

# Characterizing Fixational Eye Motion Variance Over Time as Recorded by the Tracking Scanning Laser Ophthalmoscope

Shivany Y. Condor Montes<sup>1,\*</sup>, Daniel Bennett<sup>1,\*</sup>, Ethan Bensinger<sup>2,3</sup>,  
Lakshmisahithi Rani<sup>1</sup>, Younes Sherkat<sup>4</sup>, Chao Zhao<sup>1</sup>, Zachary Helft<sup>5</sup>, Austin Roorda<sup>3,2</sup>,  
Ari J. Green<sup>1,6</sup>, and Christy K. Sheehy<sup>1,5</sup>

<sup>1</sup> University of California - San Francisco, Department of Neurology, San Francisco, CA, USA

<sup>2</sup> University of California Berkeley, Vision Science Graduate Group, Berkeley, CA, USA

<sup>3</sup> University of California Berkeley, School of Optometry, Berkeley, CA, USA

<sup>4</sup> University of California Berkeley, College of Engineering, Berkeley, CA, USA

<sup>5</sup> C. Light Technologies, Inc. Berkeley, CA, USA

<sup>6</sup> University of California - San Francisco, Department of Ophthalmology, San Francisco, CA, USA

**Correspondence:** Christy K. Sheehy, UCSF Department of Neurology – Green Lab, 675 Nelson Rising Lane, San Francisco, CA 94158, USA. e-mail: [christy.k.sheehy@gmail.com](mailto:christy.k.sheehy@gmail.com)

**Received:** October 17, 2021

**Accepted:** January 5, 2022

**Published:** February 24, 2022

**Keywords:** fixational eye motion; microsaccades; retinal eye-tracking

**Citation:** Condor Montes SY, Bennett D, Bensinger E, Rani L, Sherkat Y, Zhao C, Helft Z, Roorda A, Green AJ, Sheehy CK. Characterizing fixational eye motion variance over time as recorded by the tracking scanning laser ophthalmoscope. *Transl Vis Sci Technol.* 2022;11(2):35. <https://doi.org/10.1167/tvst.11.2.35>

**Purpose:** The purpose of this study was to characterize the benign biological variance of fixational microsaccades in a control population using a tracking scanning laser ophthalmoscope (TSLO), accounting for machine accuracy and precision, to determine ideal testing conditions to detect pathologic change in fixational eye motion (FEM).

**Methods:** We quantified the accuracy and precision of the TSLO, analyzing measurements made by three operators on a model eye. Repeated, 10-second retinal motion traces were then recorded in 17 controls, 3 times a day (morning, afternoon, and evening), on 3 separate days. Microsaccade metrics (MMs) of frequency, average amplitude, peak velocity, and peak acceleration were extracted. Trace to trace, interday, and intraday variability were calculated across all subjects.

**Results:** Intra-operator and machine variation contributed minimally to total variation, with only 0.007% and 0.14% contribution for frequency and amplitude respectively. Bias was detected, with lower accuracy for higher amplitudes. Participants had an average (SD) microsaccade frequency of 0.84 Hz (0.52 Hz), amplitude of 0.32 degrees (0.11 degrees), peak velocity of 43.68 degrees/s (14.02 degrees/s), and peak acceleration of 13,920.04 degrees/s<sup>2</sup> (4,186.84 degrees/s<sup>2</sup>). The first trace recorded within a session significantly differed from the second two in both microsaccade acceleration and velocity ( $P < 0.05$ ), and frequency was 0.098 Hz higher in the evenings ( $P < 0.05$ ). There was no MM difference between days and no evidence of a session-level learning effect ( $P > 0.05$ ).

**Conclusions:** The TSLO is both accurate and precise. However, biological inter- and intra-individual variance is present. Trace to trace variability and time of day should be accounted for to optimize detection of pathologic change.

## Introduction

Fixational eye motion (FEM) describes the involuntary movements of the eye that occur while trying to keep one's gaze as stable as possible. FEM is generally categorized into three major categories: drift (a low frequency, random walk), microsaccades (small, jerk-like motions), and tremor (a high frequency, repeti-

tive motion superimposed upon drift).<sup>1</sup> Early literature has shown that fixational eye motion helped to prevent the perception of a stationary image from fading away from view, by consistently providing a neural refresh.<sup>2–5</sup> In more recent studies, evidence suggests microsaccades have a much larger role in visual perception, modulating neural activity in cortical regions, enhancing the resolution of high spatial frequencies, and assisting in the neural processing of visual

information.<sup>4</sup> Further, in clinical research studies, measures of FEM, specifically microsaccades, have been used as biomarkers for central nervous system disorders.<sup>6,7</sup> However, despite nearly a century of research and advancement of eye tracking technology, the intrasubject variability of fixational movements for healthy individuals has not been adequately characterized.

Previous literature on FEM has suggested that substantial variance may exist between healthy controls (intersubject). Many factors are thought to contribute to this variance, with significant evidence suggesting an effect of age. Kosnik et al. reported greater variability in fixational spread along the horizontal meridian of motion compared to the vertical meridian in older observers.<sup>8</sup> Abadi and Gowen found a positive correlation between age and the amplitude of saccadic intrusions during fixation.<sup>9</sup> Most recently, in a study of 100 healthy controls, Sheehy et al. found that participants between 51 and 88 years of age had significantly higher microsaccadic frequency, mean amplitude (dominated by the horizontal contribution), mean velocity, and mean acceleration than those  $\leq 50$  years.<sup>6</sup> In addition to age, attention,<sup>10–13</sup> cognitive ability,<sup>14–16</sup> mental fatigue,<sup>17</sup> and external stimulus size and luminance<sup>18,19</sup> have all been reported to alter fixational stability. Although there is still no comprehensive explanation for why such a large degree of microsaccade variability is seen among healthy controls,<sup>20</sup> it is clear that substantial variance can be expected, and that observed variance is dependent on multiple factors. Another likely contributor to the current lack of understanding is that data have been collected in an ever-changing technological landscape, with various degrees of measurement error introduced by new devices and testing conditions.<sup>21</sup> This poses a significant challenge to both clinical interpretation and the development of microsaccade-based biomarkers of disease. To date, it remains unclear what quantity of change in FEM characteristics constitutes pathological change versus benign variability, or how to best control for external sources of variance. In this study, we aim to quantify microsaccade variability within a healthy control population, which ultimately requires the subtraction of any variation caused by the measurement device itself; the tracking scanning laser ophthalmoscope (TSLO).

The TSLO system is a custom-built retinal eye-tracking technology capable of micron-level tracking sensitivity.<sup>6</sup> The TSLO can record FEM as small as 0.25 arcminutes in size and has a high potential to record fixational biomarkers. However, before TSLO recordings can be clinically interpreted, both the measurement system accuracy and precision, as well as the

expected benign biologic variation of microsaccade metrics (MMs), must be independently characterized. In the present study, we quantified the sources of variability of the TSLO through a gage repeatability and reliability (R&R) analysis. Gage R&R studies are a common measurement analysis tool used in the field of engineering to determine what portion of the variability is due to machine error, intra-operator error, and/or a combination. We then quantified the TSLO's accuracy on predetermined signal inputs through a gage linearity and bias study.

Next, we characterized the reliability and repeatability of MMs recorded by the TSLO at multiple time points in a young (age  $\leq 40$ ), healthy population, and defined the variability of MMs within this cohort over a neurologically and visually stable period (1 week). Moreover, we used these data to characterize the expected variance in a healthy control population and determined a threshold above which MM variance may no longer be benign. Finally, we recommend ideal testing conditions, namely the time of day for the recording session and number of required traces in the TSLO device, to reduce inpatient variance and optimize the detection of pathological change.

## Methods

### Eye-Tracking Technology

Retinal imaging and eye-tracking were performed using a custom-built retinal eye-tracker, the TSLO. The system and the principles behind the software-based method for eye-tracking have been described previously.<sup>22,23</sup> The TSLO uses 840 nm light (50 nm bandwidth) from a superluminescent diode (SLD) (Superlum, Dublin, Ireland) to raster-scan the retina over a 5 degree  $\times$  5 degree field of view (FOV). Because individual frames are acquired pixel by pixel over time (approximately 33 ms), we exploit this “rolling, shutter-like” feature to monitor translational shifts of the retina. An out-of-plane optical design<sup>24,25</sup> limits system astigmatism, maximizes resolution, and ultimately optimizes image-based tracking accuracy down to 0.25 arcminutes (approximately 1  $\mu$ m). For operation, a resonant scanner (16 kHz) and galvo scanner (30 Hz) are placed at pupil planes to scan the eye horizontally and vertically, capturing 512  $\times$  512 pixel frames. Retinal images are collected by a photomultiplier tube (PMT) with a 75  $\mu$ m pinhole (1.46 Airy-disk diameters over a 4 mm pupil) placed at the retinal conjugate prior to the PMT to ensure confocality. A custom MATLAB, strip-based eye-tracking software tool is then used to extract eye motion.<sup>26</sup> For this study,

each frame was broken up into 16 strips and cross-correlated to a reference frame for a 480 Hz sampling frequency. For the reference frame, the offline analysis software builds a frame utilizing the motion of the retina over the entire video recording, thus creating a larger reference frame with limited motion artifacts. Frames were removed if they had a mean pixel value of less than 10 (pixel scans range from 0–255, and each frame is made up of  $512 \times 512$  pixels), indicating these are areas of blinks or lost frames. Strips were removed if they had a cross-correlation value of less than 0.8. Once passing the above criteria, each of the 16 strips was cross correlated to the reference frame, resulting in a high-fidelity record of the horizontal and vertical displacements of the eye.

