A Model of Optimal Oculomotor Strategies in Reading for Normal and Damaged Visual Fields

Jean-Baptiste Bernard, Fermin Moscoso del Prado Martin, Anna Montagnini, Eric Castet

A MODEL OF OPTIMAL OCULOMOTOR STRATEGIES IN READING FOR NORMAL AND DAMAGED VISUAL FIELDS

Jean Baptiste BERNARD
I.N.C.M. (UMR 6193)
Université Aix–Marseille II & C.N.R.S.
Marseille, France
email: jean-baptiste.bernard@incm.cnrs-mrs.fr

Fermín MOSCOSO DEL PRADO MARTÍN
Laboratoire de Psychologie Cognitive (UMR 6146)
Université Aix–Marseille I & C.N.R.S.
Marseille, France
email: fermin.moscoso-del-prado@univ-provence.fr

Anna MONTAGNINI
I.N.C.M. (UMR 6193)
Université Aix–Marseille II & C.N.R.S.
Marseille, France
email: anna.montagnini@incm.cnrs-mrs.fr

Eric CASTET
I.N.C.M. (UMR 6193)
Université Aix–Marseille II & C.N.R.S.
Marseille, France
email: eric.castet@incm.cnrs-mrs.fr

ABSTRACT
We present an ideal observer analysis of single word reading in normal readers and central scotoma patients. Using this technique we are able to predict the spatio-temporal pattern of saccades in terms of pixels. This enables us to contrast theories that are impossible to compare using the traditional letter-slot approaches to modelling reading.

KEY WORDS
Scotoma, Ideal Observer, Reading, Computational Model

1 Introduction

From a low-level perspective, reading consists of a succession of fixations – each of which extracts information from a text image – interleaved with saccadic movements. From this perspective, a model of reading must provide an account of the spatio-temporal properties of these fixations, and of how these relate to the physiological properties of the eye. It is well-known that normally-sighted subjects read words by placing the maximal acuity zone of the retina (i.e., the *fovea*) on different locations of the words. However, patients with macular lesions in the center of the visual field (i.e., *central scotomata*), need to place the fovea outside of the word and use the peripheral zone of the retina (i.e., the *parafovea*) to be able to effectively extract information about the word.

Current clinical data are not sufficient to identify which are the oculomotor strategies that would optimize the reading performance of central scotoma patients. Results on the 'pseudo-fovea' used by these patients – their *preferred retinal location* (PRL) – are contradictory: On the one hand, some studies suggest that there is no correlation between reading performance and PRL ([1]). On the other hand, some authors argue that such a correlation exists and that it is best to place the scotoma above the word to be read (*vertical strategy*) rather than on the text line to be read (*lateral strategy*; [2])

All currently implemented models of eye fixation behaviour during reading, rely on the assumption that fixations must always be centered on the actual line of text to be read. This enables the computational simplification that fixations can be described in terms of letter position slots. Unfortunately however, models of this type are unsuitable to investigate the optimality of the lateral and vertical strategies described above, as it not even possible to represent the latter in this way (i.e., fixations occur mostly above or below that line of text). As a consequence, the only existing computational model of reading with scotoma, *Mr. Chips* ([3, 4]), directly *assumes* that the lateral strategy is optimal, but the fact remains that the lateral was in fact the only strategy that the model could follow.

Our purpose in this study is to obtain a mathematical description of the pattern of eye fixations that would be optimal for the subjects to follow. We operationalize optimality as the maximization of the amount of information about the word identity that one would expect to obtain with a certain pattern of eye-fixations. This is in turn made explicitely dependendent on the detailed properties of the retina. Our model describes predicted eye fixation behaviour at the level of individual image pixels. This permits predicted fixations to be centered either on or outside the actual text area.

