To appear in *Proceedings of the 15th Annual Meeting of the Cognitive Science Society*, Boulder, CO.

A Connectionist Attentional Shift Model of Eye-Movement Control in Reading

Ronan Reilly
Department of Computer Science
University College Dublin
Belfield, Dublin 4, Ireland
rreilly@ccvax.ucd.ie

Abstract

A connectionist attentional-shift model of eye-movement control (CASMEC) in reading is described. The model provides an integrated account of a range of saccadic control effects found in reading, such as word-skipping, refixation, and of course normal saccadic progression.

Theoretical background

The bulk of research on eye-movement control in reading suggests that the processes controlling the *when* and the *where* of eye movements operate relatively independently; word identification appears to determine the *when* of most of the forward movement of the eyes, while low-level oculomotor factors are the main influence on where in a word the eye lands. Nevertheless, the processes controlling the *when* and *where* must interact at some level, and a number of attempts have been made to provide a coherent account of the dynamics of this interaction (McConkie, 1979; Morrison, 1984; Rayner & Pollatsek, 1989; O'Regan, 1990).

Connectionism provides a convenient framework for integrating information from several domains (e.g., language and vision) and would therefore seem well suited to the task of modelling eye-movements in reading. The model described here, CASMEC, is primarily an integration and computational implementation of the informal models of Morrison (1984) and O'Regan (1990). Morrison's proposal can be sketched out broadly as follows: Assume that the word currently fixated is word *n*. In the normal course of events this word will be correctly identified and attention will shift to word *n+1*. Note that foveation and allocation of visual attention are assumed to be decoupled. The process of shifting attention to the next word automatically results in the programming of a new saccade. In most cases, this program is executed. However, if the shift in attention has been of sufficiently long duration to allow the identification of word *n+1* without the need to foveate it, three possibilities arise: The first is that word identification takes place, the programmed saccade is cancelled, and attention shifts to word *n+2*. A new saccade is then programmed and subsequently executed. The second possibility is that identification occurs too late to delay the execution of the saccade to word *n+1*. In this case, a saccade to word *n+1* is rapidly followed by a saccade to word *n+2*. The third possibility is that the saccadic program is modified, so that the resulting saccade causes the eye to land somewhere between word *n+1* and word *n+2*. Within this framework, one can account for the skipping of high-frequency words (i.e., readily identifiable words), saccades that land between words, and the occasional very brief fixation. The attentional shift mechanism is also a way of explaining preview effects. These occur when the encoding of a word in the current fixation benefits from it having been attended on the preceding fixation. There is a considerable amount of evidence supporting the integration of some form of information across saccades which facilitates the encoding of the subsequently fixated word in both reading and non-reading tasks (Rayner & Pollatsek, 1989).

The other major element of CASMEC is based on the work of O'Regan (1990). He proposed a model of eye-movement control which is a function of low-level oculomotor constraints and lexical processes. In his Strategy-Tactics model, the eye moves forward in careful word-by-word reading, using low-spatial frequency cues to aim at the optimal viewing position (OVP) of the next word (somewhere to the left of its centre). O'Regan and his co-workers (O'Regan, Levy-Schoen, Pynte, & Brugalliere, 1984) identified the OVP as a particular location in a word where both speed of recognition and likelihood of refixation are at their lowest. Aiming at the OVP represents an overall strategy which gives way to a

within-word tactic to maximise the amount of information picked up once a word has been fixated. If the eye fails to land near a word's OVP, a typical tactic, according to O'Regan, is to saccade to the other end of the word rather than to the middle, thus maximising the combined information from both fixations.

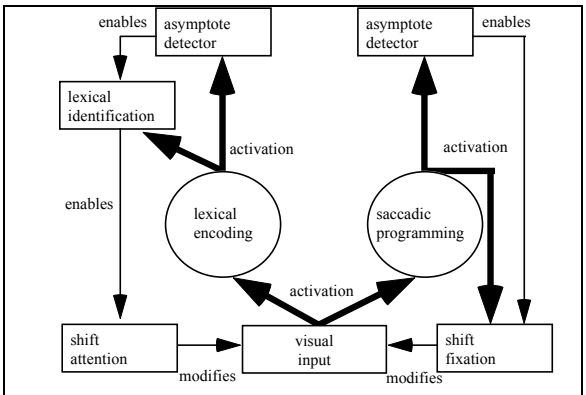


Figure 1: The connectionist attentional shift model of eye-movement control. The circles represent connectionist modules and the rectangles non-connectionist control modules. Thick lines indicate a flow of activation, thin lines a flow of control. The asymptote detectors determine when the cascading outputs from the lexical and saccadic modules have reached asymptote.