## TSLO Device Analysis

We conducted a gage repeatability and reliability (GR&R) analysis using the analysis of a variance (ANOVA) method to assess the precision of the TSLO as a measurement system and identify its sources of variability.<sup>27</sup> Twelve predetermined frequency and amplitude values of sinusoidal retinal motion representing the span of human fixation were input into the custom-built model eye via a waveform generator (Table 1). The model eye was comprised of a galvo scanning mirror (Thorlabs) placed between the relay optics and the retina. Eye motion was extracted at 480 Hz for frequency analysis and 1920 Hz for amplitude

**Table 1.** Predetermined Frequency and Amplitude Signal Values for the Gage R&R and Linearity and Bias Analyses

Signal Number	Frequency (Hz)	Amplitude (Arcsec)
1	3	6
2	30	6
3	100	30
4	1	60
5	40	60
6	4	120
7	40	120
8	4	360
9	40	360
10	1	480
11	40	480
12	100	480

Each predetermined frequency value was input into the custom-built model eye via a waveform generator. Values were chosen to span the range of expected human fixation. The second frequency input at 30 Hz was discarded due to overlap with the device framerate of 29.005 Hz.

analysis. The frequency input was determined by dividing the number of times the trace transitioned across 0 by the length of the trace. Amplitude was determined by using the known input frequency to find the period of motion and taking the median amplitude of each periodic peak.

Each of the three device operators (AA, AB, and AC) recorded a single 10-second trace for each input signal trial. This trial was repeated 3 times a day on 3 different days. Between each operator, the TSLO system was fully turned off and restarted. Operators were blinded to their past results and to the results of other operators. After obtaining the sources of variation from the GR&R (Fig. 1, Fig. 2), we conducted a Linearity and Bias analysis (Fig. 3) to assess the accuracy of the TSLO.

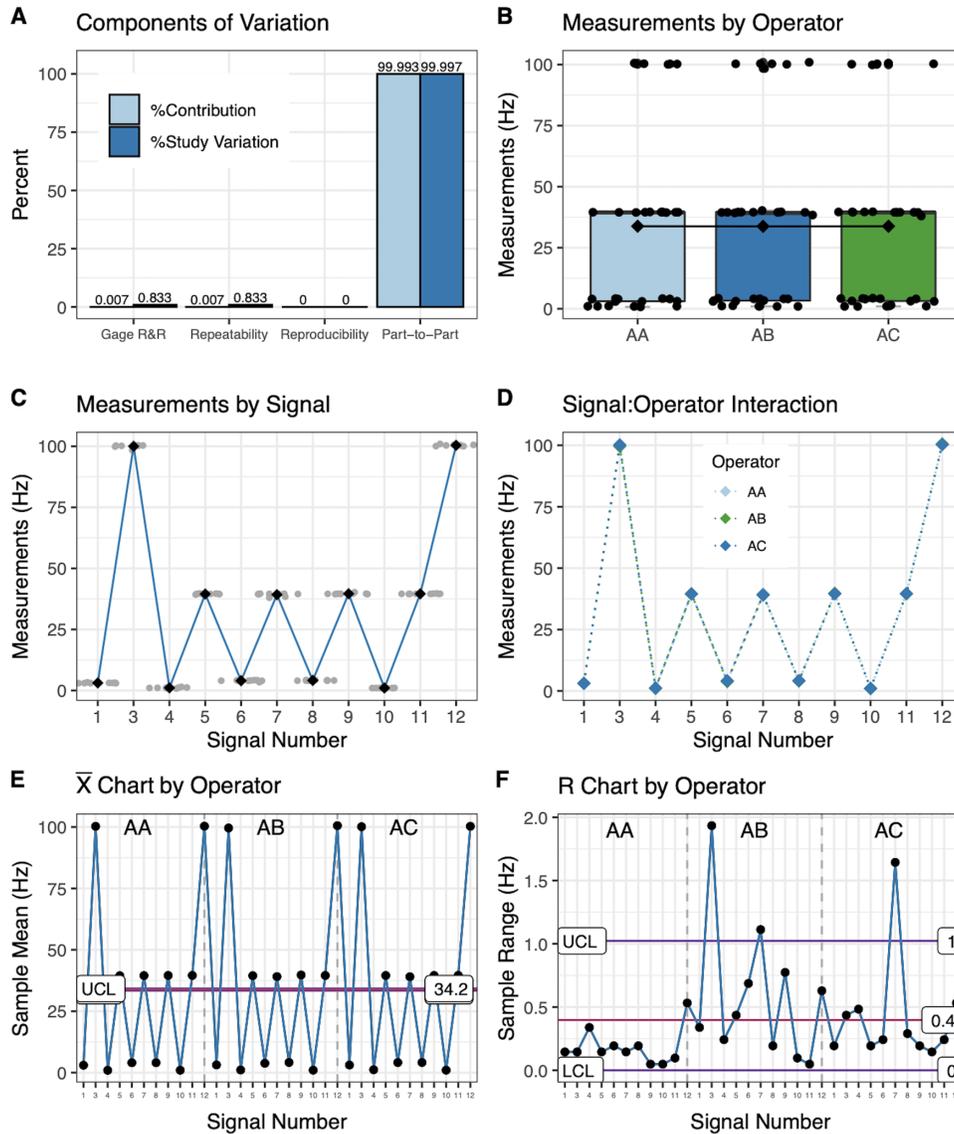
## Participants

We internally recruited 17 healthy controls (34 eyes), aged  $\leq 40$  years between March 2018 and July 2019. Inclusion criteria for recruitment were no diagnosis of retinal, ophthalmologic, or neurological disease, no retinal surgery, not pregnant, and best-corrected high-contrast visual acuity of 20/20 or better in each eye. One participant (subject 14) did not complete the testing protocol due to insufficient video quality due to post-Lasik dry eye. All participants gave written informed consent to the study. The experiment was approved by the University of California - San Francisco, Institutional Review Board, and all protocols adhered to the tenets of the Declaration of Helsinki.

## Experimental Protocol

### Healthy Control Testing

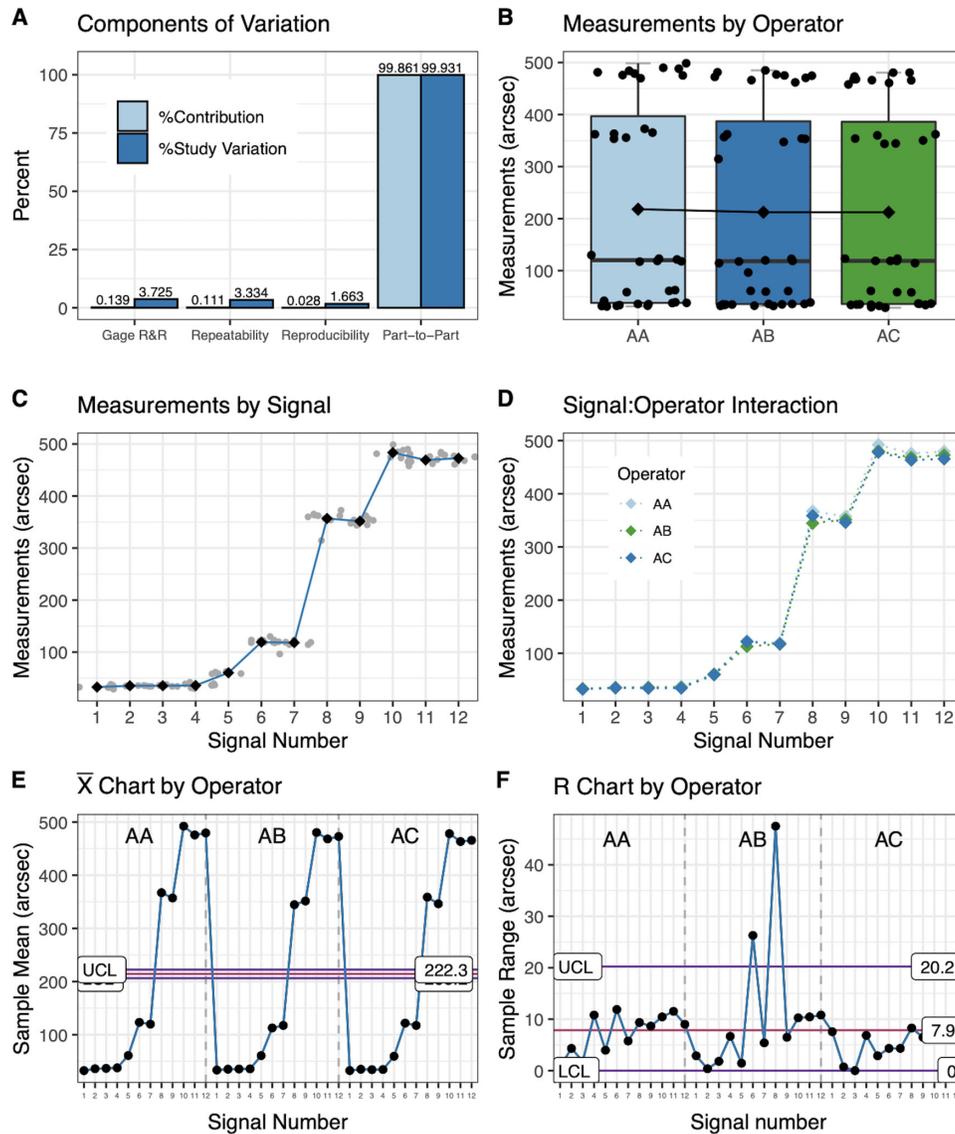
A minimum of three 10-second video recordings of fixation, hereafter referred to as traces, were acquired monocularly over 9 recording sessions. This totaled 27 traces per participant-eye and 54 traces per participant. Recording sessions were performed 3 times daily; once between 9:00 AM and 11:00 AM (session 1 – morning), once between 1:00 PM and 3:00 PM (session 2 – afternoon), and once between 4:00 PM and 6:00 PM (session 3 – evening). This daily protocol was repeated for 3 days within a single 5-day workweek, resulting in 9 visits total per participant (Fig. 4). During recording sessions, a chinrest with temple pads was utilized to limit head motion, and participants were instructed to fixate on the upper right-hand corner of the 5 degree  $\times$  5 degree infrared (IR) imaging raster. The power of the light source directed at the eye never exceeded 500  $\mu$ W.



