Vision Research 57 (2012) 18–25

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

A simple nonparametric method for classifying eye fixations

Matthew S. Mould, David H. Foster *, Kinjiro Amano, John P. Oakley

School of Electrical and Electronic Engineering, University of Manchester, Manchester, UK

ARTICLE INFO

Article history:
Received 2 September 2011
Received in revised form 20 December 2011
Available online 2 January 2012

Keywords:
Eye movements
Natural scenes
Fixation classification
Nonparametric modeling
Expert classification
Individual differences

ABSTRACT

There is no standard method for classifying eye fixations. Thresholds for speed, acceleration, duration, and stability of point of gaze have each been employed to demarcate data, but they have no commonly accepted values. Here, some general distributional properties of eye movements were used to construct a simple method for classifying fixations, without parametric assumptions or expert judgment. The method was primarily speed-based, but the required optimum speed threshold was derived automatically from individual data for each observer and stimulus with the aid of Tibshirani, Walther, and Hastie's 'gap statistic'. An optimum duration threshold, also derived automatically from individual data, was used to eliminate the effects of instrumental noise. The method was tested on data recorded from a video eye-tracker sampling at 250 frames a second while experimental observers viewed static natural scenes in over 30,000 one-second trials. The resulting classifications were compared with those by three independent expert visual classifiers, with 88–94% agreement, and also against two existing parametric methods. Robustness to instrumental noise and sampling rate were verified in separate simulations. The method was applied to the recorded data to illustrate the variation of mean fixation duration and saccade amplitude across observers and scenes.

1. Introduction

It has long been known that the natural movement of an observer's gaze over a scene is discontinuous, with distinct locations being fixated sequentially, presumably reflecting their potential value to the observer (Yarbus, 1967). During such fixations, the eye is relatively stable, but between fixations it undergoes very rapid movements—saccades—which often have speeds of over 700 deg s^{-1} (Carpenter, 1988). Since little information is collected during saccades (Thilo et al., 2004) and it is the fixations themselves that are of principal interest in scene analysis, it is customary to summarize recordings of the point of gaze (PoG) by extracting just the fixations and discarding the remainder of the data.

Unfortunately, classifying fixations in practice can be difficult, and there is no standard method. Part of the difficulty lies in the fact that even during a fixation the eye continues to make small movements. These fixational eye movements include tremor, a small rapid movement of the eye with an amplitude between about 5 and 60 arcsec; also microsaccades, which are short involuntary eye movements with an amplitude between about 1 and 120 arcmin; and drift, which is a slow eye movement with an amplitude between about 1.2 and 30 arcmin (Martinez-Conde, Macknik, & Hubel, 2004). A typical eye-movement trace over one of the images used in this study is shown in Fig. 1. Although there are three candidate fixational areas, the eye clearly continued to move within each, and the demarcation of fixations and saccades within and between these areas is not easy to determine automatically, without additional assumptions.

Existing approaches to classifying fixations include methods based on gaze stability (or, in a complementary sense, dispersion or displacement) and on gaze speed (Salvucci & Goldberg, 2000). A stability-based method defines a sequence of eye movements as a fixation if the PoG remains within a circle of given radius (the stability threshold) for a given duration (the duration threshold) (e.g. van der Linde et al., 2009); analogously, a speed-based method defines a sequence of eye movements as a fixation if eye speed remains below a given value (the speed threshold) (e.g. Kienzle et al., 2009). The two methods may be combined by incorporating both stability and speed thresholds, sometimes also with an acceleration threshold (e.g. Tatler, 2007).

The absence of a standard method for classifying fixations presents a problem, as does the absence of commonly accepted threshold values for speed, acceleration, duration, and stability. It is known that different choices of algorithms and their input parameters can lead to systematic differences in reported fixations, and consequently markedly different interpretations of gaze data (Shic, Scassellati, & Chawarska, 2008).

* Corresponding author. Address: School of Electrical and Electronic Engineering, University of Manchester, Sackville Street, Manchester M13 9PL, UK.

E-mail addresses: matthew.mould@postgrad.manchester.ac.uk (M.S. Mould), d.h.foster@manchester.ac.uk (D.H. Foster), kinjiro.amano@manchester.ac.uk (K. Amano), john.oakley@manchester.ac.uk (J.P. Oakley).

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.12.006

Fig. 1. A typical eye-movement trace. The vertical scale bar at the top right indicates 1 deg visual angle and the inset at the bottom left shows a magnified version of the lower section of the trace. The scene image has been darkened locally to show the trace more clearly.

The uncertainty about appropriate parameter values is not surprising since the spatiotemporal characteristics of eye movements may vary across observers, stimuli, and tasks (Andrews & Coppola, 1999). Appealing to biologically plausible values, which are themselves subject to debate, does not resolve the problem, nor does the hand-tuning of thresholds based on visual inspection of PoG data, a procedure which assumes experimenter independence and expertise. In addition to these factors, the problem is complicated by the fact that different eye-tracking systems may themselves generate different levels of instrumental noise, which may, for example, come from errors in image sampling and processing by a video eye-tracker, from movement of the observer's head relative to the eye-tracker, and from environmental vibration. A more objective method for classifying PoG data would therefore be desirable, especially if it were nonparametric and accommodated both individual variation and noise variation in a natural way.

There have been previous attempts to develop data-driven approaches to classifying fixations, but none has succeeded in being completely nonparametric. A method proposed by Blignaut (2009) introduced a range of values for the optimum spatial threshold, but the associated temporal threshold was derived from previous studies and was fixed at 100 ms. A method described by Santella and DeCarlo (2004) used a clustering technique preceded by a so-called mean-shift procedure that required the use of a kernel function, typically a Gaussian function, but the parameters controlling its scale had to be selected by the user. A method proposed by Engbert and Kliegl (2003) needed little user input, but it depended on a speed threshold being derived from the data by a method that, although data-driven, required a choice of multiplier of the estimated noise levels. The method was later extended by Nyström and Holmqvist (2010), but it retained a parametric dependence. Two other speed-based methods, one proposed by van der Lans, Wedel, and Pieters (2011), another by Behrens, MacKeben, and Schröder-Preikschat (2010), which additionally used acceleration data, required, respectively, a choice of minimum fixation duration and a choice of low-pass filter to smooth the speed profile.

