

The Incomplete Fixation Measure

Frederick Shic*
Brian Scassellati†
Yale University

Katarzyna Chawarska‡
Yale University School of Medicine

Abstract

In this paper we evaluate several of the most popular algorithms for segmenting fixations from saccades by testing these algorithms on the scanning patterns of toddlers. We show that by changing the parameters of these algorithms we change the reported fixation durations in a systematic fashion. However, we also show how choices in analysis can lead to very different interpretations of the same eye-tracking data. Methods for reconciling the disparate results of different algorithms as well as suggestions for the use of fixation identification algorithms in analysis, are presented.

CR Categories: J.4 [Computer Applications]: Social and Behavioral Sciences—Psychology

Keywords: eye-tracking, fixation identification, dispersion, dwell, mean fixation, Lévy flight

1 Introduction

The coupling of eye-tracking hardware with off-the-shelf scanpath analysis software has made it easy to overlook the dependence of eye-tracking results on the details of the algorithms used. In this paper, following in the steps of [Karsh and Breitenbach 1983] who remarked upon the “amorphous fixation measure”, we examine how the fundamental unit of eye-tracking analysis, the fixation, depends on both the algorithm and parameters employed. In contrast to previous work, we do not examine the scanpath directly. Instead, we examine how outcome measures are impacted by changes to fixation algorithms. We focus on mean fixation duration, a measure typically associated with cognitive load [Jacob and Karn 2003].

2 Fixation Identification

We examine simple fixation identification algorithms that are likely to be used in modern psychological and cognitive work. These algorithms are well summarized in [Salvucci and Goldberg 2000] and [Duchowski 2003], and include four dispersion-based threshold algorithms and one velocity threshold algorithm.

2.1 Dispersion-based Threshold Algorithms

Dispersion-based threshold algorithms mark a segment of consecutive points in the scanpath as a fixation if those points obey certain temporal and spatial constraints. The temporal constraint is a duration requirement: if the duration of a fixation is less than a thresh-

old time t_{min} , it is judged to be non-physiological and is marked as invalid. The spatial constraints are more variable. Here we consider four variations: (a) a pure distance dispersion algorithm; (b) the centroid distance scheme by Anliker [1976]; (c) the position variance-method, also due to Anliker; and (d) Salvucci’s I-DT algorithm [2000].

Distance Dispersion Algorithm For a fixation to be valid under the distance dispersion algorithm, each point in that fixation must be no further than some threshold d_{max} from every other point. This is perhaps the most intuitive measure, but is less popular than other simple dispersion algorithms because every fixation point must be checked against every other fixation point, resulting in $O(n^2)$ operations.

Centroid-Distance Method Anliker’s centroid-distance method requires that M of N points be no further than some threshold c_{max} from the centroid of the N points. Since the M of N criteria has denoising properties, for comparability, we have set $M = N$. Also, it is possible to use either a consistent version of this algorithm, where, whenever the fixation is being expanded [Salvucci and Goldberg 2000; Widdel 1984], we recompute the distance of all points in the fixation to the centroid, or a fast version, where we only check the distance of the new point to be added. Here, we use the more consistent version.

Position-Variance Method This method is a variant of the centroid-distance restricted algorithm where it is required that M of N points have a standard deviation of distance from the centroid not exceeding σ_{max} . Again we set $M = N$ and use the consistent interpretation of the algorithm.

Salvucci I-DT Algorithm Salvucci’s fixation identification by dispersion threshold algorithm requires that the maximal horizontal distance plus the maximal vertical distance is less than some threshold m_{max} . This algorithm is a fast approximation to the distance dispersion algorithm, as only the extreme horizontal and vertical positions need to be compared against when adding a new point to a fixation.

2.2 Velocity-based Threshold Algorithms

Velocity-based threshold algorithms focus on saccades rather than fixations. Typically a specific velocity, v_{max} , is chosen as a threshold. Points that exceed this velocity are considered saccades; points below this are potential fixations. To increase the comparability of this algorithm to dispersion-based algorithms, we also employ a minimum time duration requirement for fixations.