2 Model Description

Humans are very apt in choosing the optimal course of actions in terms of the benefit they expect to obtain from them. Subjects performing tasks where an explicit gain or penalty (in score points) is introduced, choose optimal movement strategies with respect to their expected gain ([5]). Similarly, [6, 7] have shown that, in visual search tasks, subjects also optimize their eye movement strategies with respect to a gain function. In this case, the gain function was the relevant information that the subjects expected

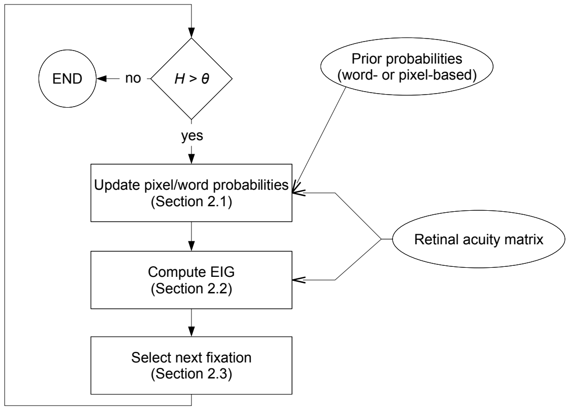

Figure 1. Schema of the model

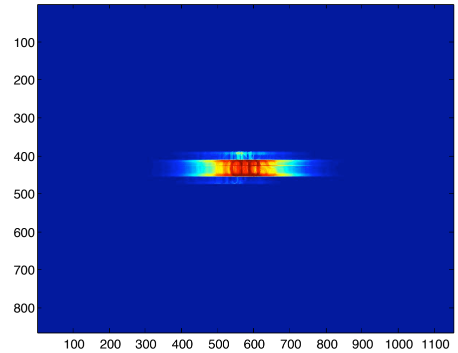

Figure 2. Pixel-based prior probabilities. Red indicates a high probability of a pixel being active (i.e., "black"). Blue areas correspond to a high probability of the pixel being inactive (i.e., "white")

to obtain by fixating on a particular point of an image, with respect to the task and the constraints imposed by the acuity of their visual fields.

Our approach to reading assumes that the optimal reading strategy is the one that optimizes what we term the *Expected Information Gain* (EIG) in each fixation. Depending on what is considered as information, the EIG can be defined in slightly different ways. On the one hand, we could consider a – suboptimal – model where the goal is to identify the image pixel values, without taking into consideration that the pixels must form letters and eventually words. In this case, the EIG would be measured in terms of information on the *individual pixel values*. On the other hand we could consider that our task in reading is the identification of words from the image, and thus introduce top-down information about the words that the image should contain. In thi second alternative, the EIG would be defined in terms of the information about the actual *word identity*. Although the latter word-based strategy would be the optimal one, some support can be found in the literature for suboptimal reading strategies that do not consider lexical top-down information ([8]). In order to consider these two possibilities, we will use both modelling strategies: a suboptimal pixel-based strategy lacking any top-down information, and an optimal word-based strategy where top-down information strongly constrains the possible images.

Figure 1 summarizes the three main steps in the model we propose. After a fixation (initially fixed at the center of the display due to the fixation cross), the model updates its probability distributions of pixel values (depending on the degree of top-down information used in the model this can either be done directly at the pixel level, or through a mediating lexical level). This is done by combining the retinal acuity matrix centered on the fixated point, with the image pixel values. This results in a noisy sample from the actual image, with the level of noise depending on the visual acuity at each particular pixel. This sample is combined with the previous knowledge about the image obtained through the previous fixations (initially the prior expectations). Further detail on this initial stage can be found in Section 2.1.

In the second step, the EIG for each possible next fixation is computed. For this purpose, the effect of a subsequent fixation in each possible point is evaluated. The pixel probabilities are transformed into pixel-level mutual informations (either about pixel values or word identities). These mutual information are combined (i.e., convoluted) with the retinal acuity matrix to obtain an estimate of how much information would be obtained by fixating each point of the image. See Section 2.2 for more details on this step.

Finally, in the last step, the EIG distribution obtained in the previous step is normalized into a probability density function. This implements the assumption that the probability of fixating a point is directly proportional to the EIG from fixating that point. The next fixation is then sampled from this probability distribution (one could also choose the maximum of this distribution, but this would result in a deterministic strategy, that does not correspond well to human behavior). Section 2.3 provides more details on this. These three steps are iterated until a predefined level of certainty (θ) about the value of the pixels (or the identity of the word) is reached.