The aim of the present work was to exploit some of the more general distributional properties of eye movements to construct a simple, completely nonparametric method of classifying fixations. The method was primarily speed-based, but, by contrast with existing methods, the optimum speed threshold for classifying saccades—and therefore fixations—was derived automatically from the data for each observer and stimulus individually. The derivation was founded on Tibshirani, Walther, and Hastie's (2001) 'gap statistic' for identifying the optimum number of clusters in a set of data. Because speed-based methods can generate what appear to be unphysiologically short fixations (Nyström & Holmqvist, 2010), usually attributed to instrumental noise, the proposed nonparametric method was extended to include an optimum duration threshold—also derived automatically from the data—as part of the classification. The complete method was tested on data recorded with a video eye-tracker sampling at 250 frames a second from seven experimental observers performing a target-detection task in each of 20 natural scenes. Since a key feature of the proposed method was its adaptability to instrumental noise, its classification accuracy with increasing levels of added noise was verified in a separate simulation. The method was then applied to the recorded data to reveal how fixation durations and saccade amplitudes varied across observers and scenes, a variation that with some approaches might have been difficult to detect.

2. Experimental data

The data for analysis were obtained as part of a larger experiment examining the effects of scene structure on target detection and on eye movements in natural scenes (Amano et al., 2012). Only relevant detail is reproduced here.

2.1. Apparatus

Observers' PoG was recorded with an infra-red monocular video eye-tracker (High Speed Video Eyetracker Toolbox Mk 2, Cambridge Research Systems Ltd, Kent, UK), sampling at 250 frames a second, connected to a computer which stored and analyzed the output signals. Head movement was minimized with a forehead rest and a chinrest. Images were displayed on a 20-in. RGB CRT color monitor (GDM-F520, Sony Corp., Tokyo, Japan) controlled by a graphics workstation (Fuel V12, Silicon Graphics Inc., Mountain View, CA). The spatial resolution of the display was 1600×1200 pixels; the intensity resolution on each RGB gun was 10 bits; and the refresh rate was approximately 60 Hz. The display subtended approximately 17×13 deg at a viewing distance of 1 m.

2.2. Stimuli

In all, 20 static images of natural scenes, taken from Foster et al. (2006), were used in the experiment. They represented urban and rural settings, from near and far views. The target was a spectrally neutral shaded sphere subtending approximately 0.3 deg of visual angle at the viewing distance of 1 m and was matched in average luminance to its local surround (approximately 1 deg). In each trial, the target appeared with probability 0.5, and over successive trials, once in each of 130 possible locations, which were distributed uniformly over the scene. Repeated presentations were used to test for scene-specific effects (Amano et al., 2012; Kaspar & König, 2011). The mean luminance of the display was 3.6 cd m^{-2} (range 0–61.4 cd m^{-2}).

2.3. Observers

Seven observers, three male and four female, aged 21–31 years, took part in the experiment. All had normal color vision and normal binocular visual acuity. All except one of the observers (author MM) were unaware of the purpose of the experiment. The experimental procedure was approved by the University of Manchester Committee on the Ethics of Research on Human Beings, which

operated in accord with the principles of the Declaration of Helsinki.

2.4. Procedure

In each trial, an image with or without the target was presented for 1 s, and the observer then indicated by clicking a mouse button whether or not the target was there. Trials were performed in blocks of 260, all containing the same scene, but divided into four sub-blocks of 65 trials separated by a short break. Each sub-block lasted 5–10 min and not more than one block was performed in each experimental session of approximately 1 h.

Calibration data for the eye-tracker were collected at the start, middle, and end of each sub-block, as follows. Twenty calibration targets, arranged in a 5×4 grid, were presented on the screen at known positions. Observers were asked to click a mouse button as they fixated each target, and the corresponding PoG signals were recorded by the eye-tracker. Three sequences of 20 measured (x,y) gaze positions were obtained in each sub-block: $((x_{1,1},y_{1,1}),\ldots,(x_{1,20},y_{1,20}))$, $((x_{2,1},y_{2,1}),\ldots,(x_{2,20},y_{2,20}))$, and $((x_{3,1},y_{3,1}),\ldots,(x_{3,20},y_{3,20}))$. If for any of the three sequences a valid gaze-position signal for a target was unavailable owing to a loss in tracking, it was replaced by the corresponding gaze-position signal from another sequence (at least one replacement was necessary in 55% of sub-blocks). To transform to screen coordinates all the experimental gaze-position signals obtained between the first and second calibration measurements, the sequences $((x_{1,1},y_{1,1}),\ldots,(x_{1,20},y_{1,20}))$ and $((x_{2,1},y_{2,1}),\ldots,(x_{2,20},y_{2,20}))$ were averaged pointwise and fitted to the target positions, expressed in screen coordinates, by a linear (affine) transformation to give least-squares error. An analogous procedure was used to transform to screen coordinates all the experimental gaze-position signals obtained between the second and third calibration measurements.

For each of these calibration transformations, the RMS difference between the calibration targets and the positions of the observer's gaze on the screen provided an estimate of the calibration error (yielding two such error estimates for each sub-block). The mean value of the calibration error over observers and scenes was 0.26 deg and individual estimates did not exceed 0.51 deg. Similar levels of accuracy have been reported by Le Meur, Le Callet, and Barba (2007) and Kienzle et al. (2006). In total, data were collected from 5200 trials for each of the seven observers. The classification method was applied only to data from the 86% of trials in which there was no loss in tracking.

3. Classification method

The rationale for the proposed method of classification was that the eye reaches much higher speeds during saccades than during fixational eye movements—such as microsaccades and tremor—and that there are far fewer peaks in speed due to saccades than to fixational eye movements. These two properties of the data made possible the derivation of an optimum speed threshold that best separated the speed distributions of saccades and of fixational eye movements and instrumental noise. An optimum duration threshold was then derived that best separated the duration distributions of fixational eye movements and of instrumental noise.

