Visual Acuity, Lexical Structure, and Eye Movements in Word Recognition

J. Kevin O'Regan*

1 Why study visual word perception?

A large amount of work has been done in experimental psychology on the perception of printed words. The interest in this topic stems not only from its importance for understanding reading, but also from the fact that word recognition is a miniature universe which has essential links to visual perception as a whole, yet which is simpler and more tractable. Thus, word perception involves only two dimensions and not three: movement is irrelevant, as are (generally) colour and texture. Compared to the situation for 3D objects, it is much easier to define the visual 'atoms' which underlie word perception (letters, syllables, morphemes) and the rules that determine how these atoms can be combined to form legal words. Yet word perception retains an essential aspect which is present in object recognition, namely the interaction of sensory information with knowledge about the world (in the case of words, knowledge about words that exist). Understanding how sensory information and knowledge combine to give recognition is one of the central problems in both human and machine perception.

There is another important aspect of visual perception which is preserved in word recognition, namely the motor aspect: visual perception is an active process, involving exploratory eye movements. Word perception is a convenient place to investigate how this process is controlled and how it contributes to perception.

The present chapter will consider work we[1] have done on these aspects of word perception. Three issues will be covered: the role of lexical constraints in word recognition; the role of acuity; and the role of eye movements.

---

* Groupe Regard, Laboratoire de Psychologie Expérimentale, Université René Descartes University, Paris, France.

[1] In collaboration with A. Lévy-Schoen, A. Jacobs, C. Coeffé, F. Vitu, V. Holmes and J. Pynte.

2 Lexical constraints

Current models of word recognition

Consider a currently fashionable framework for word recognition, compatible with the work of, for example, Morton (1969), Bouwhuis and Bouma (1979), McClelland and Rumelhart (1981), or Paap, Newsome, McDonald and Schvaneveldt (1982). A word falls on the retina. At each letter position, probably after some kind of low level image filtering, features are recognized by feature 'units'. Those units that have been activated excite letter units, which in turn excite word units. The units involved are usually considered to be Perceptron-like threshold devices (cf. Rosenblatt, 1962; Selfridge and Neisser, 1960), which gave an output when a weighted sum of their inputs exceeds the threshold.

Though many people would agree with this general plan, the precise nature of each stage is very uncertain. What is the low-level processing exactly? What degree of invariance to translation, size, rotation, colour, brightness does it provide? What are the features (and/or letters): are they hooked to precise retinal locations or not? Can global contours be features? Can the absence of an element be a feature? Is there a hierarchy in the feature extraction process? Are the same battery of features (and/or letters) used for all typographies, or are they turned in and out as a function of the typography? Are the features (and/or letters) size/rotation invariant? Is there really a letter stage, or can features feed directly into word units?

There are also some questions concerning the whole organization of this model of word recognition.

1. It does not seem very parsimonious to suppose that the same battery of feature and letter detectors must be repeated several times over the central part of the retina. Could not there be some kind of internal scanning mechanism that allowed the same detectors to be used in a serial fashion over each retinal location?

2. Is it correct to suppose that letters are coded individually and then assembled into words, or are there detectors for groups of letters, considered as units (for example the following have been proposed: bigrams, trigrams, vocalic centre groups, syllables, morphemes)? Are there units that code phonemic structure?

3. Exactly how are the units connected: are there connections within a given level, or only between levels? Can connections cross over a level? Which connections are excitatory, which inhibitory? How are their strengths determined?

These questions determine how the feature/letter/word units interact or compete to reach a decision when there are ambiguities or uncertainties. What information is sent up to each successive level: are there really thresholds, and how are they determined? What level of this hierarchy is subjectively 'seen' by the reader: only the top levels, or all levels at once?

Despite such vagaries, various authors have attempted to approach some of the above questions using experimental methods. A surprising degree of success has been achieved by authors who have worked on the higher levels, ignoring the feature extraction process and what precedes it (for a review of such studies, see Carr, 1986). One classic paradigm that has been used is the 'word superiority effect', perfected by Reicher (1969) and extensively studied since, in which it is shown that in conditions of brief, masked presentation, a word can be more easily seen than each of its constituent letters. This phenomenon is explained naturally by the present model, since owing to the fact that only some letter combinations form real words, even under conditions of poor visibility, there may still be enough information to select the correct word. A second much-studied phenomenon is the word frequency effect, in which words which occur more frequently in the language are found to be recognized more easily than rarer words. Though there is debate about whether this effect actually has its locus within the present framework, or whether it should be considered a phenomenon taking place at some later decision level, many workers account for it by building thresholds of activation into the word units in such a way that frequent words are more easily activated than rare words. Detailed study of the word superiority and word frequency effects has enabled researchers to make precise suggestions about the 'writing' of the higher levels of the word-processing 'network' (McClelland and Rumelhart, 1981). In the following I will discuss a new effect which also looks promising.

Orthographic neighbourhood effects

In the word-recognition framework presented above, there is an interesting consequence of the idea that the units for frequent words should have lower thresholds. Consider the French word 'SOIN'. When it is presented to a reader, it is not the only word unit to receive activation. Some activation is received by other words like 'SOIR', 'LOIN', 'FOIN', 'SOIT', and 'SOIS', which share in the same positions in the word three out of four of the letters of 'SOIN'. Activation is also received by words sharing only two or one letter with 'SOIN', but presumably this is much weaker. Following McClelland and Rumelhart (1981), we call the words sharing three out of four letters with 'SOIN': 'orthographic neighbours' of 'SOIN' (this definition of orthographic neighbour will serve as a simple first approximation, but it should ultimately be improved by taking account of, for example, visual similarity of letters, letter frequency of the letters that are changed, as well as possibly changes in letter position, etc.). Now suppose that one or more of these words has a higher frequency of occurrence in French than the word 'SOIN' itself. This is the case for 'SOIR', 'LOIN', 'SOIT', for example. Then, depending on the 'wiring' postulated in the network, these words should compete with 'SOIN' and make it hard or impossible to see!

We tested this hypothesis in two experiments (Grainger, O'Regan, Jacobs and Sequi, *submitted*). We searched the French dictionary for words with certain kinds of orthographic neighbours. One kind of word had no orthographic neighbours (Henderson, 1982, has called these 'hermit' words). An example in French is 'BREF'. A second kind of word had several neighbours, but all of lower frequency (example: 'JUPE'; lower frequency neighbours: 'JUGE', 'JUTE', 'DUPE'). A third type had a single neighbour of higher frequency (example: 'CHAR'; higher frequency neighbour: 'CHAT'), and a fourth type had several neighbours of higher frequency (example: 'SOIN'; higher frequency neighbours: 'SOIR', 'LOIN', 'SOIT'). We had 10 words in each category, all words being of length four letters. We were interested in the time it took to recognize the words as a function of the different neighbourhood categories. In order to prevent other factors from producing spurious time differences, we carefully matched the test words for frequency (mean 30, range 5-80 per million), and experiential familiarity (mean 4.2, range 3.2-5.8 on a survey using a subjective 7-point scale). We also attempted to match them for positional bigram frequency, but this was not possible for the last category (for word types 'BREF', 'JUPE', 'CHAR', 'SOIN', bigram frequencies were 7.0, 7.0, 6.3 and 16.3 respectively over the 800 words of the four-letter French corpus). We did two experiments, one in which the words were mixed with nonwords, and the task was to make a word/nonword judgment as each word appeared on the screen. In the other experiment each word appeared on the screen accompanied by a second word which either had or did not have a semantic relation to the test word. The subject read the first word, and moved his eye to the second word to make the semantic decision. Using eye movement recording, we measured the total time the eye spent on the test word before moving on to the second word. This 'gaze duration' on the test word was taken as a measure of the recognition time for the test word.