**Figure 1. Results of Gage R&R with ANOVA method on the input frequency signals.** (A) Components of variation are part-to-part variability (defined as the variance between input signals), repeatability (inherent machine error), reproducibility (operator-to-operator error), and the sum of repeatability and reproducibility (gage R&R). %Contribution (*dark blue*): The percentage of variation due to the source compared to the total variation. %Study Variation (*light blue bars*): The percentage of variation due to the source compared to the total variation, with the added benefit of extrapolating beyond our specific data values. (B) Box plot of signal input measurements by operator. (C) Input signal number against the measured signal frequency for across all operators. (D) Input signal numbers against the measured signal frequency by operator. (E) Mean recorded measurement by signal number faceted by operators. The control limits (*purple lines*) represent the measurement system variation, and any measurements within them cannot be distinguished from random equipment error. (F) Range of recorded measurements by signal number faceted by operators. If the operators measure consistently, the points will fall within the control limits.

Our initial sample size consisted of 879 traces from 16 participants who had completed all 9 sessions. Motion traces were manually screened for quality side-by-side with their associated raw recordings by a single, trained technician (author L.R.). Criteria for further analysis were: (1) high signal strength and (2) cross-correlation values of  $\geq 0.8$  for each frame, assessed by clear visualization of retinal vasculature structure and

a motion trace matching the motion of the raw recording. A total of 704 (80%) of these traces passed manual verification. Next, participants were required to have at least two videos per eye for each session, and at least seven of the nine sessions successfully passing manual verification to be included in the final analysis. Sixty-three individual recordings, including all data for two participants (40011 and 40012) were removed, as they



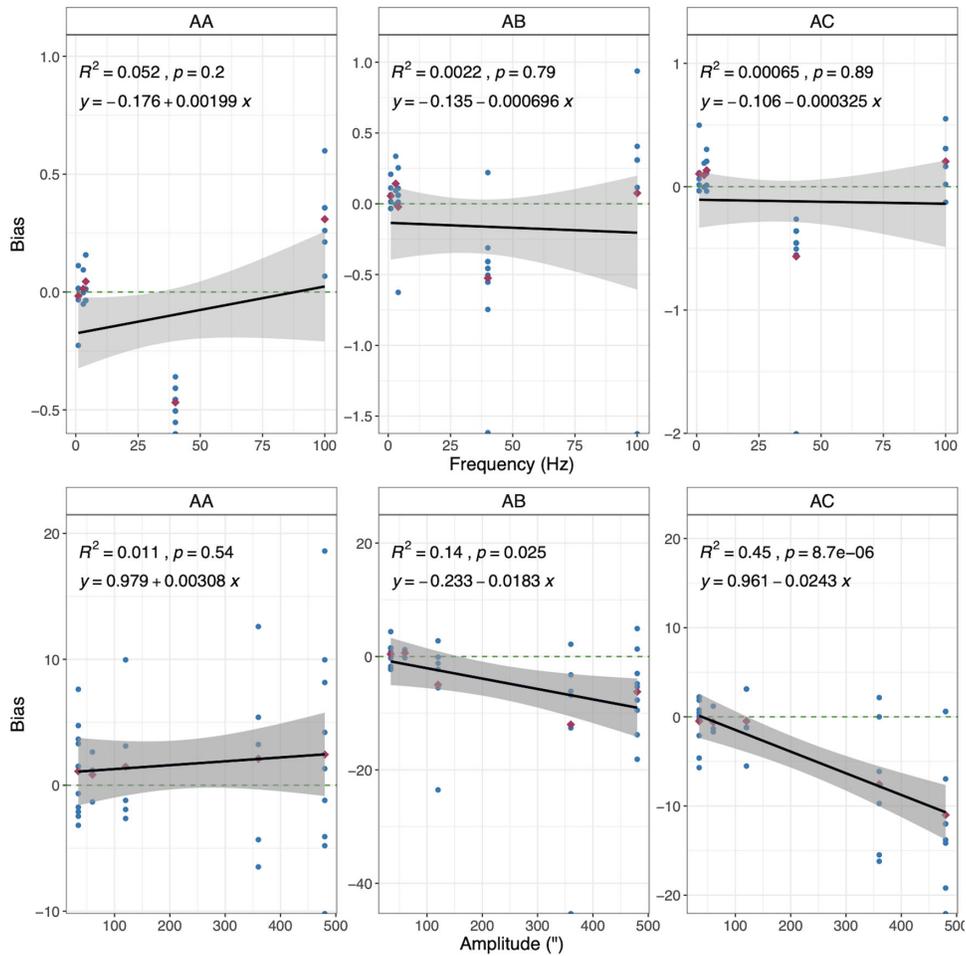
**Figure 2. Results of Gage R&R with ANOVA method on the input amplitude signals.** (A) Components of variation are part-to-part variability (variability between input signals), repeatability (inherent machine error), reproducibility (operator-to-operator error), and the sum of repeatability and reproducibility (gage R&R). %Contribution (*dark blue*): The percentage of variation due to the source compared to the total variation. %Study Variation (*light blue bars*): The percentage of variation due to the source compared to the total variation, with the added benefit of extrapolating beyond our specific data values. (B) Box plot of amplitude signal input measurements by operator. (C) Input signal number against the measured signal amplitude for across all operators. (D) Input signal numbers against the measured signal amplitude by operator. (E) Mean recorded measurement by signal number faceted by operators. The control limits (*purple lines*) represent the measurement system variation, and any measurements within them cannot be distinguished from random equipment error. (F) Range of recorded measurements by signal number faceted by operators. If the operators measure consistently, the points will fall within the control limits.

did not meet the above inclusion criteria on either eye. Consequently, our eligible sample constituted of 14 participants (13 left eyes and 14 right eyes) with 123 individual sessions and 641 traces.

**Microsaccade Extraction**

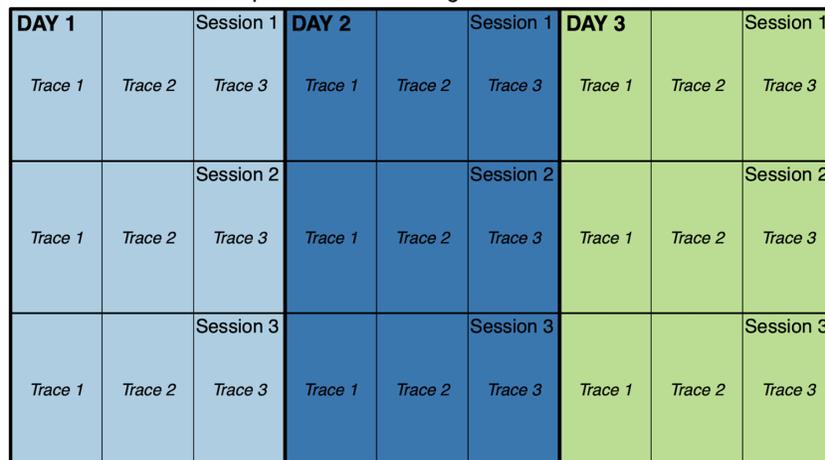
For microsaccade detection, extracted eye motion traces were smoothed using a median smooth filter (20

sample window) to remove artifacts caused by distortions in the reference frame. Next, fixational eye motion traces were flagged for the presence of microsaccades via a semi-automated custom MATLAB program utilizing a 10 degrees/s velocity threshold. Once located, the start of the saccade was marked when the velocity first exceeded 3 degrees/s and the end of the saccade was marked as 15 ms after the velocity



**Figure 3. Linearity plots for frequency and amplitude signals.** Bias is plotted against the signal input value for frequency (*top three plots*) and amplitude (*bottom plots*). Linearity was calculated through a linear regression, with equation,  $P$  value and  $R^2$  (represents the proportion of the variance for a dependent variable that is explained by an independent variable or variables in a regression model) displayed. The *red diamond* indicated the mean per signal input measurements.