2.1 Updating the Probability Distributions

Before each fixation, the model has a prior expectation on the possible color values (black or white) of the pixels in the screen. This prior expectation corresponds to the information we have obtained by the previous fixations – or just to the overall prior of the model, if there have not been any fixations yet. We will refer to this pixel-based prior after the k-th fixation as $P^{(k)}$. This expectation is a matrix whose elements are the the probabilities that each pixel takes the value of 1. The prior $P^{(0)}$ represents the probability of a pixel being active before obtaining any information through fixations. As for the moment we will only consider the situation where words are presented in a constant font at the middle of the screen, this prior will be the sum of the images corresponding to the 30,000 most frequent French words weighted by the corresponding word frequencies. Figure 2 illustrates how this prior looks like in our experiments.

In order to update this matrix using the visual information, we resort to Bayes' theorem. The probability that point p_j is active after fixating on point i is estimated as:

$$P_j^{(k+1)} = P_j^{(k)} \frac{P(d_{i,j}|p_j = 1)}{P(d_{i,j}|p_j = 0) + P(d_{i,j}|p_j = 1)}, \quad (1)$$

where $d_{i,j}$ is the value of point j that results from centering the acuity matrix at point i of the image, and adding noise in each point in inverse proportion to the level of acuity at that point (see Figure 3 for the acuity matrices that we used). The likelihood in this equation is calculated as a coming from a Bernoulli trial with the corresponding amount of noise.

2.2 Computation of the EIG

As mentioned above, we define the EIG as the mutual information between the probability distribution of pixel values (using the current estimate at each point), and the word identity.[1] More precisely, we consider the mutual informations (in the strict bi-variate sense) between each individual pixel and the word identity. These mutual informations are summed and weighted by the acuity matrix for each possible fixation. The weighted sum is easily computed as the convolution (performed in Fourier space) between the matrix of mutual informations for each pixel, and the retinal acuity matrix. Note that in a word-based strategy this will result in an overestimation of the total mutual information, as much of the information provided by one pixel is redundant with the others. A direct estimation of the amount of redundancy in each pixel is difficult to obtain. However, the mutual information between pixels in natural images decreases as a power-law of their distance ([10]), and this applies also to sequences of letters in running text ([11]). Therefore, we can correct our estimation by de-convoluting the resulting information matrix with a power-law filter with wider horizontal than vertical covariance (this is to account for the mutual information between pixels being larger within the same line of text). The application of this filter results in a high-passed version of the matrix of pixel-word mutual informations (with a stronger horizontal component).

As pixel values univocally determine word identies (we use constant fonts, sizes, and word locations) the mutual information between words and individual pixel values reduces to the plain entropy of each pixel. Thus it is easy to convert the probability matrix $P^{(k)}$ into the corresponding matrix of individual mutual informations $I^{(k)}$ at fixation k:

$$I^{(k)} = -P^{(k)} \log_2 P^{(k)} - (1 - P^{(k)}) \log_2 (1 - P^{(k)}). \quad (2)$$

In order to compute the EIG for the next fixation ($EIG^{(k+1)}$), in the pixel-based approach we only need to convolute the matrix of pixel-word mutual informations ($I^{(k)}$) with the corresponding acuity matrix (A). In the word-based approach an additional correction for redundancy is obtained by de-convoluting the result with the filter described above.

2.3 Selection of the next fixation

The expected information gain matrix ($EIG^{(k+1)}$) represents our estimation of the gain in information that will be obtained by fixating in each point of the screen. Maximizing this gain can be done in two ways. An option could be picking directly the maximum of $EIG^{(k+1)}$ as the next point to fixate, leading to a deterministic (maximum posterior) strategy. Alternatively one can sample from $EIG^{(k+1)}$ as if it were a probability distribution (after a normalization by its sum). This presents a non-deterministic strategy, which is more suitable to model non-deterministic human data, and still converges to an optimal strategy. Note that this non-deterministic strategy is equivalent to saying that the probability of fixating a particular point is directly proportional to the information we expect to obtain from it, thus more informative points will be sampled more often.