Figure 1 shows the results for the different types of neighbourhood. Both the word/non-word decision times and the gaze durations show a similar pattern. Words like 'BREF' with no orthographic neighbours are fastest, and comparable to words like 'JUPE' with several lower-frequency neighbours. But words with one (e.g. 'CHAR') or more (e.g. 'SOIN') higher-frequency neighbour are more difficult to recognize (shaded bars). These results were statistically significant both when analysed by subjects as by items. It is interesting, however, that the 'SOIN'-type words behave differently in the gaze duration data. Further work will show if this has something to do with the fact that this category had a higher bigram frequency than the other categories, and that this plays a more important role in gaze duration measurements than in word/non-word decision times.

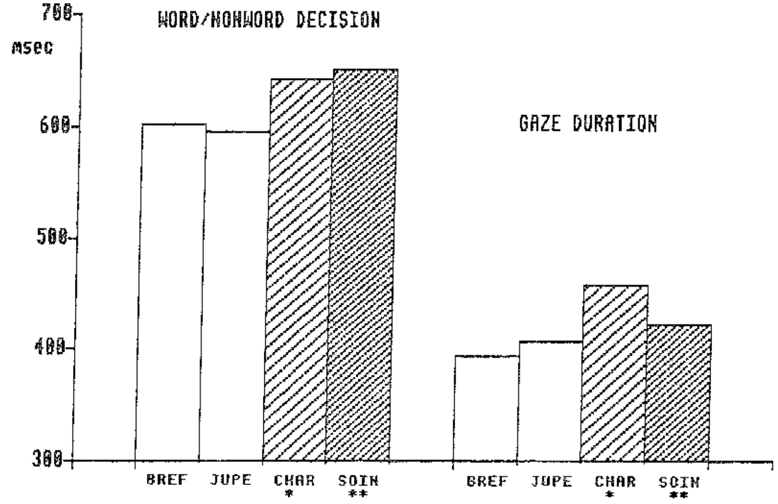

Figure 1: Results of the orthographic neighbourhood experiment (Grainger, O'Regan, Jacobs and Segui, submitted). Each bar shows the manual reaction time to word/non-word decision (left half or graph) or the eye's gaze duration (right half) for words like 'BREF' (no neighbours), 'JUPE' (several lower frequency neighbours), 'CHAR' (a single higher frequency neighbour, indicated by the asterisk) and 'SOIN' (several higher frequency neighbours, indicated by two asterisks).

Given the many questions concerning the details of the word recognition framework we have set out above, and given the simplistic definition of orthographic neighbourhood we used, it is gratifying to find our hypothesis supported. Note that several other authors (Savin, 1963; Coltheart, Davelaar, Jonasson and Besner, 1977; Chambers, 1979) have done similar studies, but with inconclusive results. We believe the reason for this is that they did not properly control for one of the three essential variables: experiential familiarity, bigram frequency, and frequency of neighbours. The point we wish to make here is that we may now be beginning to manipulate the relevant variables sufficiently well to start using orthographic neighbourhood effects as a sensitive tool with which to probe the higher levels of the word recognition process.

3 Acuity and word recognition

What can be seen at an eye fixation?

I will now turn to another aspect of word recognition, associated with the earliest, sensory, stages of processing. The word recognition framework presented above was built up to account for recognition of short words. This explains why many authors did not bother to incorporate the eye's power of resolution into their models. It is clear that in the recognition of long words this must however be done, since letters not directly fixated are seen with less resolution. In addition, word perception is not a purely receptive process, since in many cases the eye moves around inside a word before recognition is complete: there is an active exploration of the word. Much of the work we have done in the past years in our group attempts to understand the implications of these aspects of vision. In the present section I will attempt to calculate on theoretical grounds the number of letters which ought to be visible from the fixation point. Then, in the section that follows this one, I will look at the consequences of this for word recognition and eye movements.

Figure 2 is a tangential cross-section of the retina taken from Osterberg (1935). The centre of the retina is on the left. The circles are the cones and the black dots that begin to appear on the right of the figure are the rods. Imagine the image of a word projected onto this retina, as at the instant of fixation during reading. The figure is plotted to scale, assuming letters of angular size approximately 20 min of arc (equivalent to reading the journal *Vision Research* at 30 cm), taking account of 30% shrinking of Osterberg's preparation, and using the optical parameters for the eye

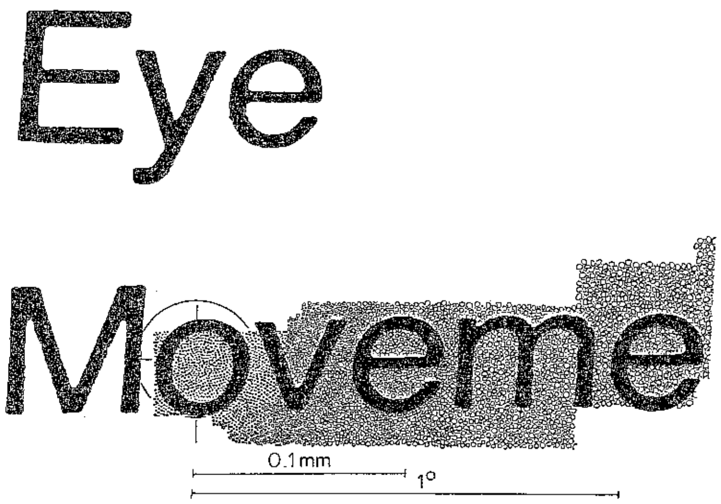

Figure 2: Tangential cross-section of the retina, including the fovea, whose centre is indicated by the cross-hairs. Small circles are cones, and dots, appearing only beyond about 1 degree, are rods. The words 'Eye movement' are superimposed as they would appear on the retina when reading them in *Vision Research* at a distance of 30 cm (except they would appear upside down and in mirror image form).

suggested by Le Grand (1964) – idealized eye with refractive index within the eye of 1.34 and a magnification at the pupil of 0.93. The letter 'o' being directly fixated is being sampled by something like a grid of 30 x 30 cones. Two letters further into periphery, the letter 'e' is being sampled by a 19 x 19 grid. Two more letters out, at an eccentricity of just over 1 degree, the next 'e' is only being sampled by a 15 x 15 grid of cones. The drop-off of resolution is very strong, even within the 1-2 degree zone usually called the fovea: at 1 degree from fixation, cones are spaced almost twice as wide as at the eye's centre. Since ganglion cell convergence gets progressively larger in parafovea, the 'effective' spacing as measured in the cortex will be even larger than this and resolution correspondingly worse. Levi, Klein and Aitsebaomo (1985) have collated recent anatomical evidence confirming this. For the studies reviewed, whereas cone spacing at 1 degree is about 1.2 times the central value, cortical visual field sizes are about 2 times the central value (this is shown by the solid lines in figure 3). It should be noted that Levi *et al.* claim that older estimates of cortical magnification (e.g. Rovamo, Virsu and Nasanen, 1978) are too small.

A striking fact about the curves in figure 3 is that they are straight lines. This expresses the fact that peripheral vision, up to at least 10 degrees, can be considered to be functionally a linearly expanded version of central vision, with an expansion factor of about $(1 + m\phi)$, where ϕ is the eccentricity in degrees, and m is about 1.7. Schwartz (1980) has suggested that this provides a cortical representation of the visual field which is almost invariant under rotation, size and projection changes. We will see an example of a kind of 'pseudo' size invariance later in the discussion of visual span.