3 Subjects and Methods

15 typically developing toddlers (mean age 27 months; range 18 to 33 months) were shown 6 color images of faces [Lundqvist et al. 1998] and 6 block designs (Figure 1) at a distance of 75 cm on a 24” (61 cm) widescreen monitor (16:9 aspect ratio). Each image, including the grey background, was $12.8^\circ \times 17.6^\circ$. Eye-tracking data were obtained simultaneously with a SensoMotoric Instruments IView X RED table-mounted 60Hz eye-tracker. Each image was shown for at least 10 seconds; here we analyze only the first

*e-mail: frederick.shic at yale.edu

†email: scaz at cs.yale.edu

‡e-mail: katarzyna.chawarska at yale.edu

10 seconds. Only data falling within the boundaries of the stimulus image and trials with at least 5 valid seconds of eye-tracking were retained. A total of 41 trials on faces and 27 trials on blocks were obtained. Failure to obtain data was typically due to the toddlers' poor affect and/or inattention; the discrepancy in obtained trials likely reflects the greater salience of faces versus abstract patterns. Results were aggregated at the level of a trial (e.g. mean fixation times are means of trial means).

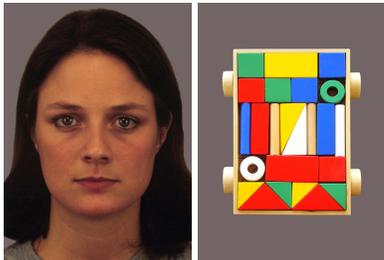


Figure 1: Examples of Experimental Stimuli

4 Results

When the mean fixation time was examined over an extreme range of parameter settings, a non-linear behavior was observed, likely due to saturation effects occurring as the result of extending the spatial extent to the entire stimulus image (Figure 2A). However, when the view was restricted to values corresponding more closely to the physiological limits of foveal vision, a remarkably linear trend was seen (Figure 2B). This was surprising, given the non-linear effects observed for both the total time spent in fixations per trial (Figure 2D) and the number of fixations per trial (Figure 2D).

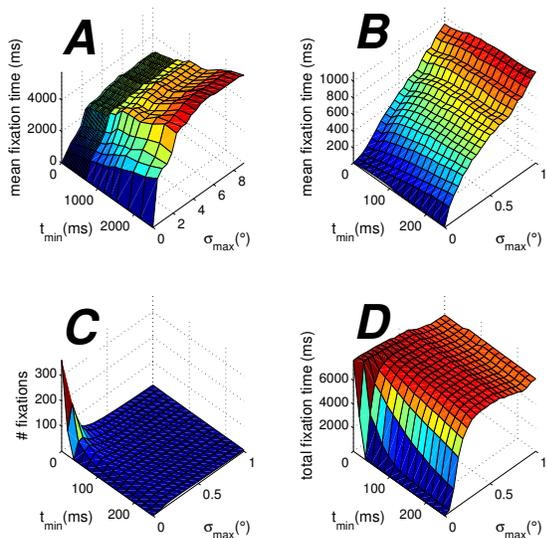


Figure 2: Parameter Dependence of Position-Variance Method on Faces (A: mean fixation time (extended range); B: mean fixation time; C: # fixations/trial; D: total fixation time/trial)

In order to characterize the behavior of the algorithms over physiologically reasonable parameter settings, we examined each algorithm under a grid of uniformly sampled temporal and spatial constraints. The temporal constraint was the minimum duration re-

quirement ($50\text{ms} \leq t_{min} \leq 250\text{ms}$, $N=13$). The spatial constraints $s \in \{d_{max}, c_{max}, \sigma_{max}, m_{max}, v_{max}\}$ were computed by matching the mean fixation time range of each algorithm to the range of the distance dispersion algorithm over ($0.6^\circ \leq d_{max} \leq 5.1^\circ$, $N=16$). A multiple linear regression (without offset) was applied to the mean fixation times (μ) as a function of s and t_{min} in order to find the spatial and temporal slopes of each algorithm ($slope_s$ and $slope_t$, respectively): $\mu(s, t_{min}) = slope_s s + slope_t t_{min}$. The results of this analysis are shown in Table 1.

