https://doi.org/10.2352/ISSN.2470-1173.2018.14.HVEI-528

# Statistical Identification of Fixations in Noisy Eye Movement Data

Jeffrey B. Mulligan;Human Systems Integration Division, NASA Ames Research Center; Moffett Field, CA / USA

## Abstract

Human eye movements typically consist of a series of fixations (during which the eye is relatively still), linked by saccades, which rapidly reorient the direction of gaze to a new location. The locations fixated usually indicate the allocation of attention, and are useful when making inferences concerning state awareness in complex information environments such as an aircraft cockpit. Identification of fixation events is straightforward when measurement noise is low (on the order of the physiological noise, typically a few arc minutes), but becomes increasingly challenging as noise increases to the levels encountered in current video-based remote tracking systems, which are suitable for installation in flight simulators. Here we present a novel method for identification of fixations and microsaccades in noisy eye position records. We assume that the data has first been processed with a velocity-based saccade detector, so that we are left with relatively short intervals of relatively constant data. The method attempts to fit the signal with a piece-wise constant function, splitting the data into two sub-intervals to produce the least RMS error in the fit. Proposed splits are accepted or rejected on the basis of a statistical t-test, with the level of significance providing a single parameter controlling the sensitivity. We compare the method to other position-based techniques, such as the classic ”dispersion” method (which grows fixations rather than splitting as in our method).

## Introduction

The analysis of eye movement data begins with the parsing of the raw signals into fixations, saccades, and smooth movements. In this paper, we focus on the identification of fixations. In particular, we are concerned with the discrimination of multiple nearby fixations, when the distance between successive fixations is comparable to the noise level of the tracking system.

A commonly-used approach is known as the dispersion algorithm [1]. In this method, fixations are grown by adding samples whose distance from the cluster mean is less than a predefined threshold. The determination of this threshold can be somewhat problematic; too small of a threshold will produce spurious fixations, while too large of a threshold will miss small saccades. While a method has been proposed to determine an optimal threshold based on repeatability [2], here we present a method allowing direct control of the false alarm rate by fixing the significance level of a statistical test used to accept a proposed saccade.

## Methods

The proposed method was designed to deal with intervals of signal that remain after segmentation at large saccades, which may be identified using a velocity criterion [3]. In this paper, we restrict our attention to the analysis of synthetic data, which consist of piece-wise-constant signals corrupted by the addition of Gaussian-distributed white noise. An example of such a signal is shown in figure 1. We first consider one-dimensional signals; generalization to two dimensions can be done in various ways, discussed at the end of this section.

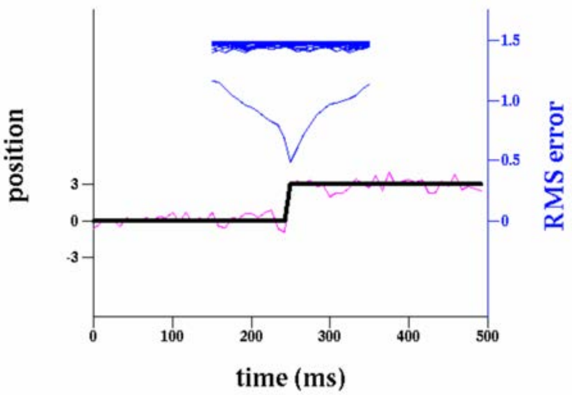

Figure 1. The heavy black trace shows a ”ground truth” signal consisting of a saccade (step) with an amplitude of 2; the overlaid red trace shows a sample measurement corrupted by noise with a standard deviation of 1. The V-shaped blue trace shows the root-mean-square error of the fit of a piece-wise constant function, as a function of the split point. The nearly flat blue traces above the V-shaped trace show the same computation performed on permutations of the input data.

We attempt to fit the input data with a piece-wise constant signal, by considering introducing a ”split” at various points within the signal. To minimize the sum-of-squared deviations, we compute the means of the samples on each side of the split, and use those values for our estimate of the underlying signal. We impose the constraint that a fixation has a minimum duration of m samples. Figure 1 shows a synthetic signal (in black) together with noise-corrupted ”measurement” (in red). The RMS error of a piece-wise constant fit to the observed data is shown by the V-shaped blue trace above the signal, with a clear minimum above the RMS error for a number of permutations of the input data In the examples shown here, we set m to a value of 9, corresponding to a minimum fixation duration of 150 ms for a sample rate of 60 Hz. Thus, the blue error curves in figure 1 begin (and end) 150 ms after (and before) the start (and end) of the data record.

