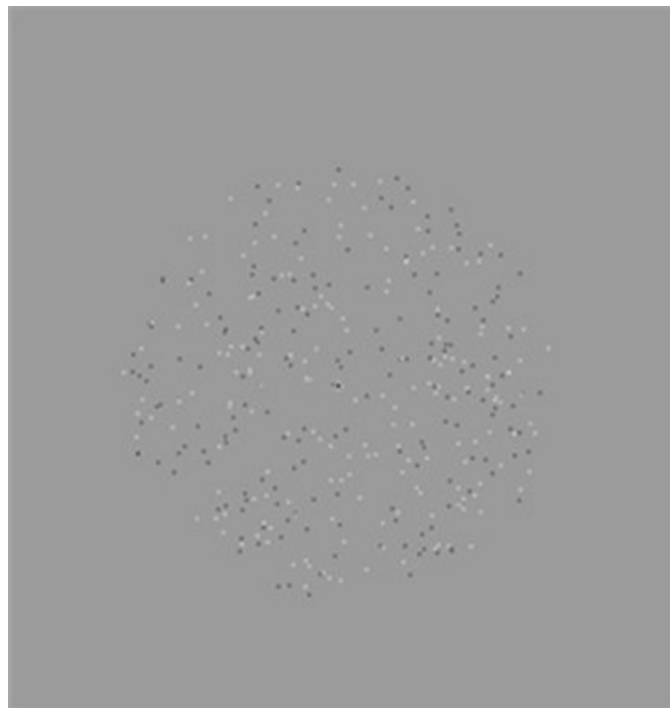


Figure 35. Motion sensitivity test stimuli.

3.6.2 Results. Because very little data exist examining sensitivity to motion, establishing normative data for the OBVA motion test is of particular interest. Figure 36 shows histograms for rotational (left) and radial (right) motion for 55 subjects using log motion coherence thresholds.

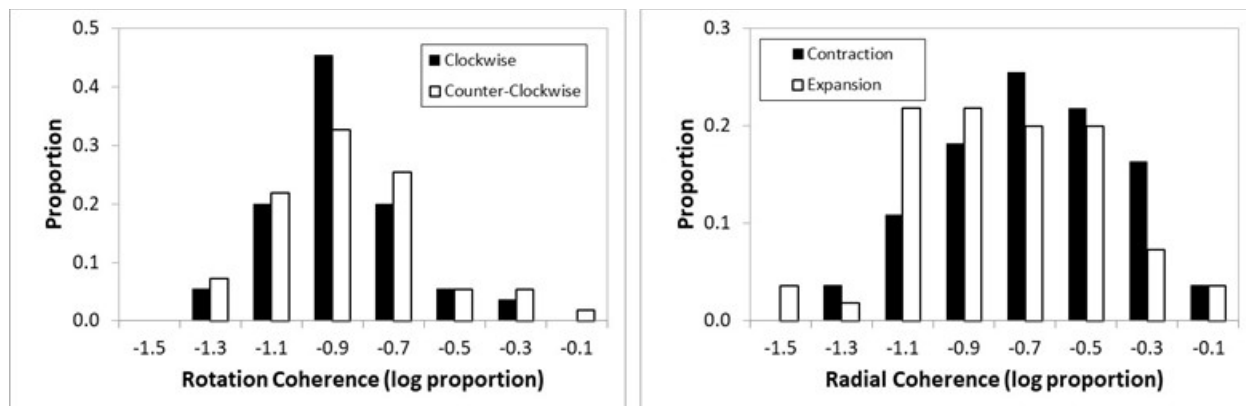


Figure 36. Histograms for rotational (left) and radial (right) motion for 55 subjects.

Figure 37 shows test-retest reliability for clockwise (top left), counterclockwise (top right), expansion (bottom right), and contraction (bottom left). Preliminary analysis suggests that motion coherence thresholds are highly reliable for clockwise ($r = 0.73$, $p < 0.001$) and counterclockwise rotation ($r = 0.84$, $p < 0.001$) and, although still highly significant, less reliable for contraction ($r = 0.43$, $p = 0.001$). However, test-retest reliability was particularly poor for contraction ($r = 0.15$, $p = 0.28$). Note that log proportion motion coherence thresholds are used.