Table 1: Mean Fixation Times of Algorithms as a Linear Function of Parameters for Faces(top) and Blocks(bottom)

Faces				
method	spatial constraint	$slope_s$	$slope_t$	R^2
distance	$0.6^\circ \leq d_{max} \leq 5.1^\circ$	151	0.98	.996
centroid	$0.4^\circ \leq c_{max} \leq 3.4^\circ$	217	1.10	.995
variance	$0.15^\circ \leq \sigma_{max} \leq 0.85^\circ$	893	1.20	.992
I-DT	$1.5^\circ \leq m_{max} \leq 8.0^\circ$	102	0.85	.998
velocity	$18^\circ/s \leq v_{max} \leq 81^\circ/s$	9.47	0.58	.993

Blocks				
method	spatial constraint	$slope_s$	$slope_t$	R^2
distance	$0.6^\circ \leq d_{max} \leq 5.1^\circ$	178	0.75	1.0
centroid	$0.4^\circ \leq c_{max} \leq 3.6^\circ$	258	0.89	.999
variance	$0.15^\circ \leq \sigma_{max} \leq 0.75^\circ$	1224	0.81	.999
I-DT	$1.5^\circ \leq m_{max} \leq 8.0^\circ$	118	0.57	1.0
velocity	$18^\circ/s \leq v_{max} \leq 81^\circ/s$	9.69	0.51	.988

The behavior of mean fixation time on parameter settings for all algorithms was strikingly linear. Though this is a simple relationship, it is not necessarily an obvious one. Doubling the spatial parameter of the distance dispersion algorithm quadruples the area. If the distribution of points was linear, we would expect a quadratic effect on mean fixation time; if it was diffusive, we would expect an effect for some characteristic length. Instead, we see a simple linear relationship, suggesting the statistical behavior of the scanpaths of toddlers in free-viewing is scale-free. In turn this supports the hypothesis that scanpaths are Lévy flights [Brockmann and Geisel 1999]. Our recent simulations have verified that the mean fixation time computed by the distance method on Lévy flight saccades, but not linear or diffusive saccades, is linear with respect to parameters.

Note also that the consistent behavior of the algorithms implies that mean fixation duration results for different algorithms can be converted to one another. However, we also note that the relationship between the spatial slopes does not necessarily follow an intuitive pattern and that the scale of each spatial parameter is different, sometimes dramatically (e.g. variance method versus I-DT).

The more common use of mean fixation time, however, is for making comparisons. In Figure 3 we compare Faces(+) versus Blocks(-) under the distance method.

We see that, depending on parameter settings, the mean fixation duration for a particular stimulus class can be either greater or lower than another class. We also see that taking into account the variation in trials also impacts analysis. For example, though the difference in fixation durations at low d_{max} in Figure 3 is small in an absolute temporal sense, the effect in comparison to the trial standard deviation is prominent. In Table 2 we examine the regions delineated by the standardized difference (i.e. the difference between the mean fixation times on faces versus blocks divided by the averaged standard deviation). For each method we localize the region corresponding to faces (large positive scores), blocks (large negative scores), and an indeterminate mixed region (small absolute scores). For each of these regions we compute the mean temporal (t) and