Implementation details

An overview of CASMEC is given in Figure 1. The visual input is processed along two main pathways, the first dealing with word recognition and the second dealing with saccadic programming. The two modules in circles represent the components of the framework that are trained using the backpropagation learning algorithm (Rumelhart, Hinton, & Williams, 1986), and both consist of single hidden-layer feedforward networks with trainable weights. The modules in rectangles are non-connectionist and are used to manipulate the inputs and outputs of the trainable modules in ways that will be described in more detail below. The thick lines connecting some of the modules represent the transmission of activation values, and the thin lines represent the transmission of triggering or enabling signals.

Visual Input Matrix

The visual input consists of a 26x20 matrix in which the rows represent letters and the columns represent spatial locations. The input matrix is intended to be analogous to a low-level cortical representation. The effects of the non-homogeneity of receptor density in the retina is represented in two ways: First, moving outward from the area projected to by the fovea, there is a decrease in the spatial resolution of letters. Second, there is a commensurate decrease in the accuracy of letter categorisation. Both these representational assumptions are well supported in the psychophysics research literature (Chastain, 1982; Levi, Klein, & Aitsebaomo, 1985).

The decrease in spatial resolution is implemented by means of a set of Gaussian distributions of varying standard deviation. These will be referred to in functional form as $y = G(\sigma, x)$, where σ is the standard deviation and x is a term corresponding to relative spatial location. Column 8 in the visual input matrix was chosen as the centre of the area projected to by the fovea. The activation of a single unit in this column represents the presence of the letter it represents. Its activation value is given by $G(0.25,0) = 1.6$. Moving away from the centre, to the columns on either side, σ increases by a fixed amount, which results in a decrease in the height of the distribution, and thereby the activation level of the units in the column. Furthermore, as the height of the distribution decreases the leakage of activation to the same letter unit in adjacent spatial locations increases. The rate at which σ increases, and consequently the rate at which the level of unit activation decreases, is based on the linear equation (due to O'Regan, 1990): $r'=r'_o(1+m\phi)$ where r' is the ratio of the acuity (in this case, level of unit activation) at some eccentricity ϕ over the acuity at the centre of the fovea, r'_o is this ratio for the centre, and m is a constant which reflects the rate of increase in the size of the cortical receptive fields as ϕ increases. Each spatial location represents an increment in ϕ of 0.25^o (i.e., four letters to a degree). A value for m of 1.6 was chosen because it gave a convenient σ increment of 0.1 and was close to the value of 1.7 estimated by O'Regan (1990) for reading on the basis of a range of psychophysical experiments.

The value x determines the amount of activation that leaks into adjacent columns of the visual matrix. On the assumption that there is perfect spatial resolution at the centre, the increment to x associated with one character space (i.e., one column) was chosen to correspond to the point at which $G(0.25,x)=0.001$; in other words, where the leakage of activation from the central location to the immediately adjacent locations is negligible. The value of x chosen using this criterion was 1.0.

For each spatial location, a Gaussian was used to represent the degree of category certainty. As one moves further away from the centre, σ is incremented, resulting in a decrease in activation for the relevant letter unit and an increase in the leakage of activation to category (as opposed to spatial) neighbours. Thus, the unit representing "a", say, activates units for visually similar letters, and does so to an increasing extent as one moves away from the centre. Visual similarity was determined by a cluster analysis of the pixel representation of a standard font.

Attentional Mechanism

A key role in CASMEC is played by visual attention. This process is operationalised by a movable inverted "spotlight" which suppresses the activation of part of the visual representation while leaving the attended area at its normal level (cf. Mozer, 1991). The neurophysiological motivation for this comes from Crick's (1984) proposal for an attentional mechanism of this sort operating in the area of the thallamus. In the implementation, the activity of all non-attended regions of the input is multiplied by 0.25. This figure was chosen to be small enough to give words that were attended to, but not foveated, a chance to compete with the foveal input. It also had to be small enough to provide the saccade-targeting mechanism with a relatively noise-free target.