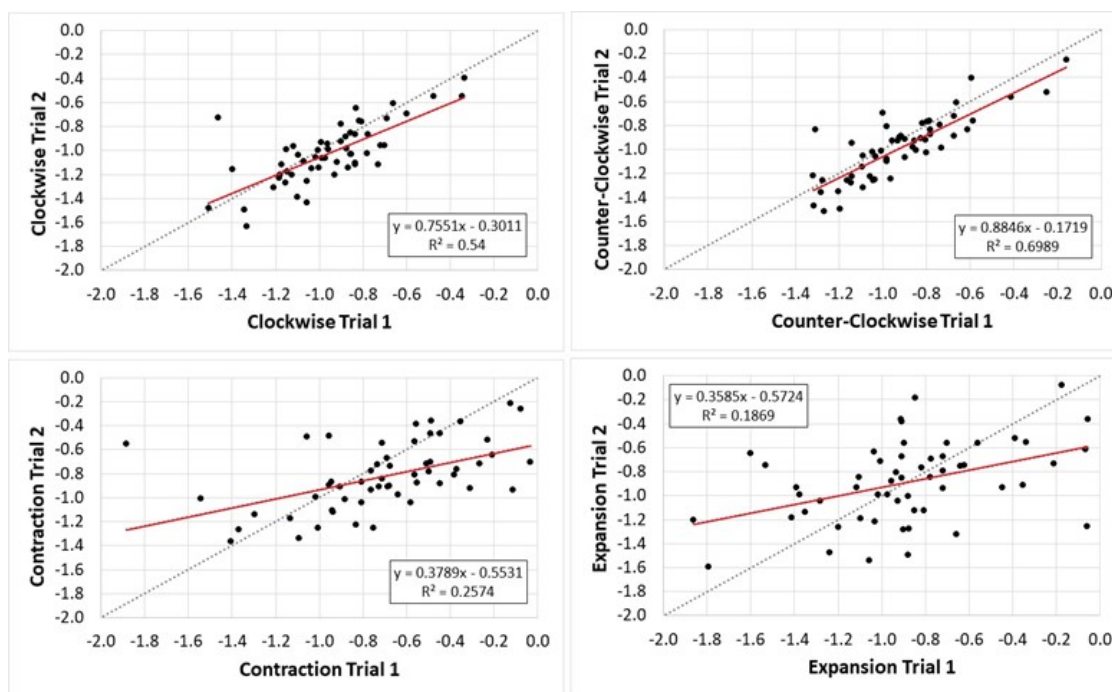


Figure 37. Scatterplots showing the relationship between repetitions of the motion perception test for clockwise (top left), counterclockwise (top right), expansion (bottom right), and contraction (bottom left).

Figure 38 shows the relationship between clockwise rotation and counterclockwise rotation motion thresholds (left) and between expansion and contraction thresholds (right). The correlation between clockwise and counterclockwise rotation thresholds was high ($r = 0.79$, $p < 0.001$). The correlation between expansion and contraction thresholds was not significant ($r = 0.13$, $p = 0.36$).

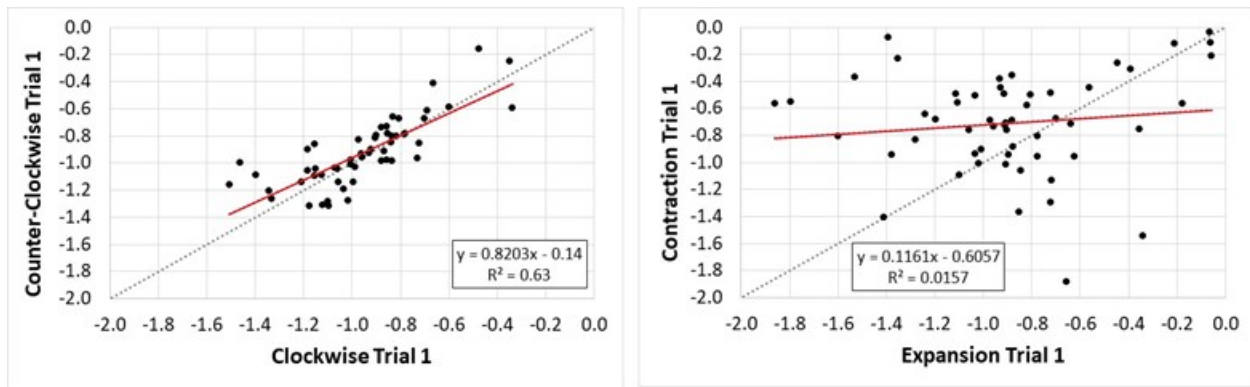


Figure 38. The relationship between clockwise rotation and counterclockwise rotation motion thresholds (left) and between expansion and contraction thresholds (right).

Figure 39 shows the relationship between clockwise rotation thresholds and contraction motion thresholds (left) and between expansion thresholds (right). The correlation between clockwise rotation and contraction was significant ($r = 0.28$, $p = 0.04$), but clearly much lower than between clockwise and counterclockwise. The correlation between clockwise and expansion thresholds was not significant ($p = 0.15$, 0.28).