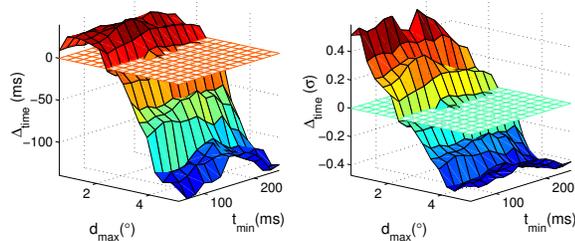


Figure 3: Mean Fixation Time Difference for Distance Method (Faces-Blocks) in ms (left) and standardized difference(right)

spatial (s) parameter. For comparability, using the slopes in Table 1, we also translate t and s to distance dispersion algorithm temporal (t_{dist}) and spatial ($s_{dist}=d_{max}$) equivalents. Finally, we report the mean temporal differences Δ_t (ms), the mean standardized difference $\Delta_t(\sigma)$, and the percentage of the physiologically-reasonable range space covered by each region (%).

Table 2: Characterization of Regions Corresponding to Mean Fixation Time Difference between Faces and Blocks

Face Region ($\Delta_t(\sigma) \geq 0.2$)							
method	t	s	t_{dist}	s_{dist}	Δ_t (ms)	$\Delta_t(\sigma)$	%
distance	133	1.19°	133	1.19°	25.2	.369	18.6
centroid	130	0.77°	145	1.11°	25.1	.350	18.8
variance	136	0.20°	167	1.20°	31.6	.324	17.9
I-DT	143	2.19°	124	1.48°	28.9	.365	24.6
velocity	124	30°/s	73	1.90°	28.1	.349	28.2

Block Region ($\Delta_t(\sigma) \leq -0.2$)							
method	t	s	t_{dist}	s_{dist}	Δ_t (ms)	$\Delta_t(\sigma)$	%
distance	152	4.71°	152	4.71°	-107	-.339	50.2
centroid	153	2.95°	181	4.28°	-110	-.380	41.5
variance	152	0.64°	165	4.42°	-158	-.520	61.0
I-DT	147	7.22°	112	4.79°	-100	-.326	40.0
velocity	-	-	-	-	-	-	0

Mixed Region ($ \Delta_t(\sigma) < 0.2$)							
method	t	s	t_{dist}	s_{dist}	Δ_t (ms)	$\Delta_t(\sigma)$	%
distance	156	2.29°	156	2.29°	-1	.016	31.2
centroid	156	1.84°	179	2.65°	-16	-.042	39.7
variance	153	.337°	178	2.17°	-4	-.012	21.0
I-DT	157	4.43°	129	3.00°	-12	-.050	35.4
velocity	160	57°/s	101	3.32°	12	.069	71.8

For all regions, translating spatial parameters to a common scale lead to greater comparability. Differences in mean fixation time were also consistent. The translation of temporal parameters increased variability, however, possibly due to edge effects and the smaller contribution of temporal versus spatial parameters. In terms of coverage, all algorithms were comparable except velocity. This was not unexpected, as dispersion algorithms share common assumptions regarding the spatial cohesion of fixations; the velocity algorithm assumes that fixations are what are left after saccades have been parsed. The two approaches have slightly different operating characteristics which are amplified by differencing

5 Discussion

A good algorithm for identifying fixations is robust, discriminative, and generates easily interpretable results. We address each of these aspects in turn.

In terms of robustness, the algorithms studied in this paper are all reliable as the effect of parameter changes is predictable and consistent. As shown by Table 1, over a minimum time duration from 50 ms to 250 ms and a spatial extent ranging from 0.6° to 5.1° (distance algorithm equivalent), mean fixation duration for all algorithms is essentially a linear function of parameters. The exact characterization of the linear effect differs, however, between algorithms and between stimuli. The change in slopes for different algorithms follows from the different assumptions each algorithm makes. For example, the centroid method's spatial constraint is the distance from the centroid. In comparison, the distance algorithm's constraint is the diameter of the cluster. From this we might assume that the spatial slope of the centroid method should be twice that of the distance method. However, though the centroid method's slope is higher than the distance method, it is not twice as high. This is likely due to the movement of the centroid as the centroid method expands its fixation window, an effect which reduces the effective scale of the radius constraint. The difference in slopes between stimuli classes is caused by changes in the scanning distribution as a result of changing image properties. We note that for faces, in comparison to blocks, the temporal slopes are higher and the spatial slopes are lower. Though a more formal analysis is necessary to localize exactly the source of this change, it is likely that when scanning faces saccades are shorter and fixations more dense, reflecting a fundamental processing shift in information retrieval mechanisms.