The anatomical data for cortical sampling of the parafoveal visual field agrees well with psychophysical acuity data. Different types of acuity judgments are known to give widely varying values, with hyperacuity for vernier displacements being measured in seconds of arc at the centre of the visual field, but resolution for optotypes like the Landolt C only minutes of arc. Nevertheless, if resolution in peripheral vision is expressed as a ratio of its central value, the curves for vernier, Landolt, stereo acuity, and phase discrimination follow similar curves, and these are in fact straight lines closely resembling those for cortical visual field sizes. This is shown as the upper set of dotted lines in figure 3, for various studies collated by Levi *et al.* and myself: at 1 degree, resolution is 2.5 times that at the centre.

Parenthetically it is worth noting that when contrast sensitivities to grating are measured in periphery, these appear not to obey the same scaling factor. It appears that grating acuity scales in a way more similar to cone

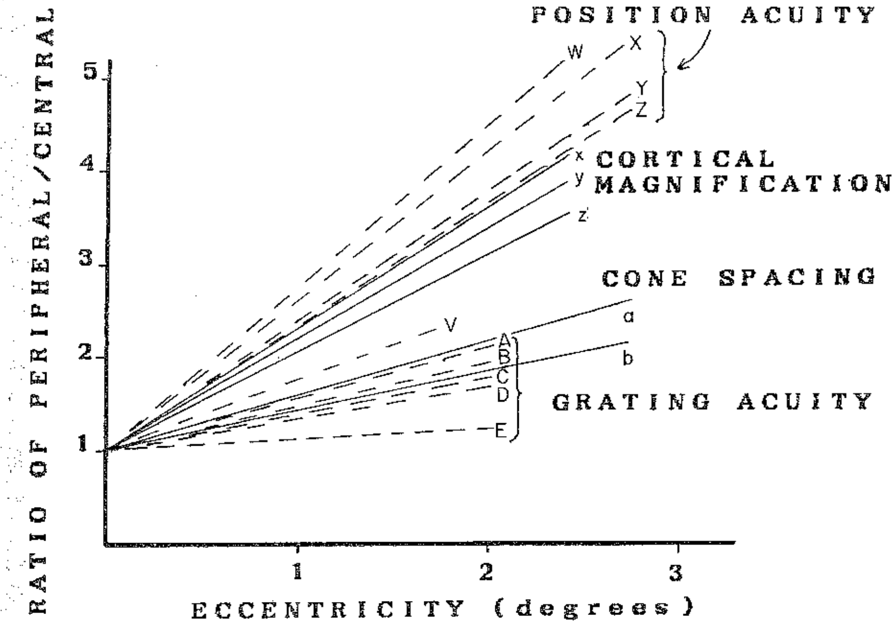

Figure 3: Ratio of peripheral and central values of various measures, as a function of eccentricity. Data compiled partially from tables given in Levi *et al.*, (1985); *C*: Westheimer, (1982); *E*: Wilson and Bergen, (1979), based on Hubel and Wiesel. Measures of position acuity: *V*: Jacobs, (1979), Landolt C without masking bars; *W*: Jacobs, (1979), with masking bars; *X* and *Z*: Levi *et al.*, (1985, vernier); *Y*: Fendick and Westheimer, (1983, stereoacuity); also *Y*: Klein and Tylor, (1981, phase discrimination). Measures of cone spacing: *a*: Osterberg, (1935, range 0-2.5 degrees); *b*: Rolls and Cowey, (1970). Measures of cortical magnification: *x*: Dow, Snyder, Vautin and Bauer, (1981); *y*: Van Essen, Newsome and Maunsell, (1984); *z*: Tootell, Silverman, Switkes and De Valois, (1982).

density rather than cortical receptive fields (cf. Levi *et al.*, 1985), that is with $m = 0.8$ (lower set of dotted lines in figure 3).

In the above paragraphs I have illustrated the strong drop-off of acuity even within the fovea. This drop-off can be expected to degrade letter recognition even a few letters from the fixation point. But there is yet another factor which further degrades letter recognition, namely lateral interaction. Optometrists have been aware of this phenomenon for a long time (e.g. Bourdon, 1902; Flom, Weymouth and Kahnemann, 1963): they call it 'crowding'. Psychologists at least since Erdmann and Dodge (1898), Korte (1923), and Woodworth (1938) have also demonstrated and studied it. More recently some influential studies have come from IPO: Bouma (1970), Andriessen and Bouma (1976). In all cases it is shown that acuity in parafovea and periphery is worse when the test stimulus is flanked by nearby contours.

What is interesting about this phenomenon is that it is generally considered not to be explicable in terms of classical psychophysics, since it extends over retinal distances which are too large. Thus Bouma (1970) estimated that interactions affecting a stimulus placed at eccentricity ϕ only become negligible if there are no contours within a distance 0.4ϕ, which corresponds to very large retinal distances. However, these may actually be rather small cortical distances.

Levi *et al.* (1985) suggest that classical psychophysics has assumed that peripheral vision scales to central vision following a scaling factor which is based on underestimations, and that the correct scaling factor is much larger. In their study of vernier acuity they confirm that masking phenomena scale correctly according to recent estimates with m=1.7. Bouma's value of 0.4ϕ for the extent of the region of interaction, would, under this scaling factor, correspond to a region of interaction at the centre of the eye having a radius of 14 minutes of arc. This is the outermost edge of the negative region of the psychophysical line spread function estimated by a number of authors (e.g., N channel in Wilson and Bergen, 1979). It is, however, about three times larger than the edge of the point spread funtion measured at IPO by Blommaert and Roufs (1981); but such differences may be due to differences in background luminance.

Thus it does not seem unreasonable to suppose that even lateral interactions of the type observed by Bouma follow the normal cortical scaling factor of $(1 + 1.7\phi)$. Though further work is needed to confirm this point, it will be useful to take it as an assumption in the following work, because it allows us to express in a single formula the dependence on eccentricity of both resolution *and* lateral masking. We define an 'effective resolution' r_0 at the centre of the field, corresponding to the resolution in the presence of masking contours. Then effective resolution at eccentricity ϕ can be taken to be $r_0.(1+m\phi)$. This assumption allows some simple, yet powerful calculations to be made concerning the number of letters visible from the fixation point.

Calculation of theoretical visual span

Assume that to recognize a letter, it is necessary to recognize a certain number of features within that letter. These features will be aspects of the letter which differentiate it from other letters in the alphabet being used. The particular features used will depend on the task: if a very restricted alphabet is used (say the only letters used are X's and O's), then fewer, coarser features will be necessary to make the distinction than if the whole alphabet of 26 letters is used. Let us therefore consider a particular, fixed task, say word recognition (rather than reading the numbers on bank checks, where more precise recognition may be necessary), and a particular type font. The features required will have a characteristic size, which we call 'g' the 'grain size'. We expect that a reasonable value for this size might be something like 1/3 to 1/10th of the character width, but this is an empirical question. Clearly the 'grain' of a typography is a measure of its visibility. The larger the grain, the easier it is to see.