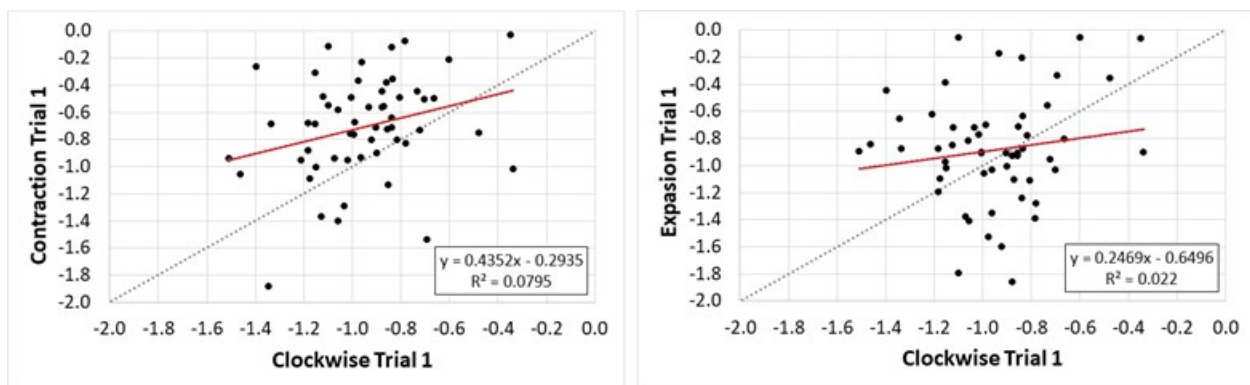


Figure 39. The relationship between clockwise rotation thresholds and contraction motion thresholds (left) and between expansion thresholds (right).

Figure 40 shows the differences between test-retest for clockwise (top left), counterclockwise (top right), expansion (bottom left), and contraction (bottom right). The average of the differences and SD of the differences are also shown for each pair of motion 1.0 tests.