In terms of the discriminative capabilities of the algorithms, as examined by differencing the mean fixation durations of faces and blocks in Figure 3 and Table 2, the effects of changing parameters on the final interpretation is shown to be quite troubling: by extending the spatial scale the differences first become insignificant, then reverse! It could be argued that the regime corresponding to higher mean fixation durations for blocks is non-physiological, being centered at the extreme range of foveal vision. However, the relationship between mean fixation time and parameters is fairly consistent, possessing an elementary quality across all algorithms. Furthermore, all dispersion algorithms yield similar results when considering the character of differencing regions. This suggests that rather than being an artifact of data processing, the effects shown are an inherent property of viewing the respective stimulus classes.

We thus propose that instead of picking a single set of parameters suitable to all analyses, an approach that is both difficult and incomplete, that the effect of the parameter space be charted, slopes of effects reported, and different regimes of dominant behavior characterized. This would provide a more complete picture as to the actual scanpath dynamics involved in observing static scenes. Furthermore, given the predictable effect of parameter changes on mean fixation time, this would also allow new results to be compared to the extant literature, and thus could offer a simple method for bridging previously inconsistent results.

Since all the algorithms behave inherently in the same manner, it is an open question as to which algorithm should be used in analysis. We note that the ratios of spatial slopes between distance, centroid, and the I-DT methods are preserved across stimulus types. This suggests that this set of algorithms is particularly comparable and that it might be possible to freely convert parameters even without *a priori* knowledge of stimulus changes. Within this set, we believe the distance method is the most transparent and interpretable, as it simply ensures that every point is within some distance to every other point. By contrast, the centroid in the centroid method tends to shift as the fixation is being calculated. Similarly, the I-DT method is asymmetric with regards to both radial distance and spanned area. For example, a series of points falling within a long, thin region could be viewed as valid as a square of maximal area, despite having an edge nearly twice as long. However, the I-DT

method is also the fastest dispersion method examined here and could be used when speed is of primary importance.

The current study is limited in many ways. For one, we only consider simple algorithms for fixation identification. We do not consider clustering, optimization, or more advanced techniques (e.g. see [Salvucci and Goldberg 2000; Privitera and Stark 2000; Santella and DeCarlo 2004]). However, many of these algorithms also incorporate at least some spatial or temporal free parameters, and therefore their study would also be amendable to the methods presented here. We should note that the data reduction method on which the dispersion algorithms here are based [Widdel 1984] is greedy and that there are benefits to considering all possible fixations and choosing the best match rather than simply the first match.

We also only consider mean fixation duration, though a host of other measures exist [Jacob and Karn 2003; Inhoff and Radach 1998]. This was done in the interests of brevity, but also because mean fixation duration is a central statistic in studies examining the relationship between cognitive processing and scene viewing. Unfortunately, we also do not take into account aspects of the image scene which could be tied directly to the observed slopes. For example, we do not consider the spatial frequency, noise, or second order statistics of images; neither do we employ any ecological or higher-level interpretations, such as the importance of facial features or the saliency of circles embedded in a field of rectangular blocks. Future work will characterize these relationships by relating the parameter space to region and image properties.

In addition, our choice of toddlers as a subject population is both a strength and a weakness. It is a strength as it shows that mean fixation duration is quite a robust statistic even when the target population is difficult, suggesting that it may be an appropriate measure in situations where data loss is a major concern. It is a weakness as the trends found might only be applicable to our particular choices of testing conditions, namely free-viewing of static images in extremely young children. It is certainly conceivable that adults would scan scenes in a more organized fashion, and it is certain that tasks could be designed which would limit the variability of the scanpath. This, of course, would itself be interesting.