Treemap of Hierarchical Organization of the Data



**Figure 4. Tree map visualization of the hierarchical organization of the data per participant-eye.** Three, 10-second video recordings were captured within a recording session in sequence per eye: trace 1, trace 2, and trace 3. Three recording sessions were conducted throughout the day, at three different time points: between morning (session 1), afternoon (session 2), and evening (session 3). Three recording sessions were conducted over 3 days within a single five-day work week - day 1 (*light blue*), day 2 (*dark blue*), and day 3 (*green*).

fell below 4 degrees/s.<sup>28</sup> The program's search window for finding the start of the microsaccade for future metric calculations was 30 ms before the selection and 45 ms after the selection for the end of the microsaccade (30 ms, plus 15 ms to account for lens wobble). A 50,000 degrees/s<sup>2</sup> acceleration cutoff was used to eliminate extraneous motion extraction artifacts. An amplitude within the 0.008 degrees to 2 degrees range was required to be selected as a microsaccade. The largest amplitude threshold (2 degrees) previously adopted in studies utilizing modern eye tracking technology was selected to minimize the likelihood of selection bias.<sup>4,29,30</sup> Selected microsaccades that were  $\leq 8$  ms between one another were combined as one individual microsaccade occurrence. Blinks were automatically detected and removed from the data set.

Simultaneous viewing of the raw eye-tracking videos with accompanying motion traces allowed for manual verification, removal, and/or addition of microsaccades. All verifications were performed by a single, trained technician (author L.R.) based on slowed-time raw video visualization. After the manual count, traces in which the automated count differed from the manual count by greater than  $\pm 1$  microsaccade were excluded (33 traces) from the eligible sample of 641 traces, for a final total of 608 traces. Trace-level data were included in the primary analytic sample (PAS), whereas the session-level analytic sample (SAS) was derived from the primary sample; session-level MMs were calculated by averaging the six traces recorded over a single session per participant. The PAS was used to assess inter-trace reliability and repeatability of the MMs, whereas the SAS was used for interday and intraday variability. Of note, 9.87% of the traces in the PAS have no values for the MMs, as these participants had no microsaccades throughout the 10-second recording time.

## Statistical Analysis

For this study, we report the following MMs: average saccade frequency, average amplitude, average peak velocity, and average peak acceleration. Descriptive statistics were used to characterize the overall measures of central tendency (mean and median) and variation (variance, standard deviation [SD], interquartile range [IQR]) for each session-level MM, along with their respective horizontal and vertical plane contributions. Because none of the MMs were normally distributed, nonparametric statistics were used for test-retest reliability and repeatability analyses. The MMs were also log-transformed to approximately conform to normality.

To assess the reliability of the MMs at different hierarchical levels (trace to trace, session to session [interday], and day to day [intraday]), a Wilcoxon matched-pairs signed rank test was used for paired test-retest values. Differences were calculated by subtracting the value from the latter time point from the former (i.e. subtract value of session 2 from session 1, value of day 3 from day 2). A *P* value of less than 0.05 was considered statistically significant.

Bland Altman statistics were conducted to assess the agreement between the test and retest estimates. The Bland Altman limits of agreement (95% LoFA), defined as two SD above and below the mean difference, identified any systematic differences between the measurements. We also calculated intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals (CI) to assess the repeatability of the MMs across three traces within a session, utilizing the primary analytic sample. ICCs were also calculated for interday and intraday repeatability, using the secondary analytic sample. These ICCs were calculated based on a mean rating ( $k = 3$ ), absolute agreement definition, calculated via a two-way mixed effects model.<sup>31</sup> The log-transformed MMs were utilized to fulfill the normality assumption.

To understand the variance of the session-level microsaccade metrics within a participant over a week, session-level coefficients of variance (CV) were calculated by dividing the within-participant standard deviations by the within participant means. We then averaged all the within-participant CVs to obtain an overall average measure of CV for each MM. To assess if there was a learning effect due to repeated testing, a linear mixed model was constructed in which the session-level CV was modeled against session number, and participant ID was included as a random effect. All the statistical methods were conducted using R version 4.0.3.

## Results

### Gage R&R, Linearity, and Bias Analyses

Eye motion traces were extracted from 11 of the 12 model eye signals using a custom MATLAB software, with a reference frame generated from the lowest frequency, highest amplitude eye motion signal to ensure a high quality, minimal artifact reference frame. The second frequency input at 30 Hz was discarded due to overlap with the device framerate of 29.005. The GR&R showed that most of the variance in the frequency and amplitude analyses was explained by the known variance of the predetermined range of input

**Table 2.** Gage Linearity and Bias Results for Frequency and Amplitude

Signal	Value	AA		AB		AC	
		Bias	P Value	Bias	P Value	Bias	P Value
Frequency (Hz)	Average	−0.11	0.06	−0.16	0.09	−0.12	0.15
	1	−0.02	0.73	0.06	0.21	0.10	0.26
	3	0.01	0.78	0.14	0.29	0.09	0.23
	4	0.04	0.28	−0.02	0.87	0.13	0.05
	40	−0.47	0.00 <sup>b</sup>	−0.53	0.00 <sup>b</sup>	−0.57	0.00 <sup>b</sup>
	100	0.31	0.01b <sup>†</sup>	0.08	0.84	0.20	0.09
Amplitude (arcsec)	Average	1.65	0.09	−4.20	0.01 <sup>a</sup>	−4.30	0.00 <sup>b</sup>
	34.5	1.11	0.27	0.41	0.47	−0.48	0.52
	60	0.84	0.54	0.60	0.30	−0.60	0.58
	120	1.47	0.48	−4.98	0.25	−0.48	0.73
	360	2.10	0.49	−12.00	0.15	−7.56	0.06
	480	2.44	0.43	−6.22	0.03 <sup>a</sup>	−11.00	0.00 <sup>b</sup>

Bias is calculated by subtracting the observed value from the expected value (predetermined signal input). The *P* values were derived from one-tail *t*-tests where the null hypothesis is that the bias is 0, and thus the measurement is accurate. We see that higher frequency and amplitude values have significant bias present.

The *P* values were based on one sample *t*-tests.

<sup>a</sup>Significant at  $P < 0.05$ .

<sup>b</sup>Significant at  $P < 0.01$ .

signals. In other words, intra-operator and inherent equipment variation were very low, with a combined contribution of 0.007% for frequency and 0.14% for amplitude (see Fig. 1, Fig. 2). The frequency linearity and bias analysis showed no correlation between frequency signal input value and bias for all operators (see Fig. 3). For amplitude, the TSLO underestimated large amplitudes, with linearity being present for operator AB ( $P = 0.03$ ) and AC ( $P = 8.7e^{-6}$ ). The maximum observed biases were measured at −0.57 Hz by operator AC at a predetermined frequency of 40 Hz and −0.18 arcminutes by operator AC at an amplitude of 8 arcminutes (Table 2).

### Healthy Control Fixation

In total, our analytic sample consisted of 14 participants, 121 recording sessions, and 608 traces, with an average (SD) age of 28.9 (5.8; Table 3). Out of the 14 healthy controls analyzed, only 1 individual – subject ID “S02” – was a trained and experienced observer. All other participants were new to the device, the task, and fixational recordings. The contribution of analytic traces from each participant is displayed in Figure 5. To visualize the mean, median, and IQR of each microsaccade metric per participant, box plots were plotted by subject ID (Fig. 6). Subjects 6 and 15 had lower values of saccade frequency, leading to a smaller sample size of microsaccades and thus a larger

spread of the remaining MMs. Overall, the participants had an average (SD) microsaccade frequency of 0.84 Hz (0.52 Hz), an average amplitude of 0.32 degrees (0.11 degrees) or 19.2 arcmin (6.6 arcmin), an average peak velocity of 43.68 degrees/s (14.02 degrees/s), and an average peak acceleration of 13,920.04 degrees/s<sup>2</sup> (4,186.84 degrees/s<sup>2</sup>). These MMs, along with their horizontal (x-axis) and vertical (y-axis) contributions are displayed in Table 3. We found the CV for all metrics to either fall within the good (10–20%) or acceptable (20–30%) ranges, with a 24.77% CV for the average saccade frequency, 16.45% for average amplitude, 19.22% for the average peak velocity, and 23.47% for the average acceleration (see Table 3).<sup>32</sup>