3.1. Classification of saccades

Let $p_i = (x_i,y_i)$ be the PoG in screen coordinates for the ith video frame F_i. The PoG speed v_i for F_i was defined as the Euclidean distance $|p_{i+1} - p_{i-1}| = [(x_{i+1} - x_{i-1})^2 + (y_{i+1} - y_{i-1})^2]^{1/2}$ between the PoGs in the immediately preceding and following frames divided by the corresponding difference in time $|t_{i+1} - t_{i-1}|$; that is, $v_i = |p_{i+1} - p_{i-1}|/|t_{i+1} - t_{i-1}|$. This estimate is equivalent to averaging estimates based on the preceding and following successive differences.

A local maximum v_{max} in PoG speed was defined as a PoG speed that was greater than that in the immediately preceding and following frames. Some of these local maxima were due to saccades whereas others were due to fixational eye movements and instrumental noise. To find the optimum speed threshold for separating them, a variable speed threshold was introduced that was allowed to range between the lowest and highest recorded local speed maxima in the data set in 250 equal steps. The number of steps used was chosen to reflect the constraints imposed by the temporal resolution of the eye-tracker.

The light-gray histogram in Fig. 2 shows, for one observer viewing one scene over one block of 223 trials, the frequency of local maxima in PoG speed exceeding a variable speed threshold that ranged between the lowest and highest recorded local maxima (in this block 37 trials were excluded because of lost tracking). Local maxima exceeding high values of the threshold were attributed to saccades and those not exceeding low values of the threshold mainly to fixational eye movements. The much greater number of the latter was reflected in Fig. 2 by the sharp peak in the number exceeding threshold values near zero. The location of the elbow in the histogram was assumed to provide the speed threshold v_{opt} that optimally separated the distribution of saccades from the distribution of fixational eye movements and instrumental noise, that is, where the probability of misclassification was the least and the same for each.

To locate the elbow, the histogram was compared with a null distribution of local speed maxima. This null distribution (Tibshirani, Walther, & Hastie, 2001) had the property that values were uniformly distributed between the lowest and highest recorded local speed maxima. The dashed line in Fig. 2 shows for this null distribution the frequency of local speed maxima exceeding threshold as a function of threshold (since values were distributed uniformly, this frequency declined linearly). The gap between the frequency of local speed maxima exceeding threshold and the frequency of null-distribution local speed maxima exceeding threshold is actually traced by a dotted curve in Fig. 2, but it is largely hidden by the solid curve, which is explained shortly. The maximum in this

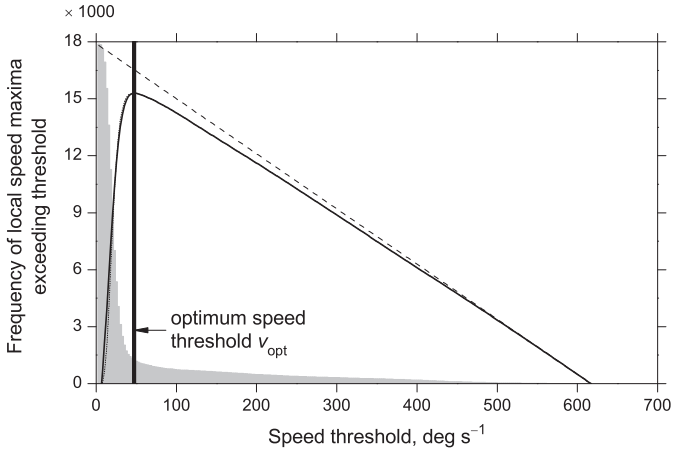

Fig. 2. Distribution of local maxima in point-of-gaze speeds. The light-gray histogram shows the frequency of local speed maxima exceeding a variable threshold that ranged between the lowest and highest recorded local speed maxima (bin width ~2.5 deg s^{-1}). The dashed line shows the frequency of local speed maxima exceeding threshold under the null distribution, according to which maxima are uniformly distributed. A dotted curve traces the gap between the two, but it is largely hidden by the solid curve, which shows the loess smooth of the gap statistic (Tibshirani et al., 2001). The vertical line marks the optimum speed threshold v_{opt}. Data were based on 223 trials by one observer viewing one scene.

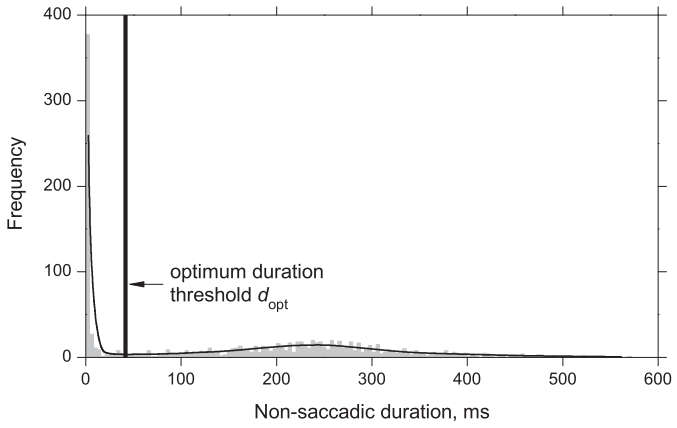

Fig. 3. Distribution of non-saccadic durations. The light-gray histogram shows the frequency of fixational durations and, at very short durations, instrumental noise events (bin width 4 ms). The solid curve shows the loess smooth of the histogram. The vertical line marks the optimum duration threshold d_{opt}. Data were based on the same trials as in Fig. 2.

gap statistic defined the location of the elbow and therefore the optimum speed v_{opt} for classifying saccades.