This procedure tells us where to segment the signal in

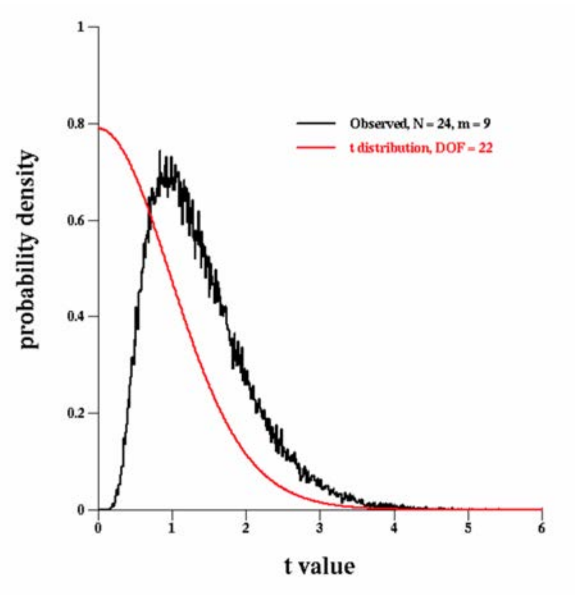

Figure 2. Distribution of the computed t statistic (in black), along with the t distribution with 22 degrees of freedom.

order to obtain the best fit to the input data. But to decide whether or not to accept this split, we must determine whether the difference between the mean positions on either side of the split is statistically significant. We examine two methods for making this decision, a permutation test, and a t-test.

The permutation test addresses the problem from the standpoint that if all of the samples were in fact drawn from a single underlying position, then the quality of fit of the best split would be about the same if the samples are randomly permuted. On the other hand, if the best-fitting split arose from a true shift in the position, then permuting the data would substantially degrade the quality of the fit. We perform the process for some large number of permutations, sort the resulting quality scores, and compare the value from the real data with the sorted list. For example, if we wish to accept splits that are significant at the p=0.01 significance level (accepting a false alarm rate of 0.01), we could 1000 permutations and accept the proposed split if the RMS error is less than the 10th smallest permutation error. Figure 1 shows the RMS error for a number of permutations of the input data

The disadvantage of the permutation test method is that the entire computation (including exhaustive search for the best split point) must be performed anew for each permutation. Thus, evaluating 1000 permutations, as in the example above, increases the computational burden by a factor of 1000. To avoid this, we instead can use a standard t-test. The t statistic is computed using the formula:

$$t = \frac{\mu_1 - \mu_2}{\sigma_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{1}$$

where $\mu_1$ and $\mu_2$ are the means of the two intervals (with lengths $n_1$ and $n_2$), and $\sigma_p$ is the pooled variance:

$$\sigma_p = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}, \tag{2}$$

with $\sigma_1$ and $\sigma_2$ representing the variances in the two intervals.

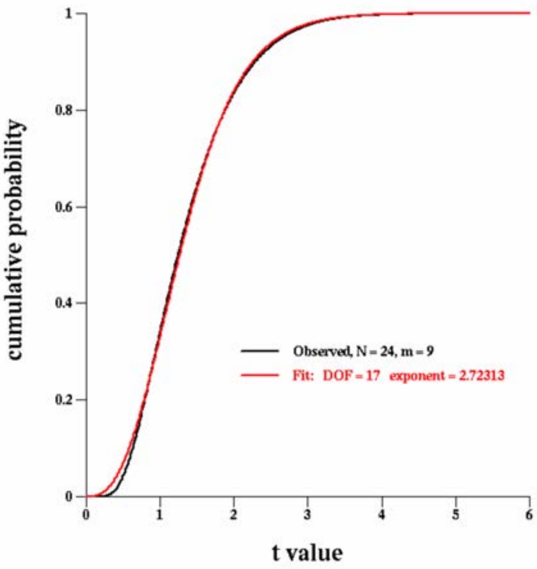

Figure 3. The cumulative distribution corresponding the the observed distribution in figure 2, together with the best fitting cumulative exponentiated t distribution. See text for details.

When we compare the t statistic to the standard t-table to test for significance, however, we obtain too may false positives when testing with a constant signal. Figure 2 shows the standard t distribution (in red) along with an empirically obtained distribution of the t statistic obtained using our procedure. We would obtain the standard distribution if we tested once, at a single predetermined split point; but in our procedure we perform the test at a split point chosen from many, to optimize the fit.