Repeated sampling from a probability distribution presents the disadvantage of a great unstability. A different point will be selected in each cycle of the algorithm (the probability of changing location asymptotes to one with growing image resolution). Ideally, we would want some points to remain fixated longer than others, as is the case in humans. This can be accounted for by introducing an additional cost for movement. During time when the eye is being moved, no information is acquired by the system. Therefore in a really optimal strategy the system would take this into account by evaluating at each point whether it is likely to obtain *more* information by moving than by just remaining on the same location, thus saving the cost of an eye movement when it is not likely to be advantegeous. Formally, if at time k we are fixating at point i, the condition that must be satisfied in order to move is:

$$\alpha EIG_i^{(k+1)} < \mathbf{E}(EIG^{(k+1)}), \quad (3)$$

where $\alpha \geq 1$ is a free parameter representing a 'conservativeness' bias. This bias ultimately reflects the time that is spent moving (which would be spent obtaining information if we did not move). The operator $\mathbf{E}(x)$ refers to the expectation of x. The expectation of the EIG after sampling is equal to the sum of the EIG at each point j ($EIG_j^{(k+1)}$) weighed to the probability that the next point to be fixated

[1] Strictly, mutual information is not defined between more than two variables. More precisely, what we approximated was the generalization of mutual information to the multivatiate case that [9] introduced as *total correlation*

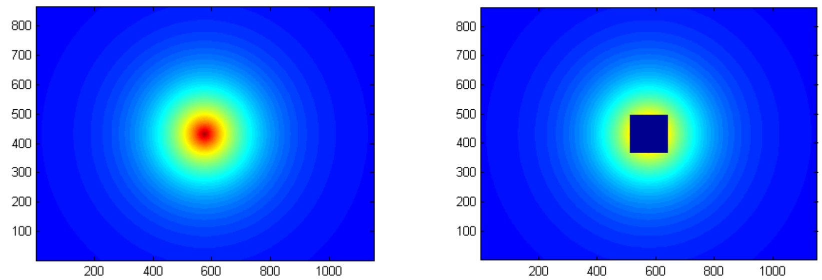

Figure 3. Retinal acuity matrices for normal (left panel) and simulated central scotoma (right panel). Red areas indicate zones of very high sensitivity (i.e., the fovea), while blue represents ‘blind’ areas. Notice that, in the scotoma the most sensitive areas correspond to the parafovea, and still these are not very sensitive relative to the normal maximal acuity zones.

is j $(P(j))$. In fact, as described in the previous section, the probability of fixating each point in the screen will be proportional to the EIG itself:

$$P(j) = \frac{EIG_j^{(k+1)}}{\sum_j EIG_j^{(k+1)}}. \qquad (4)$$

Therefore the expectation of the EIG is readily computed as:

$$\begin{aligned} \mathbf{E}(EIG^{(k+1)}) &= \sum_i \frac{EIG_i^{(k+1)}}{\sum_j EIG_j^{(k+1)}} \cdot EIG_i^{(k+1)} \\ &= \frac{\sum_j \left(EIG_j^{(k+1)}\right)^2}{\sum_j EIG_j^{(k+1)}}. \end{aligned} \qquad (5)$$

3 Results and Discussion

Figure 4 illustrates the distributions of predicted fixations that one obtains using the method described above (in a pixel-based strategy). The most apparent difference between the normal retina and the central scotoma case is that, while in the normal case fixations would mostly happen directly on the word, most fixations in the scotoma condition would fall either above or below the actual word, with only a few of them falling on the sides. Thus, according to our analysis, the optimal reading strategy in scotoma would be the “vertical” one mentioned in the introduction, which is strongly preferred over the “lateral” strategy (which is also present but in a much lesser degree). This strategy is preferred across all stages of the recognition process, from the very early ones to the last ones. Thus, an ideal observer analysis of (single word) reading, provides support for the “vertical” strategy, consistent with the experimental results of [2].