Lexical Encoding Module

The internal architecture of the lexical module is a fully-connected feedforward network with a 26x16 input units, 150 hidden units, and eight output units. The input to the module comes from the central 16 columns of the visual input matrix and is modified by the attentional spotlight, which dampens down the activation of non-attended words. Eight output units are used to represent each of the 222 words in the training corpus (described below). Words that are visually similar are given similar lexical codes.

Within a larger reading model, the lexical module would make a lexical representation available to higher-order processes. Here, however, it simply serves to store the sequence of identified words and enable a shift in attention.

Saccadic Control Module

The input to the saccadic programming module is also derived from the visual input matrix. Since low-spatial frequency information appears to be used in targeting saccadic eye-movements, the visual input *matrix* is transformed into a *vector* by collapsing over the category dimension. The elements of the resulting vector correspond to the 20 spatial locations, and the value for a given element is the maximum activation value in the collapsed column for that location.

The internal architecture of the saccadic module is a standard feedforward network. There are 20 input units, 15 hidden units, and two output units. The learning task is to saccade to the spotlighted area of the input vector. The two output units represent the directions left and right. Their activation values provide the distance to the left or right that the "eye" has to move in order to foveate the attended word "blob." The "shift fixation" module, when triggered, takes this output and uses it to modify the visual input matrix.

Modelling the Temporal Dynamics

Normally, a two-layer feed-forward network will generate an output from a given input in two time-steps. In order to derive processing time data from these networks, a technique first described by Cohen, Dunbar, and McClelland (1990) was used. During the *performance* phase of the modelling process, the standard weighted sum of input activations is replaced with the following time-averaging formula:

$$net_{j,t} = \tau \sum_i (a_{i,t} w_{ji}) + (1 - \tau) net_{j,t-1}$$

where $net_{j,t}$ is the net input to unit *j* at time *t,* $net_{j,t-1}$ input to this unit on the previous time cycle, $a_{i,t}$ is the activation of unit *i* at time *t*, w_{ji} is the strength of the connection from unit *i* to unit *j,* and τ is a time constant that determines what combination of the current and previous net inputs to the unit is to be used in the calculation of the current activation level. By using this formula, activation builds up slowly in the output units of a feedforward network and asymptotes to a stable value. The number of cycles to asymptote is used as an analogue of processing time. In Figure 1, the two modules labelled "asymptote detector" are used to check whether the output from the lexical and saccadic modules has asymptoted. When an asymptote is reached the

modules generate a signal that is used by the modules controlling fixation and attention shifts.

In the case of the lexical module, τ was chosen so that the number of cycles taken for the module to asymptote when fixating a typical word (both in frequency and length) was roughly equal to 125. This is the number of ms estimated to be needed to encode a typical word (Rayner & Pollatsek, 1989). In the case of the saccadic programming module, a value for τ was chosen so that the average number of cycles to asymptote was also around 150. The aim here was to equate cycle time with the number of milliseconds required to programme *and* execute a saccade. The saccadic programming time probably has a lower bound of 75 ms, which when combined with an efferent lag of 50-60 ms, gives a combined lower bound of 125 ms, with 150 ms assumed to be an average value. Using these criteria, the τ for the lexical module was set at 0.1 and at 0.15 for the saccadic programming module.

Training phase

In the training phase, the saccadic and lexical modules were trained using the backpropagation learning algorithm. Three stories excerpted from a school reader were used, consisting of 863 words in total, made up of 222 different words. The average word length of the text was 4.5 letters. Words occurred with varying frequency in the text, and this corpus-based frequency was used as a way of building in frequency structure that could be used in a later study of frequency effects.

The lexical module was trained to identify words randomly fixated at different locations. In training the saccadic programming module, the network was trained to make the range of saccade-types that one finds in normal adult reading. The precise proportions of progressions, regressions, and re-fixations were derived from empirical data (Rayner & Pollatsek, 1989; O'Regan, 1990).

Testing phase

For the test phase the *trained* saccadic and lexical components were assembled as shown in Figure 1, and the resulting behaviour compared with known qualitative and quantitative aspects of eye-movement control in reading.