Intuitively, we suspect that the observed distribution may arise as the maximum of a number of t values, computed for various split points. The distribution of the maximum of a number of independent samples is most easily calculated from the cumulative distribution: if $C(x)$ is the probability that a sample t value is less than $x$, then the probability that $N$ independent samples are less than $x$ is $(C(x))^N$. In the present case, however, the t values computed at nearby split points are not independent, as they share most of the constituent data points. We therefore attempted to fit the observed cumulative distribution with a cumulative t distribution raised to an arbitrary power. The cumulative distribution corresponding to the empirical distribution shown in figure 2 is shown in figure 3, along with the best fitting exponential of a standard cumulative t distribution. The fit is fairly good, but there is no obvious relationship between the degrees of freedom and exponent used to obtain the fit, and the parameters used in the simulation.

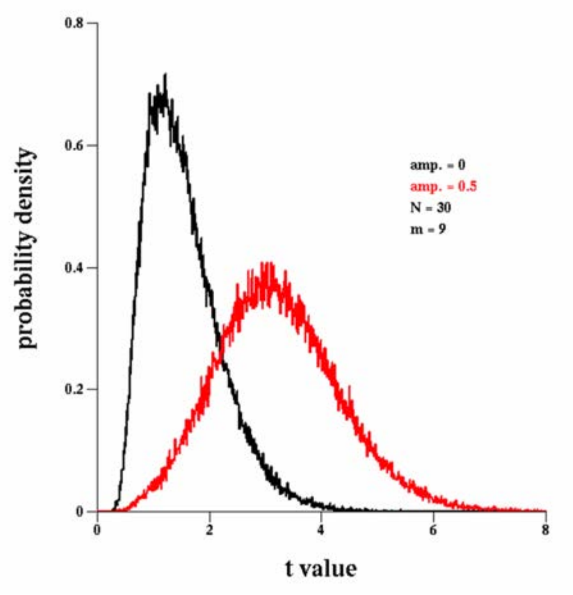

Figure 4. Observed t statistic distributions for a noisy signal with no saccade (black), and one with a saccade with an amplitude of half the noise standard deviation.

The preceding analyses have been performed using one-dimensional signals, but eye-trackers generally report two-dimensional gaze directions. The method is easily generalized to two dimensions: when evaluating a possible split point, the horizontal and vertical dimensions are fit independently, and the total squared error is obtained by summing the squared error for each of the dimensions.

Performing the statistical test for acceptance of the split is slightly more complicated. One way to generalize the method to two dimensions is simply to perform the test described above independently to the horizontal and vertical coordinates. This gives a slight penalty to oblique saccades. A slightly more complicated approach which does not suffer from this flaw is to project each two-dimensional sample onto the line passing through the means on either side of the split, and then using the position on this line as a one-dimensional coordinate, and applying the one-dimensional test as described above.

## Results

We test the performance of the method by generating test input signals containing small saccades with a range of amplitudes. Figure 4 shows the distribution of the t statistic for a signal containing a saccade with an amplitude equal to half the noise standard deviation (in red), along with the no-signal distribution (as shown previously in figure 2). In practice, we will select a criterion t value, and accept splits where the observed value exceeds the criterion, and reject those where it does not. The significance level that we choose for the test determines the false alarm rate. By varying the criterion in small increments, we can trace out the Receiver Operating Characteristic or ROC

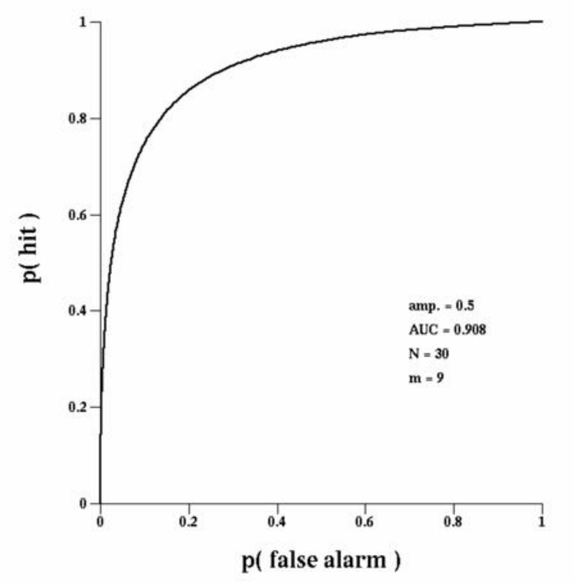

Figure 5. Receiver Operating Characteristic (ROC) curve generated from the distributions shown in figure 4.