The precise location of the maximum in the gap statistic was estimated from the data by applying a locally weighted quadratic regression (loess) (Cleveland, 1979; Fan & Gijbels, 1996) to smooth the gap statistic. The closeness of the fit was determined by a bandwidth h which controlled the size of the local neighborhood and consequently the proportion of data included in each local fit (Fan & Gijbels, 1996). In general, if the bandwidth is too large, the loess fit is likely to be biased, and with it the location of the maximum; if the bandwidth is too small, the loess fit is likely to incorporate random fluctuations in the data, and therefore produce multiple local maxima. Accordingly, the optimum bandwidth h_{opt} was defined as the smallest value of h for which the number of local maxima in the smoothed curve was one.[1] To accommodate the high gradient at low speed thresholds, speeds were log-transformed before the fit was made.

The solid curve in Fig. 2 shows the loess smooth of the gap statistic and the vertical line marks the optimum speed threshold v_{opt}. Local speed maxima greater than v_{opt} were classified as being due to saccades; all other local maxima were classified as fixational or the result of instrumental noise.

3.2. Classification of noise

Fig. 3 shows for the same set of data as in Fig. 2 the frequency of durations classified as non-saccadic, that is, for which speed remained continuously less than or equal to the optimum speed threshold v_{opt}.

There is a local duration maximum at about 250 ms, a value that is broadly consistent with other estimates of fixation duration (e.g. Castelhano, Mack, & Henderson, 2009; Henderson, 2003; Rayner, 2009). There is also another local duration maximum, which is also a global maximum, at very short durations (a few ms). As indicated earlier, this maximum was taken to be due to instrumental noise events (Nyström & Holmqvist, 2010). These unphysiological durations have, in other studies, been eliminated by imposing a minimum duration threshold (Nyström & Holmqvist, 2010; Tatler, Baddeley, & Gilchrist, 2005). In keeping with the present nonparametric approach, the frequency of durations classified as non-saccadic was assumed to reach its minimum at the duration d_{opt} that optimally separated the noise distribution from the distribution of fixational eye movements, that is, where the probability of misclassification was the least and the same for each.

[1] This method should be contrasted with those methods based on estimating a bandwidth that makes the observed number of maxima the most probable (Silverman, 1986; Foster, 2002). The concern here was with the smallest bandwidth that satisfied this constraint.

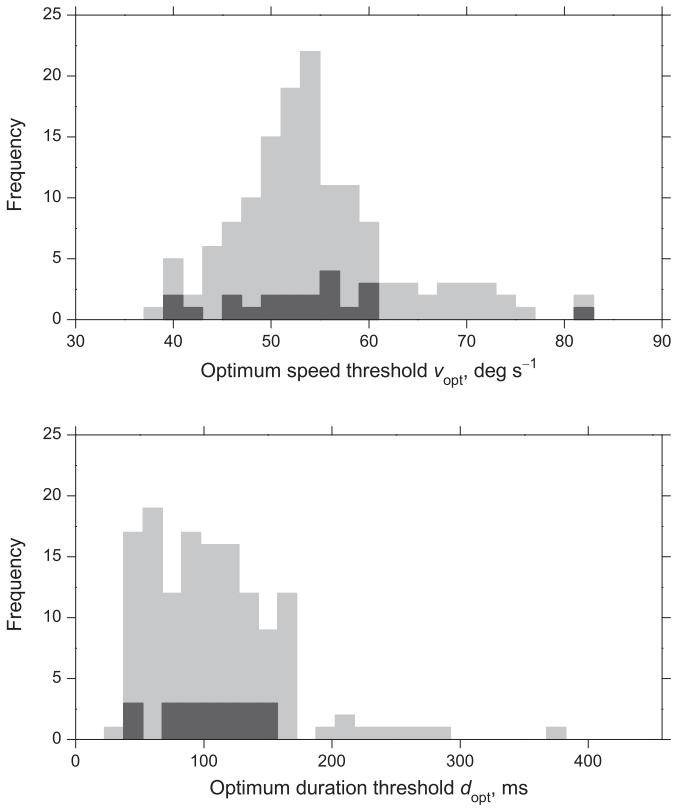

Fig. 4. Distribution of optimum thresholds. The light-gray histograms show the frequency of optimum speed thresholds v_{opt} (upper panel, bin width 2 deg s^{-1}) and optimum duration thresholds d_{opt} (lower panel, bin width 15 ms). Data were based on 140 blocks of 31,409 trials by seven observers viewing 20 scenes. The dark-gray histograms show the corresponding frequencies for blocks from which subsamples of 3–21 trials were taken for expert classification.

The location of the minimum was estimated from the data by again applying a quadratic loess to smooth the histogram. As with the speed data, the optimum value h_{opt} of the bandwidth was determined nonparametrically. Thus, the optimum h_{opt} was defined as the smallest value of h for which the frequency of maxima in the smoothed curve was two. To accommodate the high gradient at low durations, durations were also log-transformed before the fit was made.

The solid curve in Fig. 3 shows the loess smooth of the duration histogram and the vertical line marks the optimum duration threshold d_{opt}. Durations greater than d_{opt} were classified as being due to fixations; the rest were classified as being due to instrumental noise.

4. Comparison with expert visual classifications

To test its classification performance, the proposed method was applied to the sets of data described in Section 2. Estimates of the optimum speed threshold v_{opt} and optimum duration threshold d_{opt} were estimated independently and automatically for each observer viewing each image over each block of 260 trials (excluding, as noted earlier, trials with any lost tracking). The method could just as well have been applied to trials pooled over observers, scenes, or blocks. The light-gray histograms in Fig. 4 show the dis-

Table 1
Agreement (%) between fixation classifications by each experimental method and by three expert classifiers across 21 sample traces. RA classified all 21 traces (5271 frames); GRB classified three traces (753 frames); and EG classified 18 traces (4518 frames).

Experimental method	Expert visual classifier		
	RA	GRB	EG
Proposed nonparametric speed-based method	94	88	94
van der Linde et al. (2009) parametric stability-based method	92	85	95
Vig et al. (2009) parametric speed-based method	96	85	95

tribution of optimum speed and duration thresholds for the 20 scenes and seven observers tested.