Simulated reading proceeds as follows: Fixation-sized chunks of text comprising on average four words are pre-processed into a visual input matrix and then loaded into the visual input module. This module is used as a source of input for both the lexical encoding and saccadic programming modules. At some point the level of activation in one of the modules asymptotes to a stable value. In the case of the lexical module, the time taken to asymptote will vary according to the frequency of the word fixated and the fixation location within the word. When the lexical encoding module asymptotes this is detected by an asymptote detection module which sends an enabling signal to the lexical identification module which enables a shift in attention. On the other hand, if the saccadic module is the first to asymptote, *and* if the size of the proposed saccade is greater than some threshold, then a saccade is executed. Since the goal of the saccadic module is to fixate the currently attended word, a saccade at this stage will cause a refixation of the currently attended word in the manner proposed by O'Regan.

When a shift in fixation is triggered information about the size of the shift is read from the saccadic module and used to select the next chunk of text to be fixated. Attention is allocated to the word at the centre of the foveal projection If the centre falls on a space between words, the word to the right is chosen as the focus of attention (this assumption requires empirical verification). The text is pre-processed in the usual way by the visual input module and passed along the saccadic and lexical pathways. Note that the lexical module is not reset at this point, only the input layers of each module are changed. There will still be some residual activation in the hidden and output layers from the previous fixation which can help accelerate convergence in the current fixation, thus permitting the integration of information across fixations. It is a debatable point whether or not the saccadic module should also be reset. Are there, for example, the equivalent of preview effects in saccadic programming, whereby a saccade of equal length to the previous one is programmed more rapidly, thus shortening the current fixation duration? Again, this is an open empirical question. For the present, it is assumed that a reset does takes place.

Performance of the Model

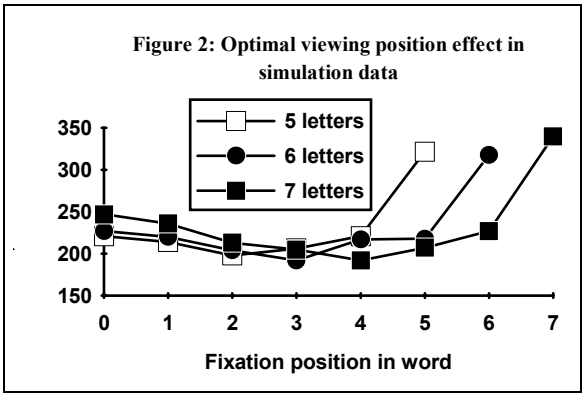
The mean fixation duration for one pass through the training text entailing over 800 fixations was around 200 ms and the mean saccadic length was 6 characters. Both figures are smaller than one would expect from adult readers. Fixations were shorter,

because the refixation tactic used by the model generated a relatively large number of brief fixations. The distribution of fixation durations tended to be bi-modal. This suggests that some of the timing assumptions, particularly with regard to saccadic module may have to be reviewed. The main reason for the shorter saccades was because fewer words were skipped than in normal reading and also because the mean word length of the text is below average at around four characters.

Shift invariance

The model demonstrated a surprising degree of shift-invariance in the recognition of words. Ten test passes were made through the corpus in which each word in the text was fixated at some random position in the range comprising the word, plus three characters prior and two after. Words were correctly identified on 95% of fixations, and of the 222 unique words 90% were correctly identified. It seems that the spatial and category "blurring" of the input representation, has the beneficial effect of making the input identifiable at different horizontal displacements.

Optimal viewing position effects



A typical OVP pattern was found in the simulation data. In Figure 2, the OVP for each word tends to be left of centre, and is more pronounced for longer words, as in real reading (O'Regan, 1990). Note that the zero location in Figure 2 represents the space prior to the word and that "cycles" is an analogue of fixation duration. Note also that this effect is not a training artefact, since each letter was equiprobable as a fixation location during training.