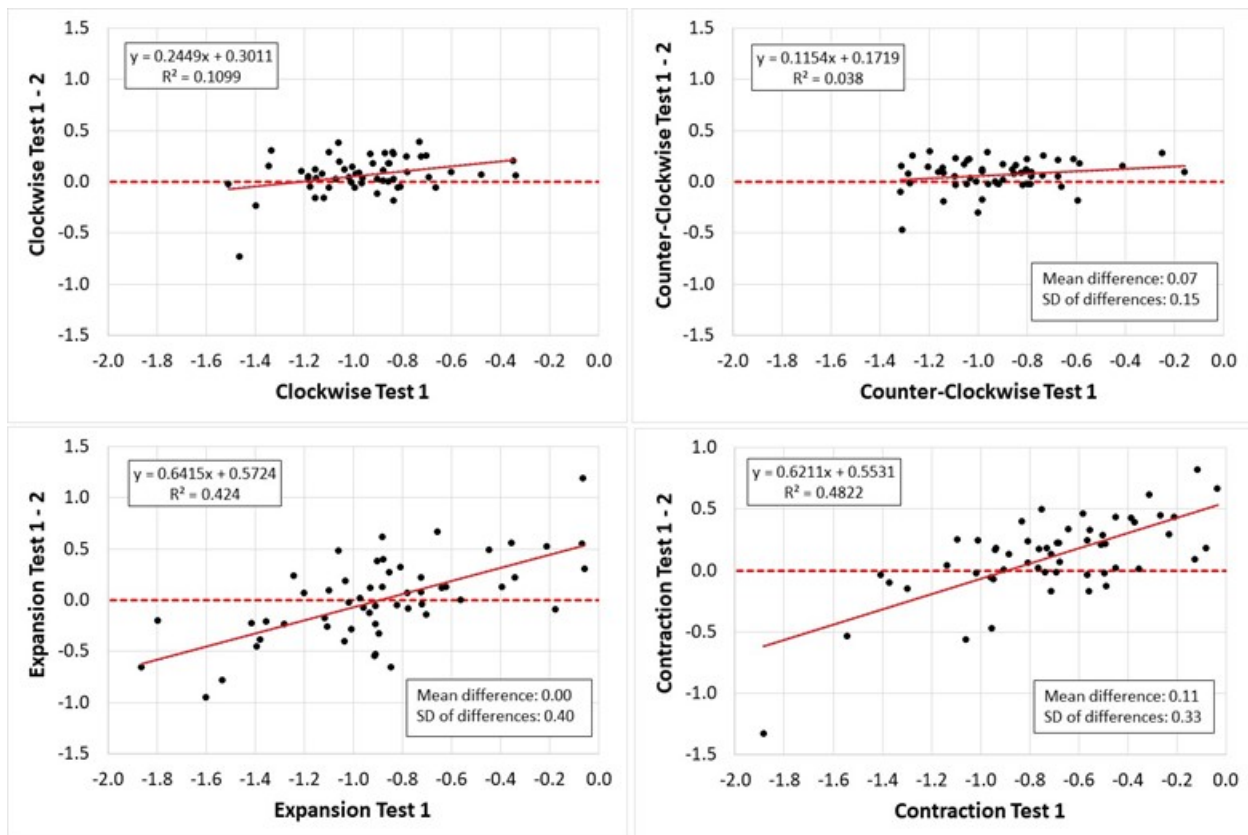


Figure 40. The differences between test-retest for clockwise (top left), counter-clockwise (top right), expansion (bottom left), and contraction (bottom right). The average of the differences and SD of the differences are also shown for each pair of motion 1.0 tests.

3.6.3 Discussion. Rotational motion coherence thresholds appear to have large differences between observers and are highly reliable. Differences between observers are about 1 log unit (10x). This is comparable to or larger than the range of visual acuity scores, and motion perception may actually be more predictive of performance in operationally relevant flight tasks. Although the correlation between repeated expansion/contraction motion coherence tests was significant, thresholds varied substantially between the two repetitions of this test. These data revealed that expansion and contraction thresholds were uncorrelated. This was an unexpected result since previous research indicates that global motion perception may be mediated by two independent mechanisms – one rotational and one radial [13]. Thus, we expected that expansion and contraction thresholds would be correlated. The test-retest data show that expansion thresholds in particular had low test-retest reliability for this implementation of the motion coherence test. Based on the variability in these results, the motion test was redesigned; the new motion test and initial results are summarized below.

3.7 OBVA Motion Coherence Test 2.0

3.7.1 Methods. The OBVA motion coherence 2.0 was redesigned to use a 2AFC task – with rotational and radial motion thresholds assessed in separate blocks of trials. Additionally, in version 1.0 both the speed and direction of the incoherent dots were randomized, whereas in version 2.0 the direction was randomized but the speed was constant and the same as the coherent dots (2 deg/s). The redesigned test has been deployed at USAFSAM and the results are summarized below. The new motion test will be provided to international partners to collect additional normative data, evaluate test-retest reliability, and examine the relationship between motion perception and operational performance.

3.7.2 Results. Figure 41 shows histograms for radial motion thresholds (black bars, $n = 30$) and rotational motion thresholds (white bars, $n = 32$). An accurate threshold could not be estimated for several subjects for radial motion, and so they were excluded.

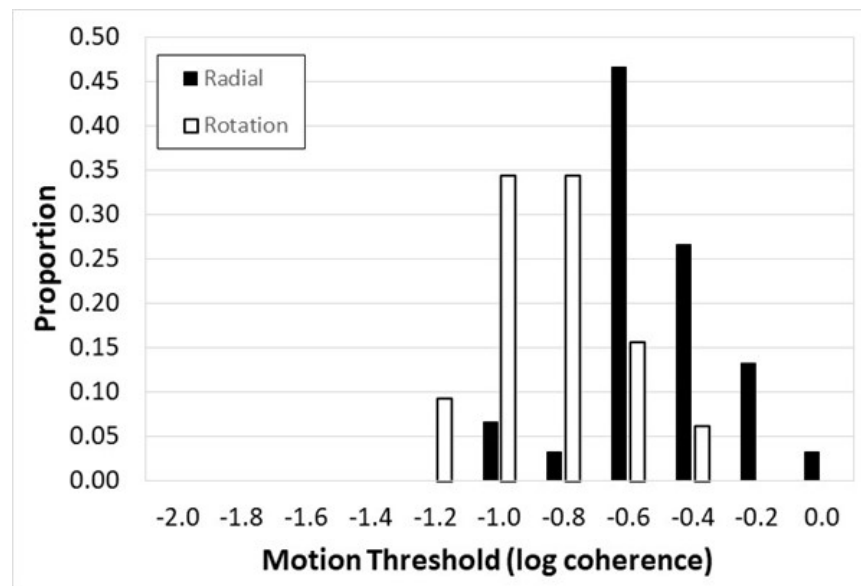


Figure 41. Histograms for radial motion 2.0 thresholds (black bars, $n = 30$) and rotational motion 2.0 thresholds (white bars, $n = 32$).