A subsample of the resulting fixation classifications was compared with visual classifications made independently by three expert classifiers, RA, GRB, and EG.[2] Twenty-one PoG traces were selected at random from the 31,409 available (three traces from each observer, with each trace containing 251 frames over 1 s). All 21 traces were then classified by RA, and, owing to time constraints, three by GRB and 18 by EG. The agreement between classifications by the proposed method and those by the three expert classifiers was summarized by the proportion of frames with common classification. The total number of frames available from the 21 traces was 5271. Results are shown in Table 1, with classification accuracy ranging from 88% to 94% (first row of entries). An example of a trace with fixations classified by the proposed method and by one of the expert classifiers is shown in Fig. 5.

To help set the performance of the proposed method in context, classifications by each of two existing parametric methods, the one based on gaze stability, the other on speed, were also compared with the classifications by the expert classifiers. The stability-based method was due to van der Linde et al. (2009), in which a sequence of PoG positions was defined as fixational if they remained within a circle of radius 1 deg for 100 ms. These parameters were taken from recommendations by Applied Science Laboratories (Bedford, MA). The speed-based method was due to Vig, Dorr, and Barth (2009), in turn derived from work by Böhme et al. (2006), in which a sequence of PoG speeds was defined as saccadic with respect to two speed thresholds. Thus, when PoG speed first exceeded the higher threshold of 137.5 deg s^{-1}, saccade detection was initiated, and then the onset and offset of the saccade were defined, respectively, as the first samples where PoG speed rose above and fell below the lower speed threshold of 17.5 deg s^{-1}. These thresholds differed a little from those used by Böhme et al. (2006) as they were hand-tuned to match visually classified saccades (Vig, Personal Communication, 23 February 2010). The use of two speed thresholds was intended to increase the noise resilience of this method. The periods between saccades were assumed to contain only fixations (and possibly noise).

As with the proposed method, the agreement between classifications by the two parametric methods and by the three expert classifiers was summarized by the proportion of frames with common classification. Results are again shown in Table 1, with classification accuracy ranging from 85% to 96% (second and third rows of entries).

To provide an estimate of the best performance that could be expected from a fixation–classification method, the pairwise agreement between the three expert classifiers was summarized by the proportion of commonly classified frames. The mean agreement over the 24 comparisons was 93%. Since expert classification provides the best available estimate of a ground truth, this value sets an upper limit on average classification accuracy.

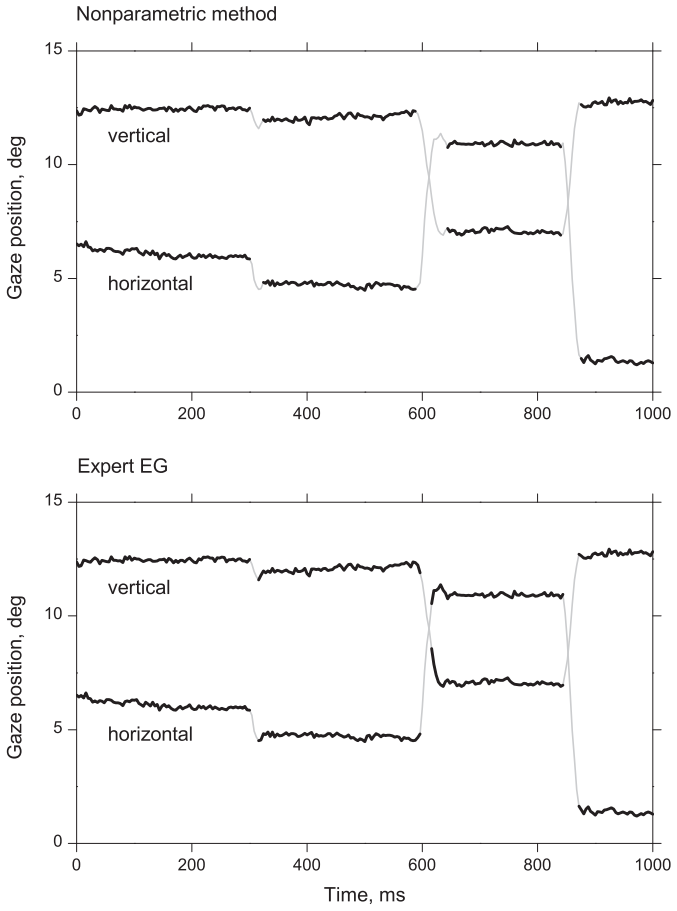

Fig. 5. An example classification of a typical eye-movement trace, sampling frequency 250 Hz. Vertical and horizontal gaze positions are plotted as a function of time from the beginning of the sample. Fixations classified by the proposed nonparametric method (upper panel) and by expert classifier EG (lower panel) are indicated by darker curves. The classification agreement between these two plots was 95%.

5. Robustness to added noise

One of the advantages of a classification method that derives thresholds automatically from data over a method that uses a fixed threshold is its adaptability to different levels of instrumental noise. If data are especially noisy, a speed-based classification method may misrepresent noise events as saccades. Misclassifications may be reduced by increasing the value of the speed threshold after visual inspection of the data, but these adjustments can be uncertain and time consuming, particularly if noise levels vary across observers.

To demonstrate the robustness of the proposed method to increased noise, a simulation was undertaken in which different levels of Gaussian noise were added to the recorded data before analysis by the proposed method. Fig. 6 shows for each noise level σ, the automatically generated optimum speed threshold v_{opt} averaged over all 140 blocks (upper panel). The threshold v_{opt} increased smoothly with increasing σ, reaching almost twice its value without added noise. Crucially, the total number of fixations classified by the method (lower panel) remained almost constant with σ.

[2] The three classifiers were Dr. R. Ackerley (Sahlgrenska Hospital and Institute for Neuroscience & Physiology, University of Gothenburg, Sweden), Professor G.R. Barnes (Faculty of Life Sciences, University of Manchester, UK) and Dr. E. Gowen (Faculty of Life Sciences, University of Manchester, UK). The lengths of their individual experience in the design, analysis, and interpretation of eye-movement experiments were, respectively, five, 40, and 11 years.