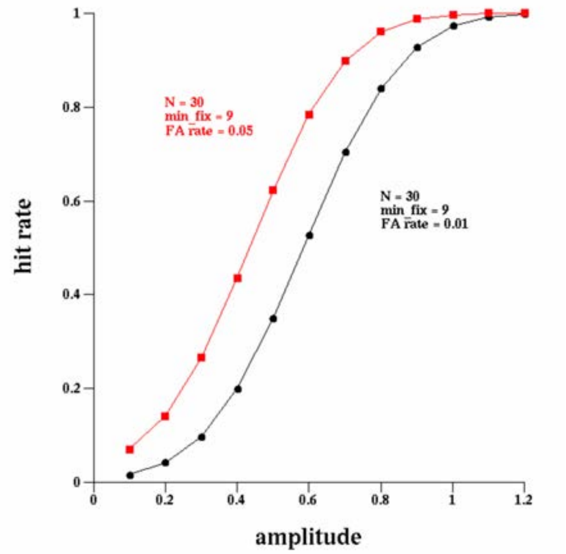

Figure 6. Hit rate versus saccade amplitude (expressed in units of noise standard deviation), for two different false alarm rates.

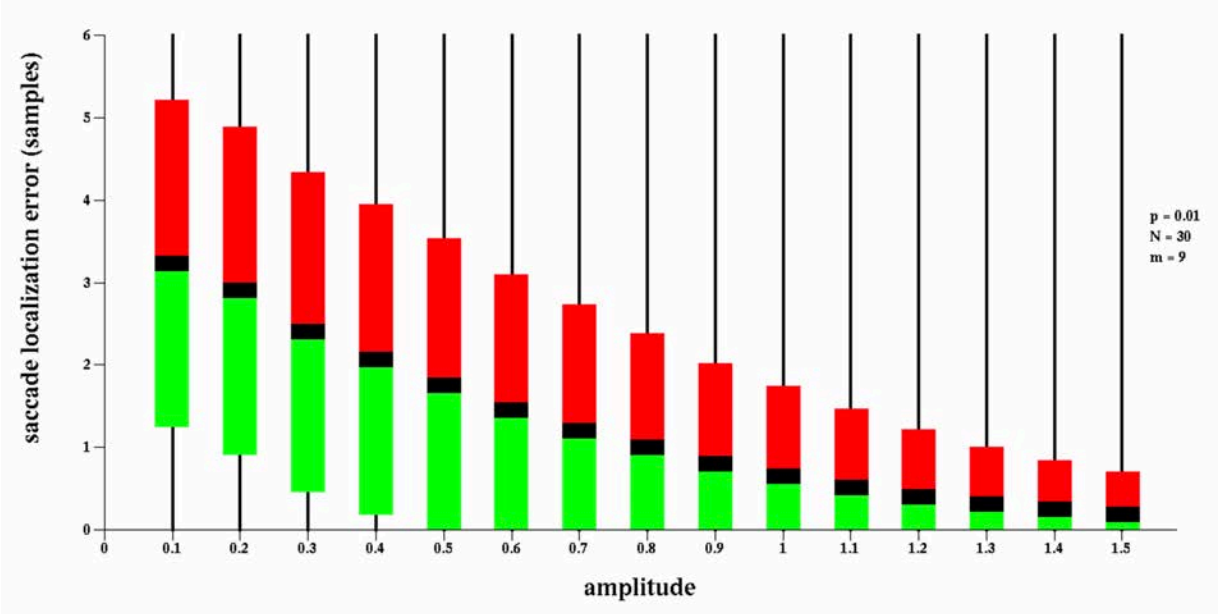

Figure 7. Distributions of saccade timing error magnitudes as a function of saccade amplitude, expressed in units of noise standard deviation. For each saccade amplitude, the horizontal black bar indicates the mean timing error (in samples), while the red and green boxes show plus and minus two standard deviations; the whiskers show the limits of the observed range.

curve, shown in figure 5 for the distributions shown in figure 2.