Figure 42 shows test-retest reliability for rotation (top left, $r = 0.87$, $p < 0.001$) and the relationship between rotational and radial thresholds (bottom, $r = 0.91$, $p < 0.001$).

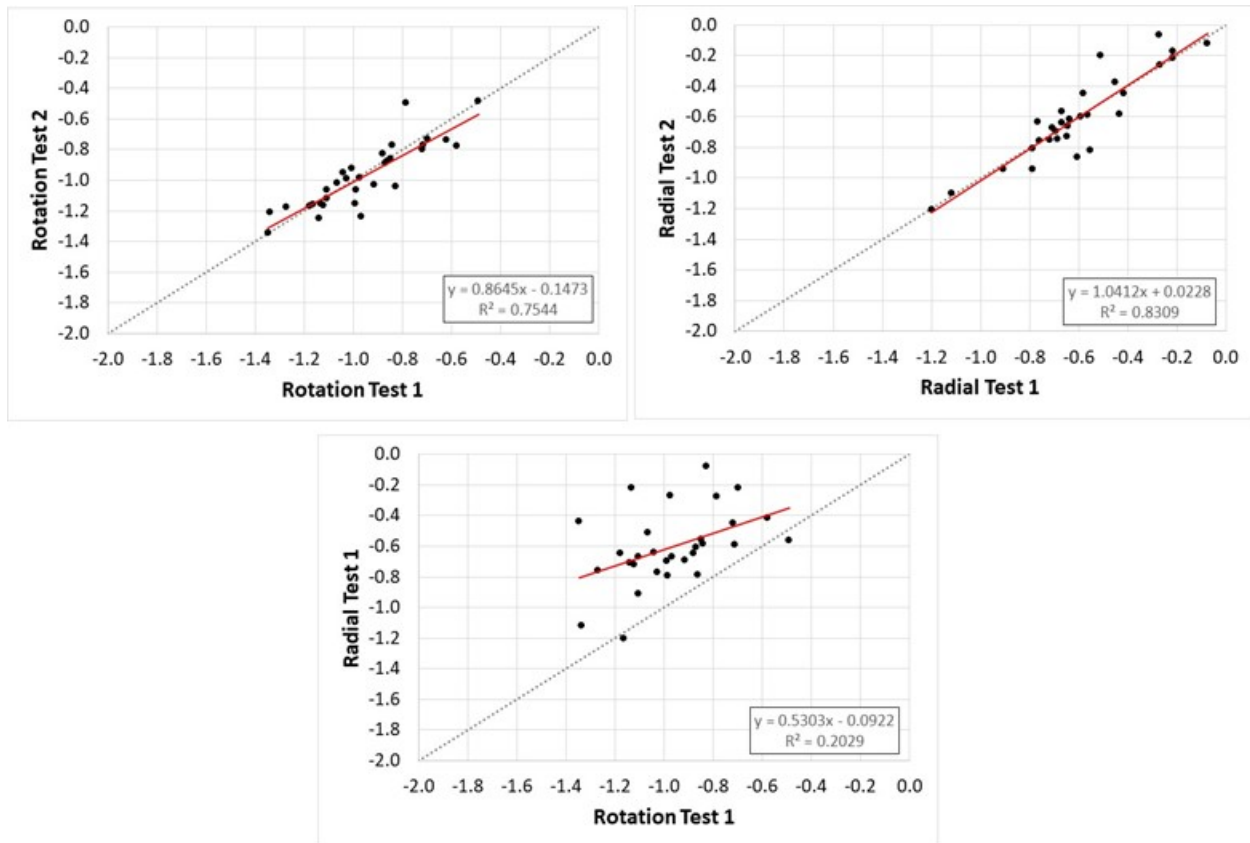


Figure 42. Test-retest reliability for rotation (top left) and radial motion (top right) and the relationship between rotational and radial thresholds (bottom).

For comparison, the average threshold for clockwise and counterclockwise directions of motion was computed for each of two test completions for OBVA motion test 1.0. Similarly, the average motion 1.0 threshold for expansion and contraction directions of motion was computed for each of two test completions. Motion 1.0 test-retest is shown in Figure 43 for rotational ($r = 0.83$, $p < 0.001$) and radial motion ($r = 0.5$, $p < 0.001$). Table 2 summarizes the mean, SD, and test-retest correlations for motion 1.0 and motion 2.0.