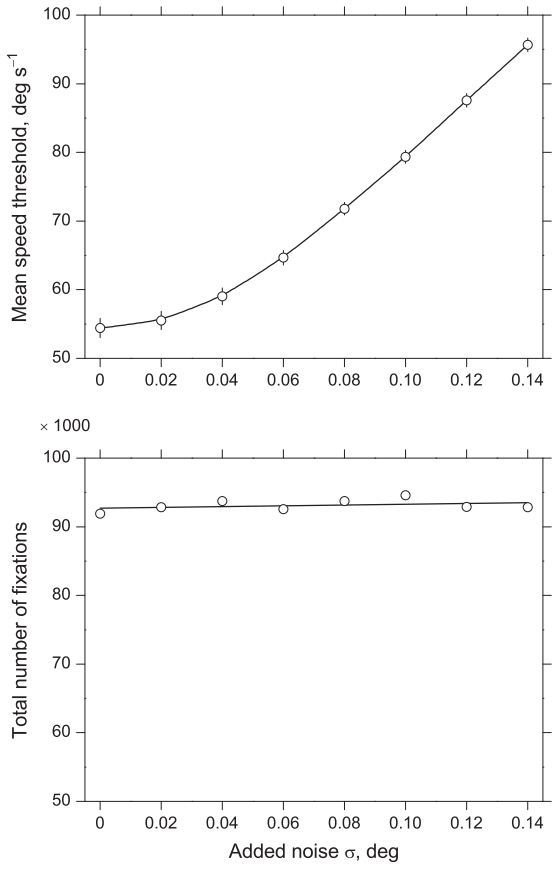

Fig. 6. Effect of added noise. Optimum speed threshold v_{opt} averaged over all 140 blocks (upper panel) and corresponding total number of fixations classified by the proposed nonparametric method (lower panel) are plotted against added noise σ in degrees of visual angle. The continuous curves are, respectively, local and global linear regressions. For clarity, the vertical bars (upper panel) show ±2 SEM.

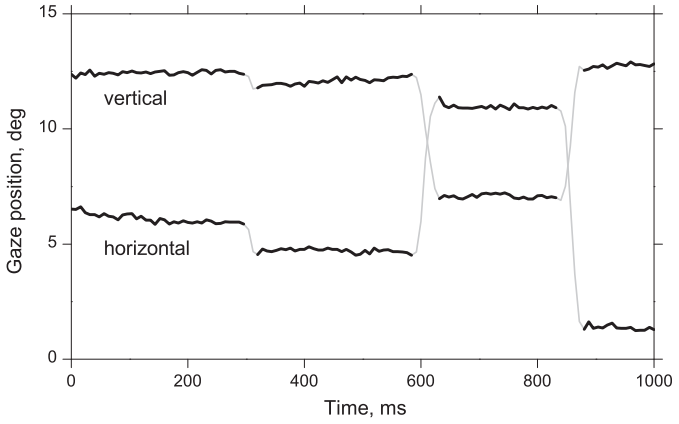

Fig. 7. An example classification of a typical subsampled eye-movement trace, sampling frequency 125 Hz. Vertical and horizontal gaze positions are plotted as a function of time from the beginning of the sample. Fixations classified by the proposed nonparametric method are indicated by darker curves (compare with Fig. 5, upper panel).

6. Robustness to sampling rate

Another advantage of the proposed method is its adaptability to different sampling rates (Andersson, Nyström, & Holmqvist, 2010). To demonstrate this robustness, a simulation was undertaken in which eye-movement traces were subsampled to reduce the sampling rate from 250 Hz to 125 Hz. The frequency of fixations classified by the method using the subsampled data was 99,063, compared with 91,907 classified by the method using the data at full resolution, showing that the frequency of fixations remained stable to within 8%. A further illustration of the robustness of the method to sampling rate is shown in the trace of Fig. 7. Data were subsampled from the eye-movement trace of Fig. 5 with fixations classified by the proposed method. Despite the reduction in sampling rate, it can be seen by comparison with the upper panel of Fig. 5 that there is little effect on the classification of fixations.

7. Revealing variation in fixation duration and saccade amplitude

The distributional properties of the thresholds used to classify fixations (Fig. 4) should not be confused with the distributional properties of the fixations themselves, which may vary with observer, stimulus, and task. Thus, differences have been reported in mean fixation duration across different observers and stimuli (Harris et al., 1988) and in mean fixation duration and in mean saccade amplitude across different observers and stimuli (Andrews & Coppola, 1999). But in the former case classifications were based on visual inspection of the PoG traces (Harris et al., 1988) and in the latter on a fixed displacement threshold and speed threshold (Andrews & Coppola, 1999).

A complication in such analyses is that the distributions of fixation durations and of saccade amplitudes are generally long-tailed (Harris et al., 1988; Velichkovsky et al., 2000), with correspondingly large standard deviations. In this study, SDs of fixation durations and saccade amplitudes averaged over observers and scenes were 119 ms and 2.42 deg, respectively. Given these large spreads, is it possible to reveal reliable differences in the means of the distributions across individual observers and scenes with the proposed method?

The dot plot of Fig. 8, taken from the analysis of Section 4, shows the ranked values of the mean fixation duration (upper panels) and mean saccade amplitude (lower panels) for all twenty scenes and for the three observers with the least, the median, and the greatest differences across scenes. The horizontal error bars show ±1 SEM. The variations in mean fixation duration and in mean saccade amplitude across observers and scenes are evident.

An analysis of variance confirmed the existence of highly significant effects of observer and scene, and their interaction, for both fixation duration (respectively, $F(6, 91767) = 454$, $p < 0.001$, $\eta_p^2 = 0.029$; $F(19, 91767) = 61.4$, $p < 0.001$, $\eta_p^2 = 0.013$; and $F(114, 91767) = 31.2$, $p < 0.001$, $\eta_p^2 = 0.037$) and saccade amplitude (respectively $F(6, 60662) = 1053$, $p < 0.001$, $\eta_p^2 = 0.094$; $F(19, 60662) = 15.3$, $p < 0.001$, $\eta_p^2 = 0.005$; and $F(114, 60662) = 21.1$, $p < 0.001$, $\eta_p^2 = 0.038$). Although the effect sizes represented by η_p^2 are modest, there are important differences across observers and scenes. For three of the 20 scenes, the ratios of largest to smallest mean duration and mean amplitude over all seven observers exceeded, respectively, 156% and 285%, and for one of the seven observers, the ratios of largest to smallest mean duration and mean amplitude over all 20 scenes were, respectively 181% and 252%.