Imagine that the eye, situated at distance d, is looking at a letter in a string of letters; the spacing from letter to letter is w. How far from the point of regard will a letter within the string be recognizable? Say the N-th letter is just visible. It is at eccentricity Nw/d radians (taking $d \gg w$). Assuming acuity and lateral interactions in central vision give an effective resolution of r_0 radians, resolution at this eccentricity Nw/d will be scaled to the value $r_0.(1+mNw/d)$. The letter will be just visible if this resolution is sufficient to resolve the distinctive features of the letter, which have angular size g/d radians:

$$\frac{g}{d} = r_0 \cdot (1 + \frac{mwN}{d}) \tag{1}$$

from which we obtain

$$N = \frac{g}{r_0 mw} - \frac{d}{mw} \tag{2}$$

The dependence of N on d is shown as a graph in figure 4. Note that the equation is only an approximation, valid for $d \gg w$. It makes no provision

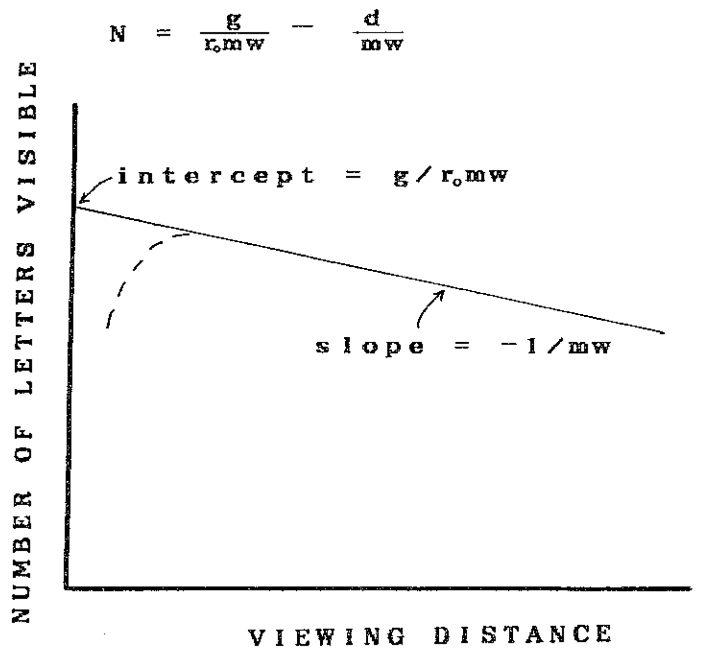

Figure 4: Predictions of the 'grain' calculations for the number N of characters visible on each side of the fixation point, as a function of viewing distance. g: grain; r_0: resolution in radians at eye's centre; m: scaling factor; w: distance between centres of successive letters; d: viewing distance.

for possible differences between left visual field and right visual field. There is also an error in assuming that lateral interactions scale according to $(1 + m\phi)$, since this would be true only if the distance between test and masking contours were scaled also. In fact letters in peripheral vision remain the same distance apart, so the 'effective' distance of test to contour is smaller. To take account of this problem, larger values of m would have to be taken as an increasing function of eccentricity, rather than a constant, as assumed above. A further correction is necessary owing to the fact that letters on a flat surface will subtend a smaller angle when they become eccentric. Both these corrections apply for large angles (small distances), and modify the equation so that for small viewing distances the critical value of N drops. This is shown as a dotted line in the graph.

Several interesting things about this graph are worth noting. First of all, except for small distances, the graph is a straight line with y-intercept g/r_0mw and gradient $-1/mw$.

If w and m are fairly large, the slope will be very flat, and dependence on viewing distance will be small. This means that the number of letters which are visible on each side of the fixation point will not depend strongly on the viewing distance. This may be part of the reason that people can place their heads quite freely with respect to the material being read. Also, since on the retina an effect equivalent to changing viewing distance can be obtained by changing character size, the same point can be made about character size: the number of letters visible will be fairly independent of character size.

Note that while the *number* of letters visible is insensitive to distance and size changes, the *angular eccentricity* over which there is clear vision is strongly dependent on distance, since it is equal to $Nw/d = constant * w/d$. This explains why it is important to measure visual span in number of characters, and not in degrees. The same is true of saccades in reading (cf. O'Regan, Lévy-Schoen and Jacobs, 1983): they should be measured in number of characters, otherwise there will be a strong dependence on viewing distance or letter size[2].

It will be useful to give an intuitive explanation of where this size invariance comes from. Consider a string of letters, being viewed from a certain distance. If I double this viewing distance, the letters will subtend an angle

---

[2] It appears that even when measured in letter-spaces, saccades are actually *less* dependent on viewing distance than predicted from the idea that the eye moves to the edge of the region of visibility. We think this is for reasons connected with cognitive and oculomotor factors (O'Regan and Lévy-Schoen, 1987).

half as large, and so they will be twice as hard to see. But because they are half as large, they are twice as close to the centre of the retina, where because of the linear relation of resolution to eccentricity, resolution is twice as good! The net result is that I can see them all equally well again! The key to this reasoning lies in the approximately linear scaling of central to peripheral vision. In fact Schwartz (1980), who has extended this kind of argument to rotation and projection changes, argues that the correct scaling should be logarithmic rather than linear. But this only makes a difference for large eccentricities. Also, the fact that at the eye's centre resolution is finite and not 0 minutes of arc (i.e. acuity is not infinite at the eye's centre) means that the intuitive argument is not quite right, and explains the slight fall-off in number of letters visible with increasing viewing distance.

Empirical data for visual span

It is difficult to empirically measure the eccentricity at which a letter can just be seen in a string. You cannot just present people with a string of letters and a fixation point, and ask them to read off the letters without moving their eyes: people have an irresistible tendency to move their eyes, and find it very difficult to suppress this[3].

We tried flashing up a string of letters for a time too short for eye movements, and asking subjects to report a letter at a given eccentricity. Unfortunately when we tried to use a nearby marker to indicate which letter was to be reported, presumably because of lateral masking or poor peripheral position coding, the subjects were totally unable to tell which letter the marker was pointing to. We resorted to using only a group of three characters in peripheral vision, with the central one to be reported. In order to avoid confusions the outer characters were numbers rather than letters: this provided contour masking similar to what other letters would produce, and yet avoid confusions in reporting the central test letter. In using this technique we were building on the method Bouma (1970) had already used to study contour interactions, except instead of having x's as flankers and studying the effect of spacing these closer or nearer to the test letter, we used numbers as flanker and investigated the effect of changing the viewing distance. We used only upper case characters.

---

[3] Nevertheless Townsend, Taylor and Brown (1971) managed to train two subjects to do this, but did not investigate the effect of size of distance on the number of letters that were readable. For the record, they found for characters subtending 20 min of arc, that people could recognize and locate 5 letters with better than 50% accuracy. Erdmann and Dodge (1898) had also found a similar figure.

We used a computer-controlled adaptive psychophysical method to obtain the percent correct letter reports as a function of eccentricity. A separate curve was plotted for each viewing distance. From each curve we extracted the eccentricity at which accuracy of report was 50% (which we arbitrarily defined as the threshold of visibility). We plotted these eccentricities as a function of viewing distance in figure 5. Of course eccentricity is indicated in number of characters (N), not in degrees, since this allows comparison with the formula obtained above. The results are means over four subjects.