The graph in Figure 5 shows the predicted reading latencies (measured in fixation cycles, which may or may not correspond to actual different fixations, depending on

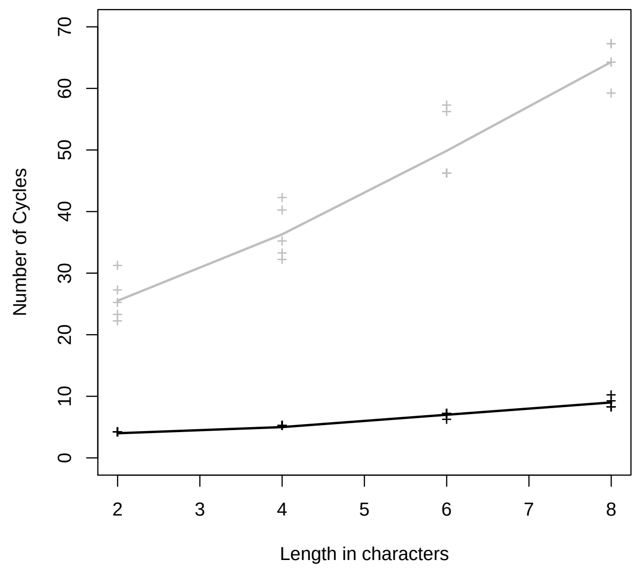

Figure 5. Comparison of recognition times (number of cycles in the system) as a function of word length for the scotoma case (grey line) and normal case (black line). The lines represent non-parametric regressions.

the condition) for the normal and scotoma cases as a function of word length. Two issues are noteworthy. First, the scotoma case is predicted to be overall much slower than than the normal retina case. Second, although both cases are strongly affected by word length in a roughly linear manner, with longer words being slower to be recognized, this effect is much more pronounced in the scotoma case. Both of these predictions are consistent with experimental results.

We have presented a simple ‘ideal-observer’ analysis of single word reading that is able to model fixation loca-