### Trace-to-Trace Reliability and Repeatability

We recorded three 10-second retinal videos per eye per session with minimal time gaps (<30 seconds) between recorded traces for a single eye. To assess inter-trace reliability, we utilized the primary analytic sample and conducted a Wilcoxon-Signed Rank test on three comparisons: trace 1 to trace 2, trace 1 to trace 3, and trace 2 to trace 3. We found that the trace 1 to trace 2 comparison had the highest number of statistically significant median differences, with the average peak velocity being higher in trace 1 by 3.93 degrees/s ( $P = 0.02$ ), and the peak acceleration higher in trace 1 by 1,312.30 degrees/s<sup>2</sup> ( $P = 0.023$ ). Their respective horizontal and vertical components were

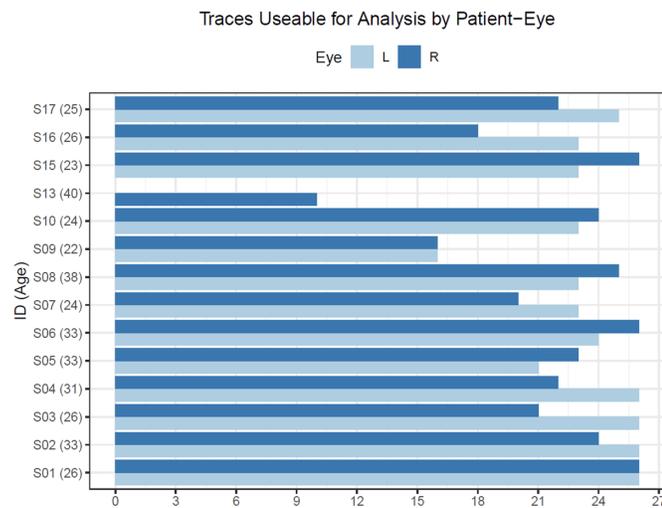
**Table 3.** Descriptive Statistics of Healthy Control Data

Variable	Mean	SD	Median	IQR	Overall CV
Age	28.86	5.76	26.00		
Saccade frequency (Hz)	0.84	0.52	0.78	0.92	24.77
Average amplitude (degrees)	0.32	0.11	0.29	0.11	16.45
Average horizontal amplitude	0.26	0.11	0.23	0.08	23.11
Average vertical amplitude	0.13	0.07	0.12	0.09	30.41
Average peak velocity (degrees/s)	43.68	14.02	40.64	14.41	19.22
Average horizontal peak velocity	38.20	14.52	35.76	14.72	24.16
Average vertical peak velocity	18.96	7.15	17.96	10.36	25.36
Average peak acceleration (degrees/s <sup>2</sup> )	13,920.04	4186.84	13,246.04	5187.40	23.47
Average horizontal peak acceleration	11,441.92	4348.52	10,812.03	5013.38	27.70
Average vertical peak acceleration	9594.98	3637.42	8972.80	4084.92	30.19

Mean, standard deviation (SD), median, IQR, and CV are reported for all microsaccade characteristics and their horizontal and vertical components. Session-level coefficients of variance (CV) were calculated by dividing the within participant standard deviations by the within participant means. We then averaged all the within participant CVs to obtain an overall average measure of CV for each MM as shown below.

Interquartile range (IQR): difference when subtracting first quartile from the third quartile.

Coefficient of variance (CV): calculated as standard deviation divided by mean.



**Figure 5.** Traces available for analysis by participant eye, with age. Left eye (light blue) and right eye (dark blue) number of traces shown.

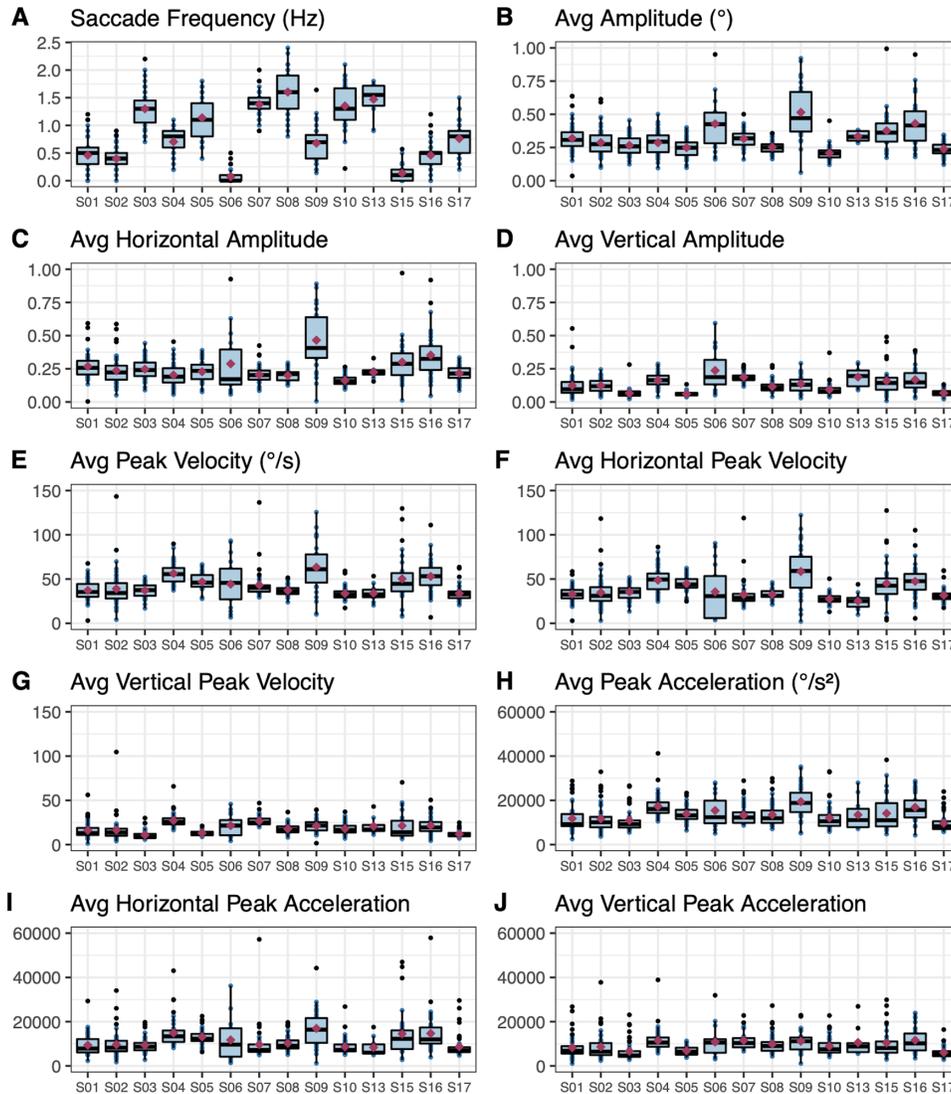
also statistically significant in their differences, with trace 1 having a higher value than trace 2. We also found that trace 1 was statistically different from trace 3 with a higher average peak acceleration in trace 1 by 1,657.14 degrees/s<sup>2</sup> ( $P = 0.009$ ). Last, we found that the vertical velocity measurement was higher in trace 3 when compared to trace 2, but only by 1.75 degrees/s ( $P = 0.017$ ; Table 4).

Next, we looked at the ICC to determine measurement reliability for each MM across the three traces obtained within a session. Following the ICC thresholds set by Koo et al.,<sup>31</sup> we found that saccade

frequency had excellent reliability, with an ICC value of 0.92 (0.91, 0.94). Additionally, the average vertical amplitude had good reliability with an ICC value of 0.81 (0.77, 0.84). All other metrics had ICCs within the moderate reliability range of 0.5 to 0.75. The average horizontal amplitude was the only metric with poor reliability, with an ICC of 0.41 (Fig. 7). The Bland Altman plots revealed a bias of larger differences with an increasing average between the test-retest estimates, more pronounced in the average amplitude measurements and the derived MMs (average velocity and average acceleration), than the saccade frequency measurements (Supplementary Fig. S1). This is in line with the results from the gage linearity and bias analyses.

### Intraday Reliability and Repeatability

Recording sessions were performed 3 times daily; session 1 between 9:00 AM and 11:00 AM (morning), session 2 between 1:00 PM and 3:00 PM (afternoon), and session 3 between 4:00 PM and 6:00 PM (evening) to assess whether the time of day had an impact on fixational measurements. Utilizing the session-level analytic sample, we conducted a Wilcoxon-Signed Rank test and found that microsaccade frequency was the only MM to have a significant difference, with the evening value being significantly higher than the morning value at 0.098 Hz ( $P = 0.007$ ; Table 5). Additionally, we found that microsaccade frequency had an excellent session-to-session reliability, with an ICC (95% CI) of 0.97 (0.955, 0.982). Average vertical



**Figure 6. Box plots of trace-level microsaccade metrics by participant.** Box length represents interquartile range. The black line across the box indicates the median, and the red dot within the box indicates the mean. Individuals with lower values in saccade frequency (S06 and S15) have larger spreads for other MMs, which may have contributed to the higher CV measures.

amplitude and average vertical velocity all had good session-to-session reliabilities, with ICCs of 0.88 (0.82, 0.93), 0.79 (0.67, 0.87), and 0.77 (0.65, 0.86), respectively. All other metrics had ICCs within the moderate range (see Fig. 7) and accompanying plots show the mean differences between session-level (Supplementary Fig. S2).