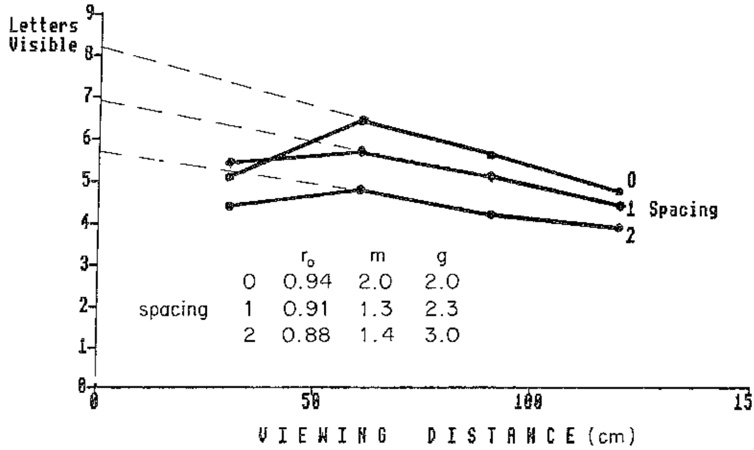

Figure 5: Results of experiment to measure number of characters visible on each side of the fixation point as a function of viewing distance. Results for normal (0), double (1) and triple (2) spacing. The table shows values of m and g calculated from the graphs' slopes and intercepts using independent measure of central resolution r_0.

As a test of the calculations above, in the experiment we used three different spacings between the characters:

(0) NORMAL SPACING
(1) O N E B L A N K S P A C E
(2) T W O   B L A N K   S P A C E S

In doing this, we assumed the grain of the characters remained constant, but that the effective spacing w from character to character increased from

1/3 to 2/3 to 1 cm.

Let us first look at the general aspects of these curves. It is clear that they follow the expected trend. The variation in visual span in going from a viewing distance of 30 to 120 cm (equivalent to using characters four times as small) is never more than 1-2 letters. Note that if I had plotted angular size of the characters rather than distance, I would have obtained a curve showing that visual span is largest for medium-sized characters, with very small and very large characters giving smaller visual spans. Bouma, Legein, Mélotte and Zabel (1982) have also found this.

If my theory is correct, the slopes of the straight portions of the curves should be equal to $-1/mw$, so since we know w, we can estimate m. The values obtained are shown in the figure. The values are exactly in the range we expect from the independent anatomical and psychophysical estimates of cortical magnification discussed above.

The intersects of the straight portions of the curves on the y-axis should be at values g/r_0mw. If we knew r_0, we could deduce the value of the grain g. We therefore did a second experiment on the same subjects in which we measured central acuity r_0, in the presence of laterally masking letters at the three different spacings 0, 1 and 2. The values we obtained are indicated in the figure. As expected from lateral interaction, we observed that resolution improved slightly as the adjacent letters moved farther away.

Reading off the intersects from the curves in figure 5, and putting these equal to g/r_0mw for each spacing condition, we get, after calculation and conversion into pixels, three estimates of the grain: 2, 2.3 and 3 pixels for spacing 0, 1 and 2 (also indicated in the figure). Considering that the characters were made in an 8×8 matrix, these values seem eminently reasonable. It is interesting that the subjects appeared to be using coarser distinctive features in the wider spaced condition. Perhaps this is because when lateral masking is weaker, global contours interact less and are more useful.

The pleasing aspect of the analyses I have presented is that they make the connection between anatomical, psychophysical and psychological data by the use of the simple notion of 'grain'. Since this notion allows predictions to be made about the visibility of characters as a function of their size, spacing and of the subject's viewing distance, it may have use in ergonomics. The present results and approach (as well as some other work we have done on the effect on character recognition of defocussing the retinal image, as occurs in presbyopia), would certainly benefit from being compared to the work that was done at IPO on character size and poor acuity (Bouma *et al.*, 1982), and by Van Nes and Jacobs (1981) on contrast in character recognition.

4 Eye movements in word recognition

The optimal viewing location

I have discussed two factors that strongly limit perception even quite close to the fixation point. Because of these, it may be that in word recognition it is advantageous for the eye to fixate a particular 'convenient' or 'optimal' place in words: for example, in the case of long words, a fixation somewhere near the middle of a word might be a good vantage point, from which the letters at the word's beginning or end might still be sufficiently visible for the word to be easily identified. On the other hand, fixating near the beginning or end of a word might render the letters at the other end of the word hard to see, and hinder recognition. However, note that a further factor, namely lexical structure, may modulate these considerations. For example, in many cases the lexical structure of a word is such that the beginning of the word carries more information than the end. This might shift the 'optimal' position more towards the word's beginning. Thus, the truly optimal position might depend not only on visibility considerations, but also on the word's lexical structure.

These ideas are confirmed by our experiments. The paradigm we use is as follows. The subject is sitting in front of a computer screen wearing a special photoelectric device mounted on spectacle frames that allows the horizontal eye position to be measured. He looks at a fixation mark at the centre of the screen, and when the computer detects accurate fixation, it displays a test word at the eye's fixation point, and one or more additional words further to the right. The subject's task is to read the short phrase formed by the whole group of words, and make some judgment about it. Depending on the experiment, we have asked subjects to decide whether the short phrase makes sense or not, or whether two words displayed are the same. To prevent interference between processing of the test word and the remaining words, while the subject's eye is on the test word, the other words in the phrase are masked out by being spattered with random dots. As soon as the subject's eye leaves the test word the fog of dots lifts from the other words and definitively descends upon the test word (thereby preventing further processing of it).

The variable manipulated in these experiments is the position relative to the eye's fixation point at which the test word appears. Thus, the test word may appear shifted to the right or to the left relative to the fixation mark in such a way that on appearance, the eye is fixating in different zones of the word. The eye is free to move around in the test word before going on to the next word, and it may make one or more fixations before doing so.

Figure 6 (*top*) shows the observed gaze durations on the test word: that is the total time spent by the eye on the word before going on to the next word. The data (replotted from O'Regan, Lévy-Schoen, Pynte and Brugaillère, 1984) show that the position where the eye starts in the word is very critical in determining the total time spent on it. There is a 'convenient viewing position', somewhat left of centre, where the eye should start in the word for recognition to be most rapid. Notice that the penalty for not starting at the convenient viewing position is rather large: about 20 ms per letter of deviation. The effects are highly significant. The lower graph shows the mean number of fixations made on the word as a function of the position the eye started fixating when the word appeared. The curves show that when the eye starts fixating at the optimum position, fewer fixations are made than when it starts fixating in the word at a different position. We have replicated these results several times now (cf. O'Regan and Lévy-Schoen, 1987), and similar effects are beginning to be found by other authors both for isolated words (Underwood, Hyona and Niemi, 1987) and also in normal reading (Blanchard and McConkie, cf. O'Regan and Lévy-Schoen, 1987).

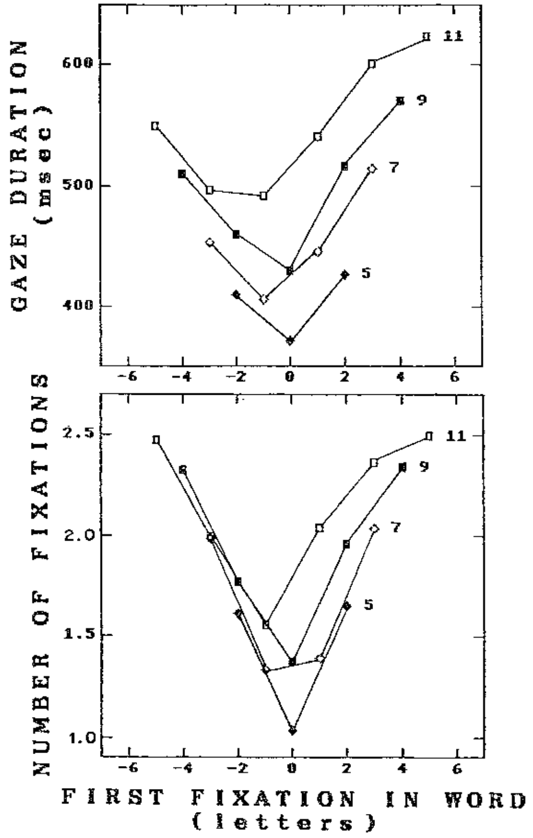

Figure 6: *Top*: gaze duration, and *Bottom*: number of fixations made on a word, as a function of the position the eye starts fixating in the word. Data for different word lengths (5, 7, 9 and 11 letters) are shown (replotted from O'Regan *et al.*, 1984).