## Detection performance

Figure 6 shows the hit rate as a function of saccade amplitude, for two false alarm rates: 0.01 (black), and 0.05 (red). The figure illustrates that, for either criterion, most saccades having an amplitude greater than the noise standard deviation are detected, while detection is falling off rapidly at an amplitude of half of the noise standard deviation.

## Timing accuracy

In the previous section, we counted a ”hit” whenever a saccade was detected in a signal that we created by adding noise to a two-level signal. In figure 1, the saccade amplitude is larger than the noise standard deviation, and the saccade in the fit aligns perfectly with the ground truth signal. But as saccade amplitude is decreased, the estimated saccade time (location of the optimal split point) will often deviate from ground truth. Figure 7 shows the magnitudes of these errors (in samples), plotted as a function of saccade amplitude. The red and green bars represent plus and minus two standard deviations from the mean (shown in black), while the whiskers show the extreme values (which in this case span the entire range of possibilities).

## Amplitude accuracy

For the smaller saccades, accuracy is degraded not only in the estimate of the time of occurrence but also for the magnitude of the saccade. Figure 8 shows a histogram of estimated amplitudes for saccades having a ground truth amplitude of 0.1 relative to the noise standard deviation.

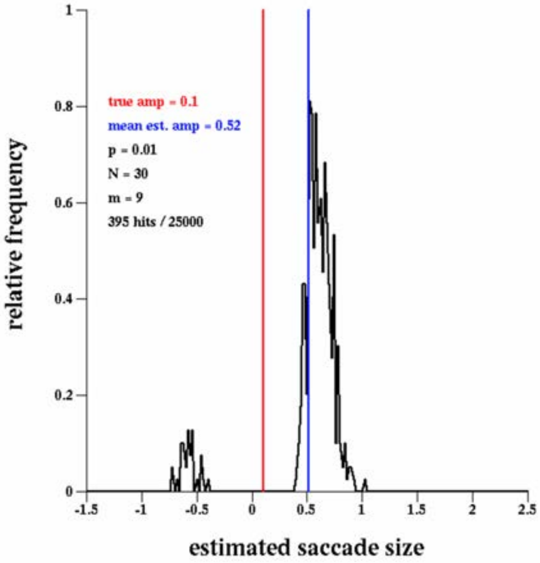

Figure 8. Histogram of estimated saccade amplitudes, for a true amplitude of 0.1 times the noise standard deviation. The estimated amplitude always overestimates the true amplitude.

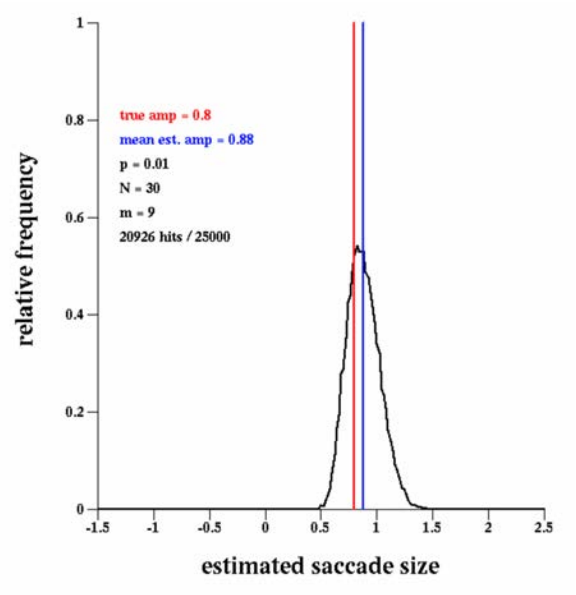

Figure 9. Histogram of estimated saccade amplitudes, for a true amplitude of 0.8 times the noise standard deviation. The estimated amplitude is close to the true amplitude.