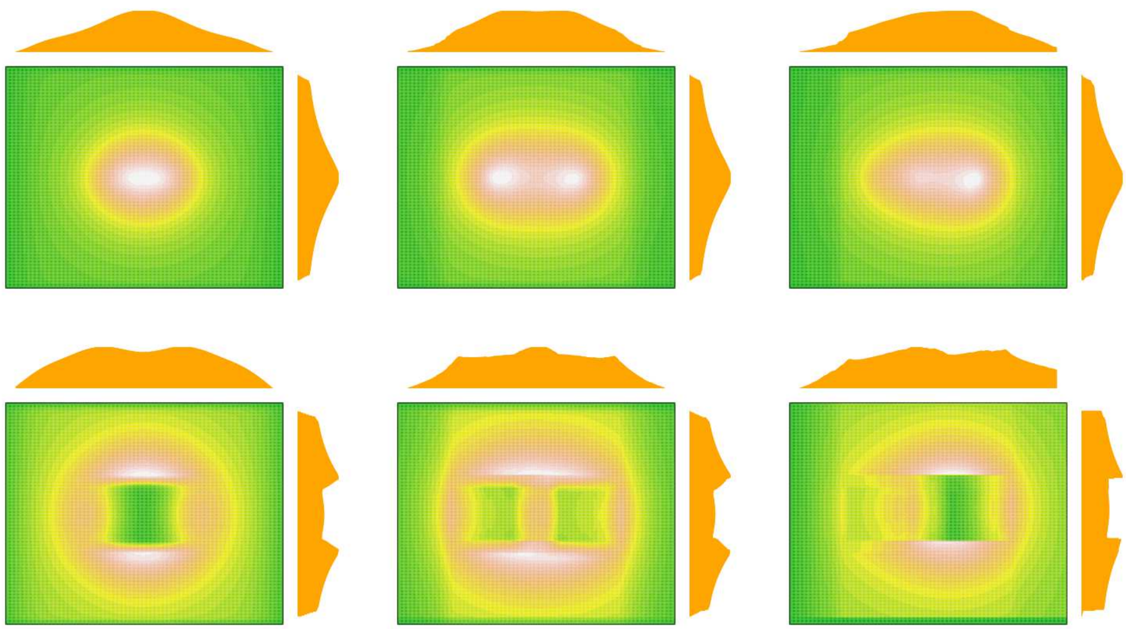

Figure 4. Probability density functions of fixation locations predicted by the model (pixel-based) for the word “responsable”. The upper panels depict the fixation distributions using the “normal” retinal acuity matrix, while the distributions obtained using the “simulated scotoma” retinal acuity matrix are shown in the lower panels. The leftmost column plots the distribution of predicted first fixations (after the initial fixation at the center). The mid column plots the distribution of fixation midway through the recognition process (5 fixations for normal and 21 for scotoma). The rightmost panels plot distribution of predicted fixations after the last one. The histograms on the margins of each panel show the horizontal and vertical marginal probability density functions.

tions and recognition latencies for both normal readers and central scotoma patients. Our analysis supports that, in the single word case, a vertical reading strategy is preferable for central scotoma patients, consistent with the results presented in [2], and not with those of [1]. Despite being overall succesfull, our analysis also fails to account for some additional facts reported in the literature. These differences between the model and the actual human behaviors are of great interest. In a sense, the ideal observer methodology presents the behavior that participants should show if they were following an optimal strategy. It is the deviations from optimality that present evidence for the need of additional neurophysiological constraints. Of particular interest in our case is that our analyses it appears that – for the first fixation – both the lateral and vertical strategies should be symmetrical (equal preferences for above or below and right or left of the word). However, actual scotoma patients (and experimental participants in 'simulated scotoma' experiments) tend to show a slight preference for PRLs respectively to the left and below the scotoma (in the visual field). This may suggest additional mechanisms in the system or, alternatively, a modification of the priors (for instance to account for the fact that reading in French mostly involves following the text left-to-right, top-down in a page).

The model we have presented has additional predictive power, as it enables us to estimate how much information about word identity a subject has obtained through a particular fixation (instead of sampling from the distributions, actual experimental data can be fed into the model to compute on-line the optimality of their movements). This enables us to compare the goodness of fit to experimental data of different strategies. For instance, we can evaluate to which extent does top-down lexical or letter level information plays a role in the determination of eye movements.

References

[1] Donald C. Fletcher, and Ronald A. Schuchard. Preferred retinal loci relationship to macular scotomas in a low-vision population. *Ophthalmology*, 104:632–638, 1997.

[2] Keziah L. Petre, Charlotte A. Hazel, Elisabeth M. Fine, and Gary S. Rubin. Reading with eccentric fixation is faster in inferior visual field than in left visual field. *Optometry & Vision Science*, 77:34–39, 2000.

[3] Gordon E. Legge, Timothy S. Klitz, and Bosco S. Tjan. Mr. Chips: An ideal-observer model of reading. *Psychological Review*, 104:524–553, 1997.

[4] Gordon E. Legge, Thomas A. Hooven, Timothy S. Klitz, J. Stephen Mansfield, and Bosco S. Tjan Mr. Chips 2002: New insights from an ideal-observer model of reading. *Vision Research*, 42:2219–2234, 2002.

[5] Julia Trommershäuser, Michael S. Landy, and Laurence T. Maloney. Humans rapidly estimate expected gain in movement planning. *Psychological Science*, 17:981–8, 2006.

[6] Wilson S. Geisler, Jeffrey S. Perry, and Jiri Najemnik. Visual search: the role of peripheral information measured using gaze-contingent displays. *Journal of Vision*, 6:858–73, 2006.

[7] Jiri Najemnik and Wilson S. Geisler. Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8:1–14, 2008.

[8] Denis G. Pelli, Bart Farell, and Deborah C. Moore. The remarkable inefficiency of word recognition. *Nature*, 423:752–756, 2003.

[9] Satosi Watanabe. Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, 4:66–82, 1960.

[10] Daniel L. Ruderman. The statistics of natural images. *Network: Computation in Neural Systems*, 5:517–548, 1994.

[11] Werner Ebeling and Thorsten Pöschel. Entropy and long-range correlations in literary English. *EPL (Europhysics Letters)*, 26:241–246, 1994.