8. Conclusions

The advantage of a nonparametric data-driven method for classifying eye fixations—and saccades—is that it automatically adapts itself to individual differences and to the effects of scene content and task. Unlike other methods, it requires no subjective hand-tuning of thresholds and therefore no intervention by an independent eye-movement expert. Because it is automatic, it can be applied to

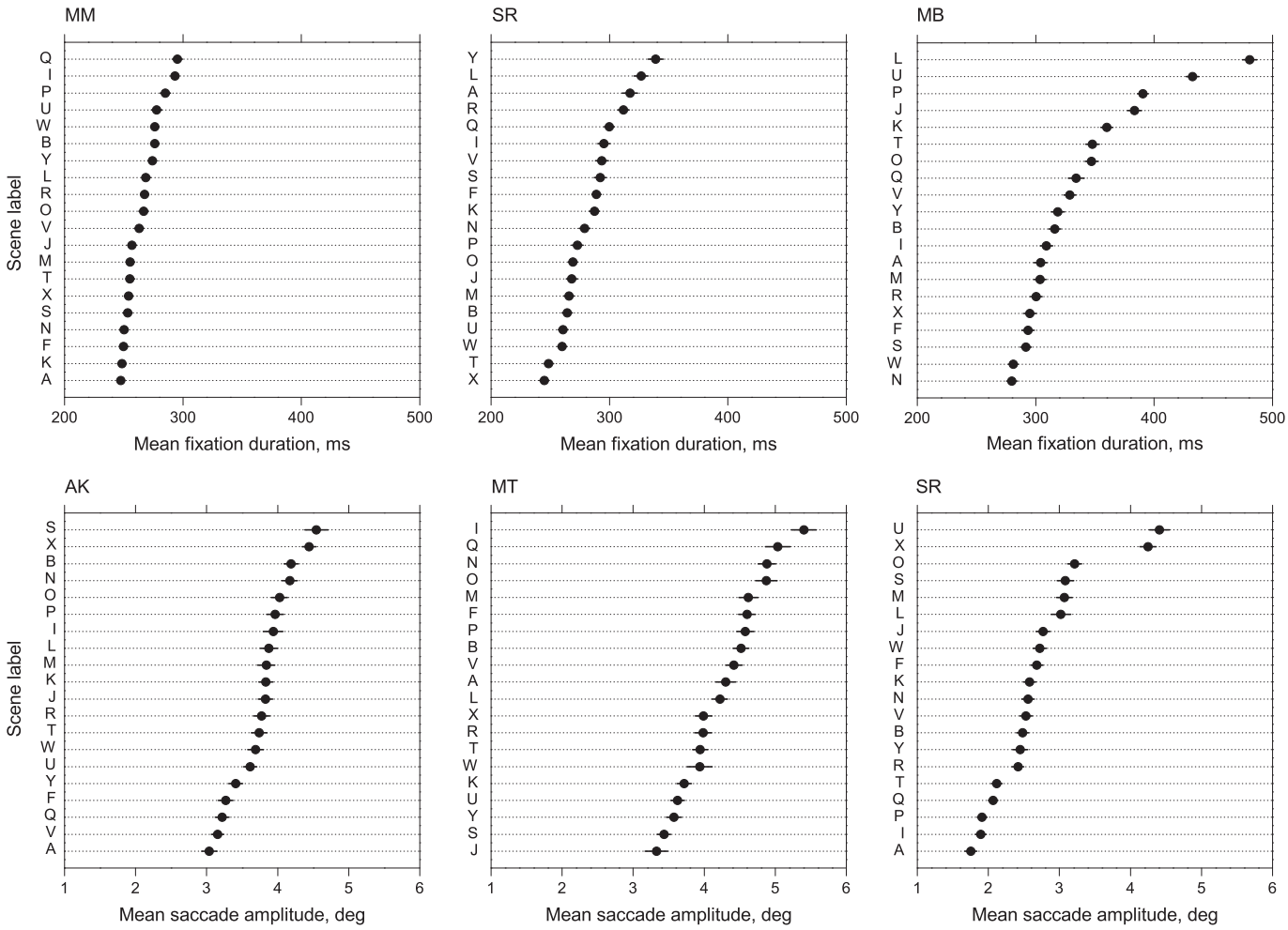

Fig. 8. Mean fixation durations (upper panels) and saccade amplitudes (lower panels) for each of three observers with the least, median, and greatest differences across twenty natural scenes. The same abscissae have been used to aid comparisons. Horizontal error bars indicate ±1 SEM where sufficiently large. The SDs of fixation durations and saccade amplitudes averaged over all observers and scenes were 119 ms and 2.42 deg, respectively.

large quantities of data, with corresponding gains in classification rate and subsequent analysis.

Although the approach adopted in this study was based on the general distributional properties of eye movements, only the existence of extrema in the distributions was assumed, i.e. a maximum or a minimum. No particular assumption was made about the model of the distribution or its parameters. Even so, as with any data-driven procedure, classification performance necessarily depends on the size of the sample. With the samples of 260 one-second trials used in the present analysis, extreme estimates of the optimum speed and duration thresholds demarcating fixations were only rarely encountered, as Fig. 4 made clear. Of course, reducing sample size is likely to lead to greater uncertainty in these estimates, although the goodness of fit illustrated in Fig. 2 suggests that rather smaller samples might still yield good estimates. As the method requires that examples of saccades and fixations be present in the data, minimum trial durations should be long enough to contain both.