**Interday Reliability and Repeatability**

Fixation was recorded over the course of 3 days within a 5-day workweek. We aimed to assess whether day-to-day differences had an impact on fixational measurements for the session-level analytic sample. Using the Wilcoxon-Signed Rank test, we found no significant differences from day to day in microsaccade

characteristics ( $P > 0.05$ ; Table 6). For day-to-day reliability, microsaccade frequency retained the highest ICC values at 0.966 (0.948, 0.979), and average amplitude, average vertical amplitude, and average vertical velocity all retained their good reliability scores, with ICCs of 0.84 (0.76, 0.9), 0.76 (0.64, 0.85), and 0.77 (0.65, 0.86), respectively. All other metrics had ICCs within the moderate range (see Fig. 7). All differences between test-retest estimates centered around 0, with no obvious display of bias with increasing values (Supplementary Fig. S3).

**Learning Model**

To assess whether there was a session-level learning effect impacting the measurements, CVs were

**Table 4.** Trace-to-Trace Variability of Microsaccade Metrics

Microsaccade Metric	Trace 1 to Trace 2			Trace 1 to Trace 3			Trace 2 to Trace 3					
	MD	95% LofA	Points%	P Value	MD	95% LofA	Points%	P Value	MD	95% LofA	Points%	P Value
Saccade frequency (Hz)	0.01	-0.73, 0.75	93.56	0.83	-0.01	-0.74, 0.72	93.64	0.38	0.01	-0.71, 0.72	96.49	0.97
Average amplitude (degrees)	0.01	-0.24, 0.26	94.01	0.54	0.01	-0.25, 0.26	91.89	0.96	-0.01	-0.28, 0.26	93.06	0.38
Average horizontal amplitude	0.01	-0.25, 0.27	93.41	0.85	0.00	-0.25, 0.25	91.89	0.92	-0.01	-0.28, 0.26	93.75	0.62
Average vertical amplitude	0.00	-0.15, 0.15	94.01	0.56	0.01	-0.13, 0.14	95.95	0.75	-0.01	-0.14, 0.13	93.75	0.25
Average peak velocity (degrees/s)	3.93	-32.76, 40.61	93.41	0.02 <sup>a</sup>	1.66	-38.12, 41.45	95.27	0.30	-2.59	-40.8, 35.61	93.75	0.14
Average horizontal peak velocity	3.27	-30.59, 37.13	94.01	0.03 <sup>a</sup>	1.04	-36.91, 38.98	95.27	0.60	-1.80	-39.5, 35.89	93.06	0.44
Average vertical peak velocity	1.85	-19.64, 23.35	92.81	0.03 <sup>a</sup>	1.29	-21.82, 24.4	95.95	0.08	-1.76	-24.51, 20.99	96.53	0.02 <sup>b</sup>
Average peak acceleration (degrees/s <sup>2</sup> )	1312.30	-15,175.11, 17,799.71	91.62	0.02 <sup>a</sup>	1657.14	-12,665.42, 15,979.7	91.22	0.01 <sup>b</sup>	-244.67	-14,744.03, 14,254.68	93.06	0.48
Average horizontal peak acceleration	2191.08	-14,490.69, 18,872.84	95.21	0.00 <sup>b</sup>	1212.21	-16,464.08, 18,888.5	94.59	0.22	-1143.68	-15,522.09, 13,234.74	94.44	0.05
Average vertical peak acceleration	948.00	-12,331.95, 14,227.96	92.81	0.01 <sup>a</sup>	1171.69	-11,321.02, 13,664.41	95.27	0.07	-346.40	-12,739.84, 12,047.04	93.75	0.45

Mean differences (MDs) and Bland Altman Statistics (95% Limits of Agreement) were calculated using the primary analytic sample. The P values were derived from Wilcoxon Signed Matched Pairs. A significant P value signals to a statistically significant median difference between the test and retest estimates, indicating that repeatability may be impacted by the trace number.

Mean difference (MD; i.e. subtract the values of trace 2 from trace 1).

95% LofA: Limits of agreement, 2 standard deviations above and below the MD.

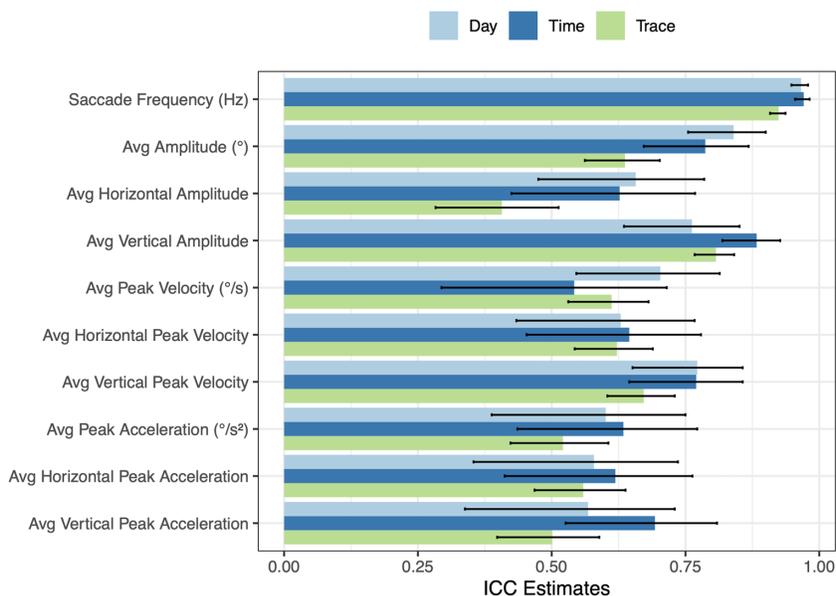
P value: P value based on Wilcoxon Signed Matched Pairs Test.

Points%: proportion of points that fall within the 95% Limit of Agreement.

<sup>a</sup>Significant at P < 0.05.

<sup>b</sup>Significant at P < 0.01.

Repeatability ICC Estimates of Microsaccade Metrics



**Figure 7. ICCs (2,3) were calculated using log-transformed MMs.** The primary analysis sample was used for trace level ICC, whereas session level analysis was used for time and day level ICC. Values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively.

**Table 5. Session-to-session Variability of Microsaccade Metrics**

Microsaccade Metric	Time 1 to Time 2				Time 1 to Time 3				Time 2 to Time 3			
	MD	95% LofA	Points%	P Value	MD	95% LofA	Points%	P Value	MD	95% LofA	Points%	P Value
Saccade frequency (Hz)	-0.06	-0.53, 0.42	95.00	0.11	-0.10	-0.51, 0.32	97.44	0.01 <sup>a</sup>	-0.05	-0.49, 0.39	92.50	0.22
Average amplitude (degrees)	0.01	-0.25, 0.26	92.50	0.35	0.02	-0.24, 0.27	94.87	0.34	0.01	-0.13, 0.14	95.00	0.81
Average horizontal amplitude	0.01	-0.27, 0.3	90.00	0.53	0.01	-0.27, 0.29	92.31	0.61	-0.01	-0.16, 0.16	92.50	0.49
Average vertical amplitude	0.00	-0.12, 0.12	92.50	0.30	0.01	-0.09, 0.11	94.87	0.23	0.01	-0.1, 0.12	97.50	0.69
Average peak velocity (degrees/s)	-1.11	-32.32, 30.1	95.00	0.55	0.69	-32.92, 34.31	92.31	0.98	1.63	-22.5, 25.75	95.00	0.30
Average horizontal peak velocity	-0.56	-35.23, 34.11	92.50	0.73	0.02	-35.12, 35.16	94.87	0.87	0.48	-22.15, 23.1	95.00	0.90
Average vertical peak velocity	-0.04	-14.76, 14.67	92.50	0.95	1.24	-10.52, 13.01	94.87	0.30	1.05	-14.13, 16.23	90.00	0.65
Average peak acceleration (degrees/s <sup>2</sup> )	-850.98	-9453.92, 7751.96	97.50	0.16	129.85	-10,031.39, 10,291.09	92.31	0.67	856.04	-8052.63, 9764.7	95.00	0.30
Average horizontal peak acceleration	-150.84	-9767.28, 9465.6	90.00	0.51	3.86	-11,832.76, 11,840.47	89.74	0.83	28.14	-7797.09, 7853.36	97.50	0.79
Average vertical peak acceleration	-656.00	-10,195.87, 8883.87	90.00	0.50	211.67	-6167.16, 6590.51	92.31	0.99	804.26	-8340.41, 9948.93	92.50	0.19

Note: Mean differences (MDs) and Bland Altman Statistics (95% Limits of Agreement) were calculated using the session-level analytic sample. The P values were derived from Wilcoxon Signed Matched Pairs. A significant p-value signals to a statistically significant median difference between the test and retest estimates, indicating that repeatability may be impacted by time of day.

Mean difference (MD; i.e. subtract the values of trace 2 from time 1).

95% LofA: Limits of agreement, 2 standard deviations above and below the MD.