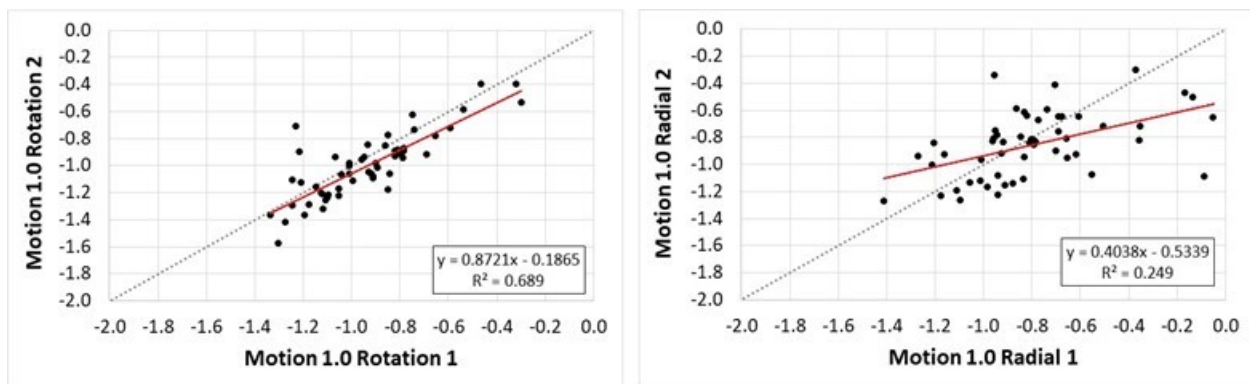


Figure 43. Motion 1.0 test-retest for average rotational (left) and radial motion (right).

Table 2. Mean, Standard Deviation, and Test-Retest Correlations for Motion 1.0 and Motion 2.0

Motion	Average Threshold		SD		Test-Retest Correlation	
	Radial	Rotation	Radial	Rotation	Radial	Rotation
1.0	-0.80	-0.94	0.29	0.23	0.50	0.83
2.0	-0.60	-0.96	0.25	0.21	0.91	0.87

Figure 44 shows the differences between test-retest for rotation (left) and radial (right) motion 2.0 thresholds. The average of the differences and SD of the differences are also shown for each pair of motion 2.0 tests.

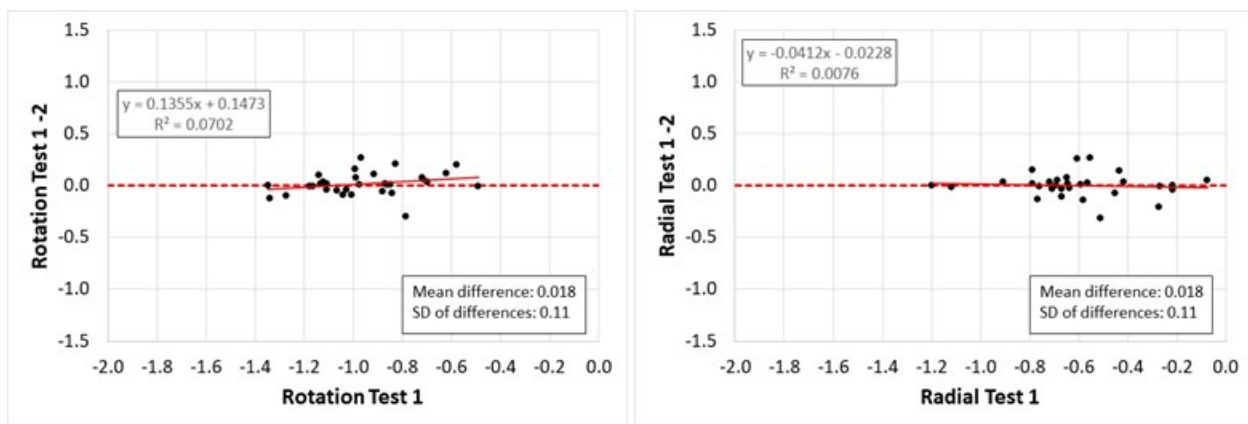


Figure 44. Differences between test-retest for rotation (left) and radial (right) motion 2.0 thresholds. The average of the differences and SD of the differences are also shown for each pair of motion 2.0 tests.

3.7.3 Discussion. Test-retest reliability was roughly comparable for rotational motion thresholds for motion 1.0 vs. motion 2.0 tests. Test-retest reliability was substantially improved for radial motion using the revised motion 2.0 test. Data collection has been discontinued for the motion 1.0 in favor of the motion 2.0 test. The OBVA Lab will continue to collect data using the motion 2.0 test to build a normative database, evaluate test-retest reliability, and examine the relationship between motion perception and operational performance. The new motion test will also be provided to international partners to support those objectives.