I call the phenomenon I have just described the 'convenient viewing position effect' or the 'optimum viewing position effect'. It consists in the strong dependence of total time spent by the eye on a word as a function of where the eye starts fixating in the word. I was surprised by the strength of the effect, and by its stability. It can also be obtained without measuring eye movements by merely asking the subject to *name* words, but using naming is not such a pure measure as gaze duration, since it involves the complication of setting the subject's articulatory apparatus into motion and brings with it a number of objections. The strength of the effect, and the fact that it presumably depends on a combination of visual and lexical factors, suggest that it might be a good way of probing the process of visual word recognition and lexical access.

The optimum viewing position and lexical factors

If, as I have suggested, the optimum viewing position depends on a combination of visual and lexical factors, then it ought to be possible to modulate it by modulating each of these factors. The work we have done recently has been concerned with modulating lexical factors, in particular word length, frequency, and morphemic structure. Studying how the optimum viewing position phenomenon depends on these factors is a new way of gaining insight into word recognition processes. Here I will discuss work we have done on 'informativeness': a factor that has strong resemblance to what Marslen-Wilson calls 'cohorts' in auditory word recognition (e.g. Marslen-Wilson and Welsh, 1978).

We selected in the French dictionary two groups of 10 words: the 'informative beginning' group, and the 'informative end' group. The groups were matched for length (10-12 letters) and for frequency (mean: 3.5 per million). Each word in the 'informative beginning' group had the property that given its length plus or minus a few letters, it was uniquely determined in the dictionary by its first six letters (examples: *perquisition, attroupement, arrestation, auxiliaire, hirondelle*). An ideal lexical access machine therefore should be able to recognize the word by looking only at these six letters. Words in the 'end' group had the property that they were uniquely determined by their last six letters (examples: *circonspecte, interrogatif, transversal, approfondi, architecte*).

The task was a semantic judgment: the test word appeared as the first word of a short phrase that could make sense (e.g., 'perquisition brutale') or not (e.g., 'gymnastique de la solidité'). Oculomotor behaviour was measured as a function of the position where the eye started fixating in the word, using the same paradigm as above.

In the hypothetical case that lexical access can make use of word end constraints as effectively as word beginning constraints, the optimum viewing position for the 'beginning' words should be near the beginning, and for the 'end' words should be near the end of the word.

Figure 7 (*top*) shows the gaze durations as a function of position initially fixated in the word (see also O'Regan and Lévy-Schoen, 1987). The typical U-shaped optimum viewing position curves are found again. For the 'beginning' words, the optimum viewing position is near the fourth letter. The 'end' words do not however have their optimum viewing position in the expected place at their ends. Instead, there seems to be a much less well-defined optimum, spread out over the central part of the word. Nevertheless there is a clear, statistically significant difference in the behaviour of the two types of words, showing that lexical factors intervene in determining the optimum viewing position. The lower curves, showing number of fixations made, follow exactly the same trends.

It is interesting to ask why the 'end' curves do not have a clear minimum over the ends of the test words. A possibility is as follows: suppose that the mental dictionary must be accessed via the first letters of words. Consider what happens if the eye starts fixating the letters at the beginning of a word. If the first letters are informative (that is, they allow the end of the word to be deduced), then the eye has no need to go to the end of the word, and recognition is faster than if the beginning is not informative. Now consider what happens when the eye starts fixating at the end of the word. In all cases the eye must move to the beginning of the word to be recognized: it is irrelevant whether or not the end of the word is informative.

We further tested this idea by modulating the degree of informativeness

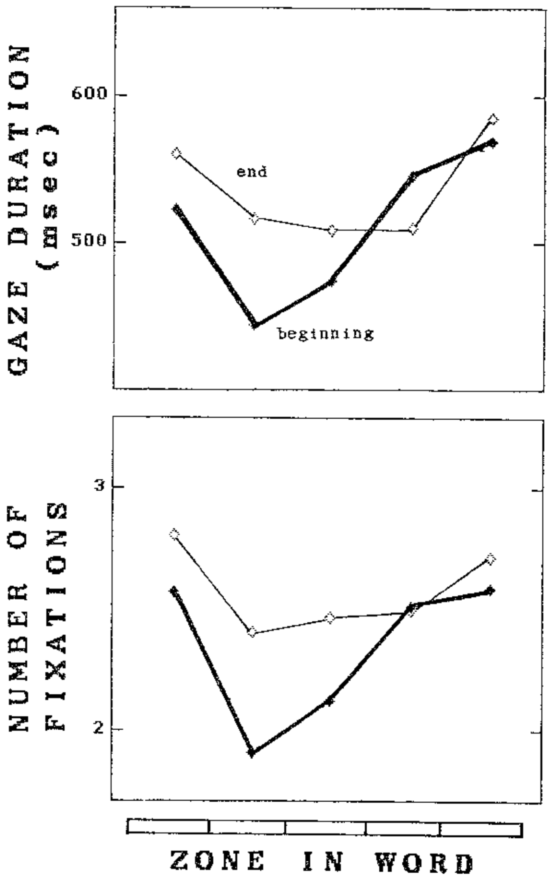

Figure 7: *Top*: gaze durations, and *Bottom*: number of fixations made on a word, as a function of the zone in the word the eye starts fixating. Data for words with informative beginnings and with informative ends are shown (replotted from O'Regan and Lévy-Schoen, 1987).

of the beginnings and ends of the words (O'Regan and Holmes, *in preparation*). We chose the same 'beginning' words as before. In addition, for each such word, e.g., 'attroupement', we found a word which could be called 'uninformative beginning'. This word was visually quite similar, had the same length and the same frequency in the language, but it had the property that the beginning was not informative this time: now there were one or several other words beginning in the same way. These other words were also more frequent than the test word itself. An example of such a test word was 'accoutrement', whose beginning is shared with two more frequent words: 'accoutumance' and 'accoutumer').

The same thing was done for the 'end' words: we retained the previous 'end' words and added for each word another word of the type 'uninformative end'. These had the same form, but were such that they shared their ends with one or several other words with higher frequency: e.g., for the 'end' word 'interviewer', we found the 'uninformative end' word 'intervertir', which shares its end with the two higher frequency words 'pervertir' and 'divertir'.

We mixed all the words together and repeated the experiment. The results are shown in figure 8. The thick lines correspond to 'beginning' words, and the thin lines to 'end' words. Globally the two sets of curves show the same trends as before, both for gaze durations (*top*) and for number of fixations (*bottom*): for 'beginning' words there is a minimum near the word beginning, but for 'end' words the minimum is spread more evenly over the middle part of the word. Now consider the difference between the curves for 'informative' words (*continuous lines*) and for 'uninformative words' (*dashed lines*). Both for gaze durations and for number of fixations there is a difference between 'uninformative' and 'informative', but only for the first two letters of the 'beginning' words. There is no statistical difference between the continuous and dashed thin lines, which are for 'end' type words. It seems therefore that the degree of informativeness plays a role, facilitating word recognition. However it does so only at the beginning of a word. This seems compatible with the interpretation we gave of the previous experiment: less use is made of constraints arising from the ends of words. Word recognition cannot occur via word ends, even if these are highly constraining.