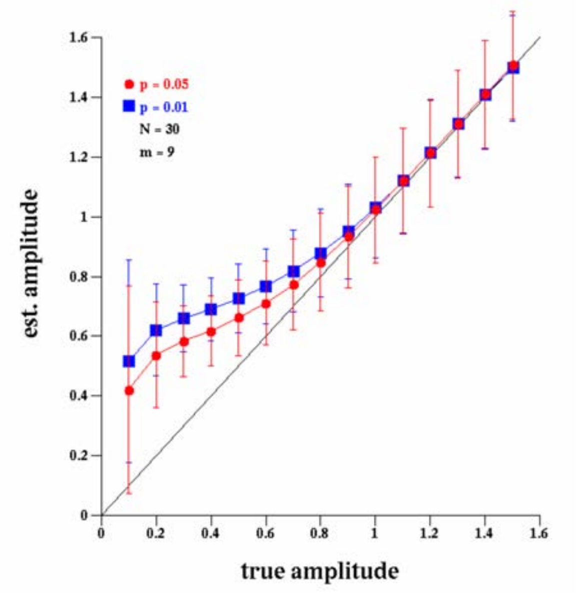

Figure 10. Average estimated saccade amplitude as a function of ground truth amplitude, for two false alarm rates. For amplitudes below the noise standard deviation, the estimates are inflated because only those trials where the noise reinforces the signal are detected.

We can see from figure 6 that very few of these saccades are detected; those that are detected are detected because the added noise fortuitously is correlated with the signal, and so aids in its detection. We also see in figure 8 that, for a small fraction of these detected saccades, the estimated amplitude is opposite in sign to the ground truth signal - these should probably be considered false alarms rather than hits!

Figure 9 shows the case for a larger saccade amplitude. In this case, the average estimated amplitude is only slightly larger than the ground truth value. Figure 10 summarizes these results by plotting estimated saccade amplitude versus ground truth amplitude for two false alarm rates, 0.05 (shown in red) and 0.01 (shown in blue). For both false alarm rates, as the ground truth amplitude drops below 1 (in units of noise standard deviation), the estimated amplitude drops more slowly, reflecting the fact that as detection performance degrades, only those trials in which the added noise reinforces the signal will be detected.

## Discussion

The preceding results have been obtained using simulated sequences corresponding to 500 milliseconds, sampled at 60 Hz. The implicit assumption is that large saccades will occur with sufficient regularity that long records can be segmented into short intervals free of large saccades, and that the proposed method can then be applied to these intervals. However, it is quite possible that records may contain intervals of several seconds containing no large saccades, for which the proposed method may be called upon to find multiple small saccades. An early implementation of the procedure [3] applied the procedure to long records, and then recursively subdivided intervals resulting from each split, until all of the intervals were too short for further splitting. This approach has several disadvantages. First, the computational burden is O(NlogN), where N is the length of the sequence. (The basic method is O(N) for a single pass, with O(logN) recursive passes.) A more serious problem is that there can occur signals containing 3 or more positions (separated by arbitrarily large intervals) for which there is no single binary split that produces a statistically significant difference between the two intervals. When the position shifts are large, a velocity-based saccade-finder can segment the signal into smaller intervals, but these types of signal can render the recursive approach ineffective when the saccades are small. Therefore, it is suggested that longer records that are free of large saccades be analyzed with a sliding window of roughly 500 milliseconds.

Our computation of the noise distribution (figure 4) assumed a noise model of Gaussian-distributed position noise. This is probably reasonable for video-based eye trackers when the main source of noise arises from the images (e.g., sensor and quantization noise), but characterization of the noise properties of a system should be done before applying the method. A more difficult problem arises from measurement errors that are not properly ”noise,” but nevertheless increase the variability of measurements. One example comes from systems that measure both the pupil position and the location of one or more ”glints” (reflections from the cornea of one or more illuminators). The glint tracking subsystem may falsely lock onto a scleral reflection, making it seem as if a saccade has occurred when in fact there was none. This can be especially devastating if it occurs during the calibration phase. Similarly, systems that compute a simple pupil centroid (rather than fitting an ellipse to the pupil margin) will suffer from systematic errors when the pupil is partially occluded by an eyelid. In all cases, these effects will increase the apparent variability of the measurements, while having characteristics that are distinct from white noise.

Positional ”noise” observed during a fixation does not come solely from measurement noise; small instabilities of the eye itself, known as fixational eye movements, are inescapable, and can contribute to the apparent noise of a system during calibration. Fortunately, these are small, usually less than 10 minutes of arc during a typical fixation [4, 5, 6], and so can be safely neglected in cases where measurement noise exceeds 0.5 degree. In future work, we hope to incorporate a model of fixational eye movements into the simulations presented here to investigate their effect on the performance of the method in low-noise situations. Unlike the positional noise used in these simulations, which was white in position, fixational drift is thought to be white in velocity, leading to a 1/f spectrum when expressed as position [7, 8].