Contingent changes in display

Apart from normal reading, the model is also capable of simulating a range of eye-movement contingent display change experiments, such as the moving window studies of McConkie and Rayner (1975). One of their conditions involved replacing the letters of words in the parafovea with Xs. They found that this manipulation actually speeded up fixation durations when compared to other replacement options, such as the use of similarly shaped letters. They interpreted this effect as due to lack of interference from letters beyond the window boundary. In the simulation, when words in the periphery were replaced with either a random sequence of consonants or a sequence of Xs, the average fixation duration was longer for the consonant sequence than the X sequence. The X sequence was close to that of normal reading. The simulation behaviour suggests that McConkie and Rayner's explanation for this effect is only part of the story: In the simulation, both the lexical processing *and* saccadic programming components are speeded up, indicating that as well as providing less interference, the Xs also present a clearer target for the saccadic module.

Other features of the model

Due to space limitations only a sample of the model's capabilities can be discussed. Among other aspects of reading behaviour reproduced by the model are refixations, frequency effects (high frequency words are more rapidly encoded than low frequency words), peripheral preview effects, and word skipping. In the latter case two words are recognised in one fixation and a saccade is programmed to word n+2. The skipped word tends to be short and of high frequency within the corpus.

Conclusion

CASMEC is capable of accounting for a range of eye-movement control behaviour in reading. It represents a rigorous alternative to the more usual, informally specified, models in the area. CASMEC exploits the single currency provided by connectionism to represent the interaction between the visual, lexical, and motor domains

The effort of implementing the model has clarified some existing findings (e.g., the X effect in

the moving window experiments of McConkie and Rayner, 1975) and raised some new empirical questions: How, for example, is the intended target word selected on a new fixation if the eye lands between two words? Are there the equivalent of preview effects in saccadic programming?

The main shortcoming of the model is that it does not match the distributional properties of fixation durations found in readers. This is due to the timing assumptions of the model. Although these have been derived from empirical data, the simulation results suggest that the interpretation of these data may need to be re-examined.

While the focus of this paper has been on just one model of eye-movement control, the connectionist implementation is potentially a framework for the exploration of a range of such models. Many of the elements of the framework are uncontroversial; what is usually at issue is how the elements interact. The framework presented here should allow a number of different interaction protocols to be tested.

References

Chastain, G. 1982. Confusability and interference between members of parafoveal letter pairs. *Perception & Psychophysics* 32:576-580.

Cohen, J. D., Dunbar, K., & McClelland, J. M. 1990. On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review* 97:332-361.

Crick, F. 1984 The function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences* 81:4586-4590.

Inhoff, A. W., & Rayner, K. 1980. Parafoveal word perception: A case against semantic preprocessing. *Perception & Psychophysics* 27:457-464.

Levi, D. M., Klein, S. A., Aitsebaomo, A. P. 1985. Vernier acuity, crowding, and cortical magnification. *Vision Research* 25:963-977.

McConkie, G. W. 1979. On the role and control of eye movements in reading. In P. A. Kolers, M. Wrolstad, H. Bouma (Eds.), *Processing of visible language I*. New York: Plenum Press.

McConkie, G. W., & Rayner, K. 1975. The span of effective stimulus during a fixation in reading. *Perception and Psychophysics* 17:578-586.

Morris, R. K., Rayner, K., Pollatsek, A. 1990. Eye movement guidance in reading: The role of parafoveal letter and space information. *Journal of Experimental Psychology: Human Perception and Performance* 16:268-281.

Morrison, R. E. 1984. Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance* 10:667-682.

Mozer, M. C. 1991. *The perception of multiple object: A connectionist approach*. Cambridge, MA: MIT Press/ Bradford Books.

O'Regan, J. K. 1990. Eye movements in reading. In E. Kowler (Ed.), *Reviews of oculomotor research: Vol. 4. Eye movements and their role in visual and cognitive processes*. Amsterdam: Elsevier.

O'Regan, J. K, & Levy-Schoen, A. 1987. Eye movement strategy and tactics in word recognition. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*. Hillsdale, NJ: Erlbaum.

O'Regan, J. K, Levy-Schoen, A., Pynte, J., & Brugaillere, B. 1984. Convenient fixation location within isolated words of different length and structures. *Journal of Experimental Psychology: Human Perception & Performance* 10:250-257.

Rayner, K. , & Pollatsek, A. 1989. *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Rayner, K. 1978. Eye movements in reading and information processing. *Psychological Bulletin* 85:618-660.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986. Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing. Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press.