P value: P value based on Wilcoxon Singed Matched Pairs Test.

Points%: proportion of points that fall within the 95% Limit of Agreement.

<sup>a</sup>Significant at P < 0.05.

<sup>b</sup>Significant at P < 0.01.

calculated at the session-level. Afterward, a linear mixed model was constructed in which the session-level CV was modeled against session number, with the participant variable included as a random effect. No significant learning effect from session to session was observed in any of the metrics considered (P > 0.05; Table 7).

## Discussion

Before characterizing the variability in a healthy control population, we assessed the TSLO’s performance as a measurement system for FEM. Through the GR&R and linearity and bias analyses, we were

**Table 6.** Day-to-day Variability of Microsaccade Metrics

Microsaccade Metric	Day 1 to Day 2				Day 1 to Day 3				Day 2 to Day 3			
	MD	95% LofA	Points%	P Value	MD	95% LofA	Points%	P Value	MD	95% LofA	Points%	P Value
Saccade frequency (Hz)	0.00	-0.39, 0.4	92.11	0.35	0.01	-0.43, 0.44	92.31	0.67	-0.01	-0.48, 0.45	95.00	0.80
Average amplitude (degrees)	0.00	-0.15, 0.14	94.74	0.60	-0.03	-0.25, 0.19	89.74	0.11	-0.03	-0.24, 0.19	90.00	0.42
Average horizontal amplitude	-0.02	-0.22, 0.18	94.74	0.33	-0.04	-0.29, 0.21	89.74	0.12	-0.02	-0.27, 0.23	90.00	0.49
Average vertical amplitude	0.02	-0.18, 0.21	92.11	0.81	0.00	-0.13, 0.14	92.31	0.52	-0.01	-0.15, 0.14	95.00	0.52
Average peak velocity (degrees/s)	-2.32	-27.98, 23.35	92.11	0.22	-4.34	-38.49, 29.82	92.31	0.08	-2.41	-33.44, 28.62	87.50	0.38
Average horizontal peak velocity	-3.33	-33, 26.33	92.11	0.07	-4.28	-41.57, 33	92.31	0.14	-1.72	-34.57, 31.14	90.00	0.39
Average vertical peak velocity	0.69	-13.97, 15.34	92.11	0.51	-1.11	-14.66, 12.44	97.44	0.28	-1.58	-16.5, 13.34	90.00	0.12
Average peak acceleration (degrees/s <sup>2</sup> )	-118.85	-8836.77, 8599.07	94.74	0.82	-1033.65	-10,807.4, 8740.11	94.87	0.38	-831.91	-11,239.72, 9575.9	92.50	0.29
Average horizontal peak acceleration	-926.51	-10,506.57, 8653.55	94.74	0.19	-678.95	-10,985.11, 9627.2	92.31	0.61	117.46	-9207.78, 9442.7	92.50	0.76
Average vertical peak acceleration	400.17	-7485.22, 8285.56	94.74	0.37	-496.52	-9446.71, 8453.66	92.31	0.49	-736.41	-10,598.99, 9126.17	95.00	0.30

Note: Mean differences (MDs) and Bland Altman Statistics (95% Limits of Agreement) were calculated using the session-level analytic sample. The *P* values were derived from Wilcoxon Signed Matched Pairs. A significant *P* value signals to a statistically significant median difference between the test and retest estimates, indicating that repeatability may be impacted by day of the week.

Mean difference (MD; i.e. subtract the values of trace 2 from time 1).

95% Lofa: Limits of Agreement, 2 standard deviations above and below the MD.

*P* value: *P* value based on Wilcoxon Singed Matched Pairs Test.

Points%: proportion of points that fall within the 95% Limit of Agreement.

<sup>a</sup>Significant at *P* < 0.05.

<sup>b</sup>Significant at *P* < 0.01.

**Table 7.** Learning Model Results

Microsaccade Metric	$P_{\beta_1}$	$\beta_1$	$B_0$
Saccade frequency (Hz)	0.77	7.14	48.25
Average amplitude (degrees)	0.62	6.67	20.80
Average horizontal amplitude	0.73	6.53	30.28
Average vertical amplitude	0.34	1.62	28.23
Average peak velocity (degrees/s)	0.84	1.09	27.11
Average horizontal peak velocity	1.00	-0.41	33.36
Average vertical peak velocity	0.98	3.68	31.86
Average peak acceleration (degrees/s <sup>2</sup> )	0.80	8.25	33.44
Average horizontal peak acceleration	0.58	-2.45	45.23
Average vertical peak acceleration	0.66	0.03	39.62

Where a linear mixed effects model with session-level CV was modeled against session number (ranging from 1 to 9). Subject ID was included as a random effect. A significant *P* value indicated the presence of a learning effect, which was not found.

<sup>a</sup>Significant at *P* < 0.05.

<sup>b</sup>Significant at *P* < 0.01.

able to quantify the repeatability, reproducibility, and accuracy of the TSLO device. We saw that the TSLO had excellent repeatability and reliability for frequency and amplitude measurements, with the main contributor to variability being the chosen variability between signal inputs. The TSLO was incredibly accurate even at high-frequency values, as there was no linearity present in any of the operators (*P* > 0.05). Nevertheless, we saw a negative correlation between amplitude signal input value and bias for operators AB (*P* = 0.03) and AC (*P* < 0.00), meaning that the TSLO tended to underestimate measurements at higher amplitude values. It is important to note, however, that

human fixation tends to follow a 1/frequency amplitude trend, with a majority of large-amplitude motion being ≤ 10 Hz.<sup>33,34</sup> This means that the bias seen at the higher amplitude value signals, as paired with high frequency values, will likely have minimal clinical significance, given that these are physically impossible for humans to perform.

In general, microsaccade frequency had the highest ICC values for all assessments, with vertical amplitude also consistently above 80%. Microsaccade frequency was on average (SD) 0.84 (0.52) Hz during the 10-second recordings, which translates to roughly 8 microsaccades (± 5). Based on the standard devia-

tion, one can infer those fluctuations of less than five microsaccades from baseline likely represent normal variance.

On the trace level, we found that trace 1 consistently had larger peak acceleration recordings as compared to both trace 2 ( $P = 0.023$ ) and trace 3 ( $P = 0.009$ ), with a higher peak velocity than trace 2 ( $P = 0.02$ ). These results may suggest that a single 10-second TSLO video recording alone may not be adequate to fully characterize microsaccade acceleration and velocity characteristics that would result from using the existing fixational paradigm. A recent study from the University of Pittsburgh utilizing the TSLO device to quantify fixation in individuals with concussion versus controls, took five, 30-second fixation recordings from control subjects and broke each recording up into three, 10-second traces for comparison. They found no statistically significant difference in the microsaccade characteristics among the three time increments.<sup>35</sup> However, it is important to note that this was measured at only a single time point and the experience between a continuous 30 second recording, versus back-to-back 10-second recordings like you see in this paper, could pose additional differences. For future studies utilizing this study's stimuli and time increment, we propose an initial 10-second video acquired for training purposes only, followed by  $\geq 2$  videos to be recorded and analyzed so as to capture the most reliable microsaccade metrics.

Next, looking at the time-of-day measurements, we found that the morning session had fewer microsaccades on average than the evening session, albeit by a small amount of 0.098 Hz (which roughly translates to only 1 microsaccade). Previous literature outlining changes in microsaccade characteristics, particularly velocity, suggests that mental fatigue could be a contributing factor to this time-of-day difference, though further studies would be required to elucidate the fatigue versus attentional components with the time of day. Nevertheless, it is important to note that with smaller sample sizes, the time of day will be an important factor to hold constant, whereas, in larger studies with more participants, this time-of-day effect would likely become negligible. Finally, in the time frame of a single workweek, we saw no evidence of a learning effect for any of the MMs ( $P > 0.5$ ).

Although a high sensitivity eye-tracking device provides a noninvasive method of recording the smallest of microsaccade characteristics, there are also limitations to this approach. First, the image-based eye-tracking method utilized in this analysis requires a reference frame to track translational shifts of the retina over time. Because the eye is raster-scanned pixel by pixel, a reference frame is not a single snapshot

in time and can have motion artifacts at roughly the frame rate (30 Hz). To combat this, we used a custom-built offline eye-tracking software that builds a single reference frame based on the movement of the eye during a full 300-frame movie. This can minimize reference frame artifacts, although not entirely remove them. Next, the fixation stimulus' size, luminance, and contrast can all impact the fixational eye motion exhibited by participants.<sup>18,19</sup> By utilizing the corner of our 5 degrees  $\times$  5 degrees imaging raster as the target, this could have led to visual fading for some participants, which can impact the characteristics of microsaccades themselves. This could have been one of the drivers for the larger CV values observed in the MMs. However, when comparing our average (standard deviation) amplitude value of microsaccades – 19.2 (6.6) arcminutes – to the average amplitude of control microsaccades obtained utilizing a 4 arcminute, 100% contrast black circle via Dual-Purkinje eye-tracking system of 20 (7) arcminutes, the values are almost identical.<sup>28</sup> Third, participants were permitted to wear glasses and contacts during their fixational eye motion recordings. For the subjects wearing glasses, the space between the lens and the pupil of the eye can cause a slight magnification of the retinal image, which directly impacts the pixels per degree calculations of movement. Fourth, whereas the exclusion criteria of the study prevented the recruitment of individuals with neurodegenerative and retinal disease, we do not exclude individuals with attention deficit hyperactivity disorder (ADHD), dyslexia, depression, anxiety, migraines, or any prescription medications. Additionally, we do not account for any caffeine intake, which may impact microsaccades.<sup>36</sup> Finally, a sample size of only 14 people may not adequately capture the full population variance that may be encountered, although it is a sufficient sample to estimate the likely parameters that we would expect to see. The individuals with zero microsaccades during their recording sessions also could have negatively influenced the ICC results and likely lead to an underestimation of TSLO repeatability.