Finally, our model of the underlying signal as a piecewise constant function has a number of limitations. First, it neglects the fact that real saccades take a finite time to execute, with larger saccades taking longer in accordance with the ”main sequence” [9, 10]. This becomes important for systems with a high sampling rate, where there may be a number of samples taken during a saccade while the eye is in flight. If these samples are included when considering a split near the saccade center, they will bias the estimated position and degrade the quality of the fit. One approach might be to discard samples near the split point when evaluating a prospective split. Alternatively, one might incorporate a model of saccade dynamics [11, 12] when constructing model signals.

Another way in which our piece-wise constant signal model fails to capture real behavior is that there is no provision for smooth pursuit. Smooth motions are also encountered in records from head-mounted recording systems during head movement, due to the vestibulo-ocular reflex (VOR). This might be addressed by generalizing the signal model from piece-wise constant to piece-wise linear.

## Summary

Large saccades are easily detected using a velocity threshold, but the detection of small saccades can be problematic. We have presented a novel method for the identification of small saccades in noisy eye movement signals. Our analysis reveals fundamental limits in the detection of small saccades having amplitudes smaller than the noise standard deviation. The smallest saccades are only detected when combined with favorably correlated noise, causing their amplitude to be over-estimated. The method is potentially useful in situations involving relatively high noise levels, such as remote camera systems installed in operational environments, with closely-spaced areas-of-interest which must be distinguished.

## Acknowledgments

This work was supported by the Technologies for Aircraft State Awareness (TASA) project of NASA's Airspace Operations and Safety program.

## References

[1] Salvucci, D. D. and Goldberg, J. H., “Identifying fixations and saccades in eye-tracking protocols,” in [Proceedings of the 2000 Symposium on Eye Tracking Research & Applications], ETRA ’00, 71–78, ACM, New York, NY, USA (2000).

[2] Blignaut, P., “Fixation identification: the optimum threshold for a dispersion algorithm,” Atten. Percept. Psychophys. 71(4), 881–895 (2009).

[3] Kalar, D. J., Liston, D., Mulligan, J. B., Beutter, B., and Feary, M., “Considerations for the use of remote gaze tracking to assess behavior in flight simulators,” Tech. Rep. NASA/TM-2016-219424, ARC-E-DAA-TN36043 (2016).

[4] Cornsweet, T. N., “Determination of the stimuli for involuntary drifts and saccadic eye movements,” Journal of the Optical Society of America 46(11), 987–993 (1956).

[5] St. Cyr, G. J. and Fender, D. H., “The interplay of drifts and flicks in binocular fixation,” Vision Res. 9, 245–265 (1969).

[6] Rucci, M. and Poletti, M., “Control and functions of fixational eye movements,” Annu. Rev. Vis. Sci. 1, 499–518 (2015).

[7] Findlay, J. M., “Frequency analysis of human involuntary eye movement,” Kybernetik 8, 207–214 (1971).

[8] Stevenson, S. B., Roorda, A., and Kumar, G., “Eye tracking with the adaptive optics scanning laser ophthalmoscope,” in [Proceedings of the 2010 Symposium on Eye-Tracking Research &#38; Applications], ETRA ’10, 195–198, ACM, New York, NY, USA (2010).

[9] Bahill, A. T. and Stark, L., “The trajectories of saccadic eye movements,” Scientific American 240, 108–117 (1979).

[10] Duchowski, A., Krejtz, K., Biele, C., Niedzielska, A., Kiefer, P., Giannopoulos, I., Gehrer, N., and Schönenberg, M., “An inverse-linear logistic model of the main sequence,” Journal of Eye Movement Research 10(3) (2017).

[11] Han, P., Saunders, D. R., Woods, R. L., and Luo, G., “Trajectory prediction of saccadic eye movements using a compressed exponential model,” Journal of Vision 13(8), 27 (2013).

[12] Harwood, M. R., Mezey, L. E., and Harris, C. M., “The spectral main sequence of human saccades,” J. Neurosci. 19(20), 9098–9106 (2015).

## Author Biography

Jeffrey B. Mulligan received an A.B. (in physics) from Harvard University, and M.A. and Ph.D. degrees (in psychology) from the University of California at San Diego. He has been a computer engineer at NASA Ames Research Center since 1987, following a year as a National Research Council post-doctoral fellow. His work has included studies of human motion perception, and the responses of the eye movement system to various types of motion stimuli. He has also worked on improving eye movement measurement technology, and the measurement of eye movements in applied settings.