The experiments described here show the use of initial viewing position as a variable to probe word processing. In addition to a number of studies we have carried out to better understand word length, frequency and morphological structure, Bouwhuis (1984) at IPO has recently also considered

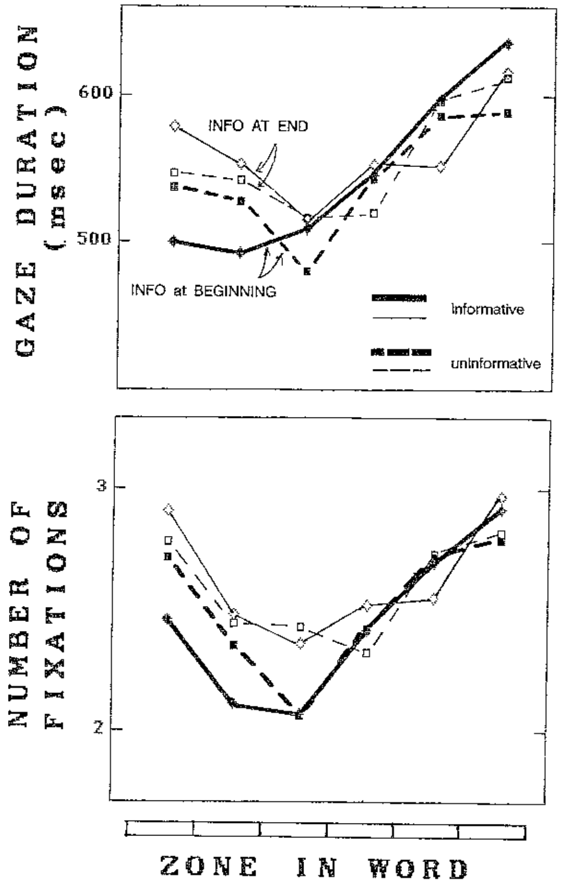

Figure 8: *Top*: gaze durations, and *Bottom*: number of fixations made on a word, as a function of zone in word the eye started fixating. Thick lines are for 'beginning' type words (*continuous*: informative, *dashed*: uninformative). Thin lines are for 'end' type words (*continuous*: informative, *dashed*: uninformative).

fixation position as a tool in studying the recognition of bisyllabic words.

Eye movement strategies in word recognition

Studying gaze duration as a function of the position initially fixated by the eye has proved a useful way of probing word recognition processes. However, further information can be gained by looking at the detailed eye movement strategies used by readers in recognizing words.

Nevertheless the analyses of these strategies have proved to be rather complicated, and I wish here only to look at a particularly interesting case, namely what happens when the eye makes exactly two fixations in the word. This case will occur frequently when the eye initially fixates a position which is not optimal. The question now arises: what does the eye do? Does it immediately move to the 'optimal' position? Actually, although at first sight appealing, this idea makes little sense, since how can the eye know in advance where the optimal position is?

Figure 9 shows what actually happens for the case of 11-letter words. These histograms are reanalysed from the data of O'Regan *et al.* (1984; see also O'Regan and Lévy-Schoen, 1987). The arrow below each histogram shows the position where the eye started in the word, and the histogram shows the distribution of places where the eye landed on the *second* fixation. It is clear that the second fixation is not always at the same position in the word. It looks as if it is placed in such a way that the two fixations taken together allow the word to be evenly covered, as though processing of the word were distributed over the two fixations.

The idea that processing might be distributed over the two fixations occurring in the word seems to be confirmed when we look at the durations of first and second fixations. The solid line in figure 10 plots an example of the duration of the first fixation as a function of where this fixation occurred. Above each x-position plotted, I have plotted the duration of the corresponding second fixation (even though this second fixation occurred somewhere else in the word!). I have done this so that durations of first and corresponding second fixation can be compared: the striking fact emerges that they are perfectly complementary. When first fixation is long, second is short, and vice versa.

This is quite compatible with the idea of distributed processing. The time spent on the first fixation is saved on the second.

Unfortunately, further work has shown that things are not so simple. The shape of the first fixation curve actually reflects purely oculomotor

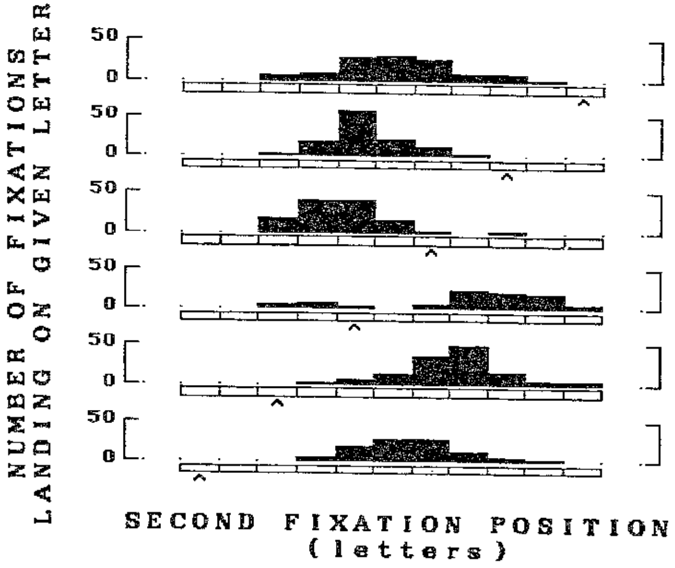

Figure 9: Histograms showing the position the eye landed on its second fixation in an 11-letter word when its first fixation was at various positions, shown by the arrow under each histogram. Data replotted from O'Regan and Lévy-Schoen, (1987).

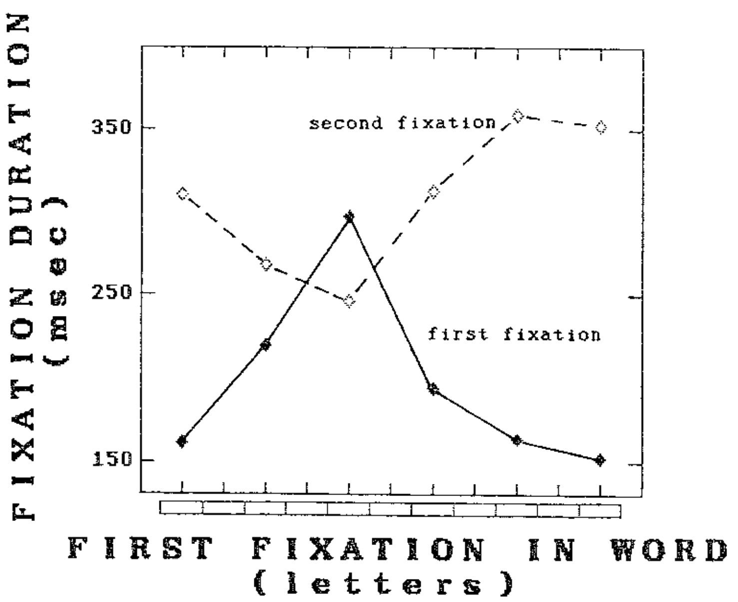

Figure 10: First and second fixation durations on a word for the case when precisely two occur. First fixation duration is given as a function of the position where that fixation occurs. But second fixation duration is given as a function of the position where the first fixation occurred. Data replotted from O'Regan and Lévy-Schoen, (1987).