Finally, we should note that the choice of fitting the mean fixation-parameter curves without an offset was done so as to match the theoretical effects of having a spatial parameter of zero and to make the parameter space more interpretable for the purposes of this study. This leads to some difficulties with the statistics of the regression as it is known that R^2 reported for linear regression without an offset overestimates the fit in comparison to regression with an offset [Gordon 1981]. We subsequently applied a linear regression with an offset and found the lowest R^2 to be 0.957, supporting the linear trend. Future work will examine the implications of this constant offset term, as it could provide an explanation for the effects observed for stimulus class differencing, as well as provide a reliable estimate for the overhead associated with cognitive processing.

6 Conclusions

We have shown that the effect of parameter changes on mean fixation duration is essentially a linear function of parameters, supporting the hypothesis that spatial scanning is scale-free. We have also demonstrated that by changing parameter settings, specific findings can be made insignificant or even reversed. However, these reversals are likely to be characteristic properties of the visual processing of the scene. For that reason, it seems that the important question is not *what are the correct parameter settings for our algorithm?* but rather *what are the relationships between parameter changes and our outcome measures?* Instead of viewing fixation identification algorithms as being analogues of physiological processes,

we should view these algorithms simply as techniques for grouping scanpath behavior at characteristic scales. We recommend, then, that for any given study using fixation statistics, that multiple parameter settings be tested so as to provide a more complete picture as to the impact of parameter manipulation. In conclusion, we suggest that the fixation measure of the past is perhaps not so much amorphous as it is incomplete.

Acknowledgements

The authors would like to thank Warren Jones and Ami Klin for their help in shaping this work. This study was supported by a National Science Foundation CAREER award (#0238334) as well as NSF award #0534610 (Quantitative Measures of Social Response in Autism). Additional support was provided by a software grant from QNX Software Systems Ltd. We also gratefully acknowledge support provided through research and training grants by Autism Speaks and NAAR. Finally, we would like to thank the parents and children involved in this study.

References

- ANLIKER, J. 1976. Eye movements- On-line measurement, analysis, and control. *Eye movements and psychological processes*, 185–199.
- BROCKMANN, D., AND GEISEL, T. 1999. Are human scanpaths Lévy flights? *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470) 1*.
- DUCHOWSKI, A. 2003. *Eye Tracking Methodology: Theory and Practice*. Springer.
- GORDON, H. 1981. Errors in Computer Packages. Least Squares Regression Through the Origin. *The Statistician 30*, 1, 23–29.
- INHOFF, A., AND RADACH, R. 1998. Definition and computation of oculomotor measures in the study of cognitive processes. *Eye guidance in reading and scene perception*, 1–28.
- JACOB, R., AND KARN, K. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises (Section commentary). *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*. Oxford: Elsevier Science.
- KARSH, R., AND BREITENBACH, F. 1983. Looking at looking: The amorphous fixation measure. *Eye Movements and Psychological Functions: International Views*, 53–64.
- LUNDQVIST, D., FLYKT, A., AND ÖHMAN, A. 1998. The Karolinska Directed Emotional Faces. *Pictorial face set available from the Department of Neurosciences, Karolinska Hospital, Stockholm, Sweden*.
- PRIVITERA, C., AND STARK, L. 2000. Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 9, 970–982.
- SALVUCCI, D., AND GOLDBERG, J. 2000. Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the symposium on Eye tracking research & applications*, 71–78.
- SANTELLA, A., AND DECARLO, D. 2004. Robust clustering of eye movement recordings for quantification of visual interest. *Eye Tracking Research & Applications (ETRA) Symposium*.
- WIDDEL, H. 1984. Operational problems in analysing eye movements. *Theoretical and Applied Aspects of Eye Movement Research*, 21–29.