3.8 AVT Descriptive Statistics

Table 3 below summarizes the descriptive statistics for each of the AVTs. The mean threshold (alpha) and SD of thresholds are shown. The psi procedure also provides an estimate of the threshold SE and an estimate of the slope. The average threshold SE and slope are also shown in Table 3. Test-retest data are also shown: mean difference between test repetitions, SD of the differences, and the test-retest correlation. Finally, the number of subjects (N) that completed each test is shown. As shown, the mean threshold SE is substantially higher for the dual ring stereo test, and the slope estimates tend to be relatively shallow in comparison to the other AVTs. In addition, average SD of the threshold differences for the dual ring stereo test is also relatively large. These results indicate that the dual ring test has significantly lower reliability than the rest of the tests evaluated.

Table 3. Descriptive Statistics for Each of the AVTs

Test	Mean Threshold	SD Threshold	Mean Threshold SE	Mean Slope	Mean Threshold Difference	SD Threshold Difference	Test-Retest (r)	N
Motion 1.0								
Radial	-0.95	0.26			0.07	0.15	0.84	55
Rotational	-0.99	0.25			0.07	0.18	0.73	55
Motion 2.0								
Radial	-0.61	0.26	0.07	0.76	0.002	0.12	0.91	30
Rotational	-0.97	0.21	0.04	0.72	0.018	0.11	0.87	32
Dual Ring Stereo								
Far Stereo	1.87	0.65	0.33	0.25	0.08	0.39	0.81	90
Near Stereo	1.74	0.65	0.32	0.29	0.12	0.36	0.83	90
LCST								
Acuity (logMAR)	-0.17	0.07	0.05		0.008	0.05	0.72	62
1.25	-0.74	0.25	0.08	0.38	0.006	0.15	0.81	66
2.5	-1.25	0.19	0.07	0.44	0.011	0.12	0.84	66
16.67	-1.95	0.10	0.06	0.48	0.011	0.08	0.71	66
OCCT								
L cone contrast	-2.08	0.30	0.07	0.45	0.05	0.11	0.93	69
M cone contrast	-2.05	0.30	0.07	0.45	0.04	0.09	0.96	69
S cone contrast	-1.10	0.17	0.08	0.37	0.09	0.14	0.65	69
Fusion Range								
Horizontal	2.18	0.49	N/A	N/A	0.05	0.28	0.86	93
Vertical	1.22	0.15	N/A	N/A	-0.008	0.09	0.78	87

4.0 GENERAL DISCUSSION

The USAFSAM OBVA Laboratory developed the computer-based AVTs described here to support threshold-level vision testing capability needed to investigate the relationship between ocular health and operationally relevant performance. It is clear that many existing standard tests simply do not support the level of accuracy and repeatability to support correlation analysis. Furthermore, more precise and repeatable vision tests are required to support U.S. Air Force Surgeon General objectives for enhanced health diagnostics and operational medicine. The AVT research grade tests also support interservice, international, industry, and academic research collaboration. The AVT software has been provided to key partners to support ongoing research collaboration, and future AVT reports will include data from our partners to develop a comprehensive normative database for the AVT. The OBVA Laboratory is also pursuing commercialization of the AVT through the Small Business Innovative Research program. The AVTs described here provide the basis for a new generation of more accurate and reliable, commercially available, computer-based vision tests to support earlier diagnosis of eye disease, more accurate monitoring of treatment and recovery, and occupational vision screening. Development of more accurate vision tests to replace 1940's era vision tests is long overdue. "Present military visual standards have existed with little real change since WWII. The design of instruments used to measure visual acuity, color vision, and muscle balance in military clinical settings remains unchanged since the original purchases over 40 years ago" [14]. Over 30 years have passed since those authors identified the need to replace the already antiquated vision testing methods. Today, the need is even greater to move beyond simply identifying the presence

or absence of disease, particularly as calls for truly occupational-/operational-based medical standards and issues concerning human rights, discrimination, and requests for exception to policy are becoming much more prominent.

Further, data obtained with AVT can enable research examining individual differences in visual ability and provide data to examine the relationships between different visual functions. A number of studies attempting to identify the key aspects of vision underlying overall visual capability have been published in recent years [15-19]. However, each study generally uses a unique set of tests, many of which are unlikely to be used for routine vision screening applications (e.g., complex forms test, glass patterns, temporal order judgment, bisection discrimination, etc.). The AVT, with tests deployed across U.S. military services and international, academic, and industry partners to generate a large normative database validated against operational task performance, may provide an opportunity to develop an accepted set of computer-based vision tests that can be used to modernize vision screening and can also support research into the aspects of vision underlying visual capability.