The proposed method was found to be successful in that it produced fixation classifications that agreed with 88–94% of the independent classifications by the three expert classifiers. This level of agreement was generally as good as—and for some expert classifications better than—the level with the parametric methods due to van der Linde et al. (2009) and to Vig, Dorr, and Barth (2009). Significantly, none of the methods of classification agreed perfectly with the expert classifications, but given the level of agreement between expert classifiers, some differences were inevitable. From the simulations with added noise, however, it seems that unlike methods with fixed thresholds, classifications by the proposed method should remain stable as the level of instrumental noise increases. This would make the method particularly useful when applied to eye-movement data obtained from observers who find it difficult to remain steady (for example, young children).

As noted elsewhere (Żychaluk & Foster, 2009), there is a general understanding that a correct parametric method will always do better than a nonparametric one. This advantage comes from the fact that the parametric model is able to assume more about the data. At present, with uncertainty about the effects on fixations of scene content, task, and the variation from individual to individual, as illustrated in Fig. 8, a correct parametric model appears not to be possible. A nonparametric approach provides an objective and efficient way of accommodating this uncertainty. As demonstrated here, eliminating parametric dependence need not degrade classification performance in practice.

Software for classifying a sequence of PoG coordinates by the proposed nonparametric method is available from the authors.

Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council (Grant No. EP/F023669/1). We thank R. Ackerley, G.R. Barnes, and E. Gowen for their help in analyzing

the experimental data, E. Vig for advice, and E. Gowen and I. Marín-Franch for providing critical comments on the manuscript. A preliminary report of the results was presented at the Vision Sciences Society 11th Annual Meeting, Naples, Florida, USA, May, 2011.

References

Amano, K., Foster, D. H., Mould, M. S., & Oakley, J. P. (2012). Visual search in natural scenes explained by local color properties. *Journal of the Optical Society of America A – Optics Image Science and Vision, 29*, A194–A199.

Andersson, R. A., Nyström, M., & Holmqvist, K. (2010). Sampling frequency and eye-tracking measures: How speed affects durations, latencies, and more. *Journal of Eye Movement Research, 3*(3), 6, 1–12.

Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research, 39*, 2947–2953.

Behrens, F., MacKeben, M., & Schröder-Preikschat, W. (2010). An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters. *Behavior Research Methods, 42*, 701–708.

Blignaut, P. (2009). Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics, 71*, 881–895.

Böhme, M., Dorr, M., Krause, C., Martinetz, T., & Barth, E. (2006). Eye movement predictions on natural videos. *Neurocomputing, 69*, 1996–2004.

Carpenter, R. H. S. (1988). *Movements of the eyes*. London: Pion.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*(3), 6. doi:10.1167/9.3.6.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*, 829–836.

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research, 43*, 1035–1045.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman & Hall.

Foster, D. H. (2002). Automatic repeated-loess decomposition of data consisting of sums of oscillatory curves. *Statistics and Computing, 12*, 339–351.

Foster, D. H., Amano, K., Nascimento, S. M. C., & Foster, M. J. (2006). Frequency of metamerism in natural scenes. *Journal of the Optical Society of America A – Optics Image Science and Vision, 23*, 2359–2372.

Harris, C. M., Hainline, L., Abramov, I., Lemerise, E., & Camenzuli, C. (1988). The distribution of fixation durations in infants and naive adults. *Vision Research, 28*, 419–432.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*, 498–504.

Kaspar, K., & König, P. (2011). Viewing behavior and the impact of low-level image properties across repeated presentations of complex scenes. *Journal of Vision, 11*(13), 26. doi:10.1167/11.13.26.

Kienzle, W., Wichmann, F. A., Schölkopf, B., & Franz, M. O. (2006). Learning an interest operator from human eye movements. In: *Proceedings of the 2006 conference on computer vision and pattern recognition workshop* (CVPRW'06). Los Alamitos, CA, USA: IEEE.

Kienzle, W., Franz, M. O., Schölkopf, B., & Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision, 9*(5), 7. doi:10.1167/9.5.7.

Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research, 47*, 2483–2498.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience, 5*, 229–240.

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods, 42*, 188–204.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. T. Duchowski (Ed.), *ETRA 2000 – Proceedings of the eye tracking research and application symposium* (pp. 71–78). Palm Beach Gardens, FL, USA: ACM Press.

Santella, A., & DeCarlo, D. (2004). Robust clustering of eye movement recordings for quantification of visual interest. In A. T. Duchowski & R. Vertegaal (Eds.), *ETRA 2004 – Proceedings of the eye tracking research and application symposium* (pp. 27–34). San Antonio, TX, USA: ACM Press.

Shic, F., Scassellati, B., & Chawarska, K. (2008). The incomplete fixation measure. In K.-J. Räihä & A. T. Duchowski (Eds.), *ETRA 2008 – Proceedings of the eye tracking research and application symposium* (pp. 111–114). Savannah, GA, USA: ACM Press.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14), 4. doi:10.1167/7.14.4.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45*, 643–659.

Thilo, K. V., Santoro, L., Walsh, V., & Blakemore, C. (2004). The site of saccadic suppression. *Nature Neuroscience, 7*, 13–14.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B, 63*, 411–423.

van der Lans, R., Wedel, M., & Pieters, R. (2011). Defining eye-fixation sequences across individuals and tasks: The Binocular-Individual Threshold (BIT) algorithm. *Behavior Research Methods, 43*, 239–257.

van der Linde, I., Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2009). DOVES: A database of visual eye movements. *Spatial Vision, 22*, 161–177.

Velichkovsky, B. M., Dornhoefer, S. M., Pannasch, S., & Unema, P. J. A. (2000). Visual fixations and level of attentional processing. In A. T. Duchowski (Ed.), *ETRA 2000 – Proceedings of the eye tracking research and application symposium* (pp. 79–85). Palm Beach Gardens, FL, USA: ACM Press.

Vig, E., Dorr, M., & Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. *Spatial Vision, 22*, 397–408.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press (Tr. B. Haigh).

Żychaluk, K., & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics, 71*, 1414–1425.