## Conclusions

The TSLO device showed promising repeatability and reproducibility with minimal intra-operator variability and equipment error contributions to the overall variance. Utilizing our fixational stimuli and 10-second recording paradigm, we observed trace-level differences in microsaccade characteristics when comparing our initial trace to subsequent record-

ings. Additionally, we saw a modest increase in the number of microsaccades depending on the time of day, with morning sessions having fewer microsaccades on average than the evening. We found no statistically significant differences in day-to-day measurements of microsaccades that were observed, revealing no significant learning effect. The goal of future experimental paradigms requires tasks that simultaneously elicit the most abnormalities and instability but do so in a repeatable fashion. The optimum intersection of these two requirements will optimize the detection of pathological change and the findings of this study will motivate future experimental paradigms.

## Acknowledgments

The authors thank Andy Zhang and Chaio-Han Chuang for contributing to the model eye data acquisition process. The model eye data acquisition project was originally conceived for a University of California, Berkeley Engineering capstone project for masters-level students with faculty mentor Austin Roorda and industry sponsor, C. Light Technologies. Subsequent Gage R&R analysis of the model eye data was completed at UCSF. This paper was presented as two posters showcasing the model eye and fixational microsaccade data separately for ARVO 2021.

Supported by National Institutes of Health TL1 TR 001871 (C.K.S.) and R41 NS100222-01A1 (A.J.G. and C.K.S.), and That Man May See (A.J.G. and C.K.S.).

Disclosure: **S.Y. Condor Montes**, None; **D. Bennett**, None; **E. Bensinger**, C. Light Technologies (I); **L. Rani**, None; **Y. Sherkat**, None; **C. Zhao**, None; **Z. Helft**, C. Light Technologies (O, P); **A. Roorda**, C. Light Technologies (I); University of California, Berkeley (P); **A.J. Green**, University of California - San Francisco (P); **C.K. Sheehy**, C. Light Technologies (E, O, P); University of California, Berkeley (P), University of California - San Francisco (P)

\* SYCM and DM performed an equal amount of work in this study.

## References

1. Ko HK, Snodderly DM, Poletti M. Eye movements between saccades: Measuring ocular drift and tremor. *Vision Res.* 2016;122:93–104.
2. Riggs LA, Ratliff F. Visual acuity and the normal tremor of the eyes. *Science.* 1951;114:17–18.
3. Ditchburn RW, Ginsborg BL. Vision with a Stabilized Retinal Image. *Nature.* 1952;170(4314):36–37.
4. Martinez-Conde S, Macknik SL, Troncoso XG, Dyar TA. Microsaccades counteract visual fading during fixation. *Neuron.* 2006;49(2):297–305.
5. Tulunay-Keesey Ü. Fading of stabilized retinal images. *J Optical Soc Am.* 1982;72(4):440–447.
6. Sheehy CK, Bensinger ES, Romeo A, et al. Fixational microsaccades: A quantitative and objective measure of disability in multiple sclerosis. *Mult Scler.* 2020;26(3):343–353.
7. Alexander RG, Macknik SL, Martinez-Conde S. Microsaccade Characteristics in Neurological and Ophthalmic Disease. *Front Neurol.* 2018;9:144.
8. Kosnik W, Fikre J, Sekuler R. Visual fixation stability in older adults. *Invest Ophthalmol Vis Sci.* 1986;27(12):1720–1725.
9. Abadi RV, Gowen E. Characteristics of saccadic intrusions. *Vision Res.* 2004;44(23):2675–2690.
10. Engbert R, Kliegl R. Microsaccades Keep the Eyes' Balance During Fixation. *Psychol Sci.* 2004;15(6):431.
11. Hafed ZM, Clark JJ. Microsaccades as an overt measure of covert attention shifts. *Vis Res.* 2002;42(22):2533–2545.
12. Rolfs M, Engbert R, Kliegl R. Microsaccade orientation supports attentional enhancement opposite a peripheral cue: commentary on Tse, Sheinberg, and Logothetis (2003). *Psychol Sci.* 2004;15(10):705–707; author reply 708–710.
13. Meyberg S, Sinn P, Engbert R, Sommer W. Revising the link between microsaccades and the spatial cueing of voluntary attention. *Vision Res.* 2017;133:47–60.
14. Siegenthaler E, Costela FM, McCamy MB, et al. Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *Eur J Neurosci.* 2014;39(2):287–294.
15. Valsecchi M, Betta E, Turatto M. Visual oddballs induce prolonged microsaccadic inhibition. *Exp Brain Res.* 2007;177(2):196–208.
16. Dalmaso M, Castelli L, Scatturin P, Galfano G. Working memory load modulates microsaccadic rate. *J Vis.* 2017;17(3):6.
17. Di Stasi L, McCamy M, Catena A, Macknik S, Cañas J, Martinez-Conde S. Microsaccade and drift dynamics reflect mental fatigue during visual search. *J Vis.* 2013;13:1346.
18. McCamy MB, Otero-Millan J, Macknik SL, et al. Microsaccadic Efficacy and Contribution to Foveal and Peripheral Vision. *J Neurosci.* 2012;32(27):9194–9204.

19. Steinman RM. Effect of Target Size, Luminance, and Color on Monocular Fixation. *J Opt Soc Am.* 1965;55(9):1158–1164.
20. Rucci M, Poletti M. Control and Functions of Fixational Eye Movements. *Ann Rev Vis Sci.* 2015;1(1):499–518.
21. Martinez-Conde S, Otero-Millan J, Macknik SL. The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nat Rev Neurosci.* 2013;14(2):83–96.
22. Sheehy CK, Yang Q, Arathorn DW, Tiruveedhula P, de Boer JF, Roorda A. High-speed, image-based eye tracking with a scanning laser ophthalmoscope. *Biomed Opt Express.* 2012;3(10):2611–2622.
23. Mulligan J. Recovery of motion parameters from distortions in scanned images. Published online December 1, 1997. NASA Ames Research Center Moffett Field, CA, United States. Available at: <https://ntrs.nasa.gov/citations/19980236600>.
24. Gómez-Vieyra A, Dubra A, Malacara-Hernández D, Williams DR. First-order design of off-axis reflective ophthalmic adaptive optics systems using a focal telescopes. *Opt Express.* 2009;17(21):18906–18919.
25. Dubra A, Sulai Y. Reflective afocal broadband adaptive optics scanning ophthalmoscope. *Biomed Opt Express.* 2011;2(6):1757–1768.
26. Stevenson SB, Roorda A, Kumar G. Eye tracking with the adaptive optics scanning laser ophthalmoscope. In: ETRA '10: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. Marcy 22 – 24, 2010. New York, NY: Association for Computing Machinery; 2010:195–198.
27. Cano EL, Moguerza JM, Redchuk A. *Six Sigma with R: Statistical Engineering for Process Improvement*. New York, NY: Springer-Verlag; 2012.
28. Cherici C, Kuang X, Poletti M, Rucci M. Precision of sustained fixation in trained and untrained observers. *J Vis.* 2012;12(6):31.
29. Poletti M, Rucci M. A compact field guide to the study of microsaccades: Challenges and functions. *Vis Res.* 2016;118:83–97.
30. Sinn P, Engbert R. Small saccades versus microsaccades: Experimental distinction and model-based unification. *Vis Res.* 2016;118:132–143.
31. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* 2016;15(2):155–163.
32. Cui ZC. Allowable limit of error in clinical chemistry quality control. *Clinical Chemistry.* 1989;35(4):630–631.
33. Ezenman M, Hallett PE, Frecker RC. Power spectra for ocular drift and tremor. *Vis Res.* 1985;25(11):1635–1640.
34. Stevenson SB, Sheehy CK, Roorda A. Binocular eye tracking with the Tracking Scanning Laser Ophthalmoscope. *Vis Res.* 2016;118:98–104.
35. Leonard BT, Kontos AP, Marchetti GF, Zhang M, Eagle SR, Reeher HM, Bensinger ES, Snyder VC, Holland CL, Sheehy CK, Rossi EA. Fixational eye movements following concussion. *J Vis.* 2021;21(13):11, 1–14.
36. Collins NE, AlKalbani M, Kavanagh P, Mahmud A, Boyle G, Coakley D. Effects of Caffeine on Fixational Eye Movements. *Invest Ophthalmol Vis Sci.* 2009;50(13):2882.