5.0 REFERENCES

1. Gaska J, Winterbottom M, van Atta A. Operational based vision assessment cone contrast test: description and operation. Wright-Patterson AFB (OH): U.S. Air Force School of Aerospace Medicine; 2016. Report No. AFRL-SA-WP-SR-2016-0007. [Accessed 1 Mar 2018]. Available from <http://www.dtic.mil/docs/citations/ADA635957>.
2. Shoda E, van Atta A, Winterbottom M, Gaska J. Operational based vision assessment automated vision test collection user guide. Wright-Patterson AFB (OH): U.S. Air Force School of Aerospace Medicine; 2017. Report No. AFRL-SA-WP-SR-2017-0012. [Accessed 1 Mar 2018]. Available from <http://www.dtic.mil/docs/citations/AD1034796>.
3. Prins N, Kingdom FA. Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Front Psychol*. 2018; 9:1250.
4. Rabin J, Gooch J, Ivan D. Rapid quantification of color vision: the cone contrast test. *Invest Ophthalmol Vis Sci*. 2011; 52(2):816-820.
5. Cooper J, Warshowsky J. Lateral displacement as a response cue in the Titmus Stereo test. *Am J Optom Physiol Opt*. 1977; 54(8):537-541.
6. Fawcett SL, Birch EE. Validity of the Titmus and Randot circles tasks in children with known binocular vision disorders. *J AAPOS*. 2003; 7(5):333-338.
7. Fielder AR, Moseley MJ. Does stereopsis matter in humans? *Eye (Lond)*. 1996; 10(Pt 2):233-238.
8. Fricke TR, Siderov J. Stereopsis, stereotests, and their relation to vision screening and clinical practice. *Clin Exp Optom*. 1997; 80(5):165-172.
9. Bach M, Schmitt C, Kromeier M, Kommerell G. The Freiburg Stereoacuity Test: automatic measurement of stereo threshold. *Graefes Arch Clin Exp Ophthalmol*. 2001; 239(8):562-566.
10. Kruk R, Regan D. Visual test results compared with flying performance in telemetry-tracked aircraft. *Aviat Space Environ Med*. 1983; 54(10):906-911.
11. Kruk R, Regan D, Beverley KI, Longridge T. Correlations between visual test results and flying performance on the advanced simulator for pilot training (ASPT). *Aviat Space Environ Med*. 1981; 52(8):455-460.

12. Wilkins L, Gray R, Gaska J, Winterbottom M. Motion perception and driving: predicting performance through testing and shortening braking reaction times through training. *Invest Ophthalmol Vis Sci*. 2013; 54(13):8364-8374.
13. Morrone MC, Burr DC, Di Pietro S, Stefanelli MA. Cardinal directions for visual optic flow. *Curr Biol*. 1999; 9(14):763-766.
14. Moffitt K, Genco LV. Criteria for a state-of-the-art vision test system. Wright-Patterson AFB (OH): Air Force Aerospace Medical Research Laboratory; 1985. Technical Report No. AFAMRL-TR-85-004. [Accessed 1 Mar 2018]. Available from <http://www.dtic.mil/docs/citations/ADA157099>.
15. Bargary G, Bosten JM, Goodbourn PT, Lawrance-Owen AJ, Hogg RE, Mollon JD. Individual differences in human eye movements: An oculomotor signature? *Vision Res*. 2017; 141:157-169.
16. Bosten JM, Goodbourn PT, Bargary G, Verhallen RJ, Lawrance-Owen AJ, et al. An exploratory factor analysis of visual performance in a large population. *Vision Res*. 2017; 141:303-316.
17. Bosten JM, Goodbourn PT, Lawrance-Owen AJ, Bargary G, Hogg RE, Mollon JD. A population study of binocular function. *Vision Res*. 2015; 110(Pt A):34-50.
18. Cappe C, Clarke A, Mohr C, Herzog MH. Is there a common factor for vision? *J Vis*. 2014; 14(8):4.
19. Halpern SD, Andrews TJ, Purves D. Interindividual variation in human visual performance. *J Cogn Neurosci*. 1999; 11(5):521-534.

LIST OF ABBREVIATIONS AND ACRONYMS

AFVT	Armed Forces Vision Tester
AOV	American Optical vectrograph
AST	Adaptive Sensory Technology
AVT	automated vision test
L	long wavelength
LAUC	log area under the curve
LCST	Landolt C contrast sensitivity test
M	middle wavelength
OBVA	Operational Based Vision Assessment
OCCT	OBVA cone contrast test
S	short wavelength
SD	standard deviation
SE	standard error
SST	stereo search test
SV	Precision Vision Rabin Super Vision
USAFSAM	U.S. Air Force School of Aerospace Medicine