processes, and not the amount of processing that has been done. Figure 11

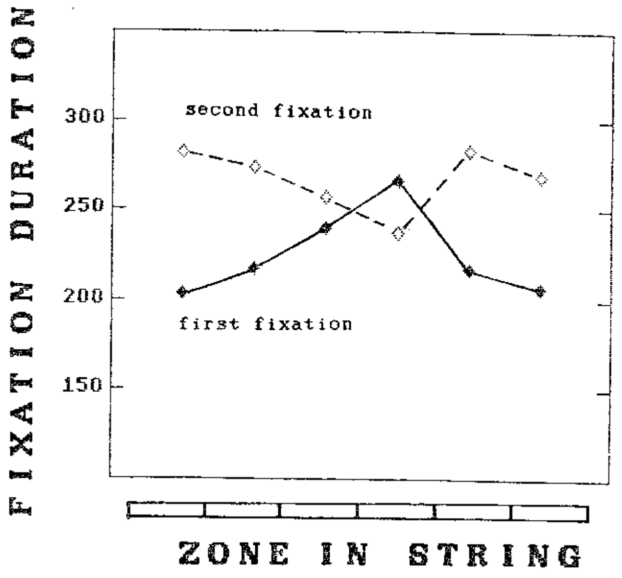

Figure 11: As in figure 10, but when scanning not a word, but a string of 13 x's with the task of fixating its middle before moving to an eccentric stimulus. The abscissa is the zone in the string where the eye was fixating when it appeared on the screen.

shows the results of an experiment in which the subject was asked to fixate the middle of a string of x's, and then to move on to another string. Just as in the optimum viewing position experiments I have been describing, the subject's eye was started off at different positions in the string. In this purely oculomotor task where no lexical processing is being done, curves very similar to those of the previous figure are found.

The previous figure (figure 10) can thus not be taken as evidence that distributed processing is occurring. An attempt to explain eye movement strategies within words was presented in O'Regan and Lévy-Schoen (1987). Here it will suffice to say that in addition to visual and lexical factors, oculomotor factors must also be taken into account when attempting to understand word recognition.

5 Conclusion

The research described in this chapter has spanned studies concerned with lexical, visual, and oculomotor factors affecting word recognition. *All three* aspects combine to create the impression of 'seeing' a word.

Because of the very strong drop-off of visual acuity even within the fovea, knowledge of the lexicon contributes in an important way to word recognition, particularly for words of length five letters or more. Work on exactly how lexical constraints intervene has been progressing rapidly over the last years. However, many questions remain to be answered, such as the kind of preprocessing done by the visual system, the precise kinds of features that are used in letter and word recognition, their degree of invariance to spatial transformations, what the higher level codes involved are, the exact way in which competing units interact, etc. The discovery of orthographic neighbourhood effects promises to provide a new tool with which to probe the 'circuitry' involved in word recognition. This in turn may provide a useful model for the first steps towards the problem of object recognition in general.

The section on acuity in word recognition showed the importance of taking into account the severe drop-off of the eye's power of resolution even within the fovea. The simple postulate of the concept of 'grain' of a typography was sufficient to make quantitative predictions about the number of letters that can be seen at a single fixation of the eye, and about the dependence of his number on viewing distance, character size and degree of defocussing. The 'grain model' was, however, a first, highly simplified step, and further work could develop the model to more accurately take into account the action of lateral masking. This might enable the model to make predictions about the effect of varying character spacing. One of the interesting insights that the model gave was the fact that, given the particular non-homogeneous structure of the retina, the amount of material that can be encompassed within a single fixation of the eye is largely independent of the viewing distance. For example, if I am looking for a ring that I have lost on the floor, then the radius of the region on the floor within which the ring will be visible is, within certain limits, independent of the height at which I place my eye. It may be that this property of the visual system constitutes a kind of pseudo-invariance that ensures that the amount of material to be processed at each fixation does not depend on the viewing distance.

Another consequence of the acuity characteristics of the retina is the existence of the 'optimal viewing position' in words: the position in a word where the eye must fixate first in order for recognition to be most rapid. This highly important discovery provides a new basis for understanding eye movements in reading. Further, the concept of an optimal position may be extendable to object recognition. Indeed, if the penalty for not first fixating in the optimal position is as large in object recognition as in word recogntion, then this concept may prove to be extremely important in understanding scene analysis, and also in the ergonomics of visual displays.

Eye movement strategies in visual perception have not received as much attention as they merit. The work described very briefly here has shown that eye movements in word recognition are a combination of automatic, preprogrammed oculomotor strategies that can act quickly on the basis of low-level sensory information, plus a kind of slower-acting control that makes use of linguistic information. Further work on the interplay between these two kinds of control will undoubtedly provide insights into reading and scene recognition.

References

Andriessen, J.J. and Bouma, H. (1976) Eccentric-vision: adverse interactions between line segments, *Vision Research, 16*, 71-78.

Blommaert, F.J.J. and Roufs, J.A.J. (1981) The foveal point spread function as a determinant for detail vision, *Vision Research, 21*, 1223-1233.

Bouma, H. (1970) Interaction effects in parafoveal letter recognition, *Nature, 226*, 177-178.

Bouma, H., Legein, Ch.P., Mélotte, H.E.M., and Zabel, L. (1982) Is large print easy to read? *IPO Annual Progress Report, 17*, 84-90.

Bourdon, B. (1902) *La perception visuelle de l'espace*, Paris: Scheicher.

Bouwhuis, D.G. (1984) Visual recognition of bisyllabic words, *IPO Annual Progress Report, 19*, 90-96.

Bouwhuis, D.G., and Bouma, H. (1979) Visual word recognition of three-letter words as derived from the recognition of the constituent letter, *Perception and Psychophysics, 25*, 12-22.

Carr, T.H. (1986) Perceiving visual language. In: K.R. Boff, L. Kaufman and J.P. Thomas, (eds), *Handbook of perception and human performance, Vol II*, New York: John Wiley.

Chambers, S.M. (1979) Letter and order information in lexical access, *Journal of Verbal Learning and Verbal Behaviour, 18*, 225-241.

Coltheart, M., Davelaar, E., Jonasson, J.T. and Bresner, D. (1977) Access to the internal lexicon. In: S. Dornic (ed.),*Attention and Performance VI*, New York: Academic Press.

Dow, B.M., Snyder, R.G., Vautin, R.G. and Bauer, R. (1981) Magnification factor and receptive field size in foveal striate cortex of the monkey, *Experimental Brain Research, 44*, 213-228.

Erdmann, B. and Dodge, R. (1898) *Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage*, Halle: Niemeyer.

Essen, D.C., van, Newsome, W.T. and Maunsell, J.H.R. (1984) The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability, *Vision Research, 24*, 429-448.

Fendick, M. and Westheimer, G. (1983) Effects of practice and the separation of test targets on foveal and peripheral stereoacuity, *Vision Research, 23*, 145-150.

Flom, M.C., Weymouth, F.W. and Kahnemann, D. (1963) Visual resolution and contour interaction, *Journal of the Optical Society of America, 53*, 1026-1032.

Henderson, L. (1982) *Orthography and word recognition in reading*, New York: Academic Press.

Jacobs, R.J. (1979) Visual resolution and contour interaction in the fovea and periphery, *Vision Research, 19*, 1187-1196.

Klein, S.A. and Tyler, C.W. (1981) Phase discrimination using single and compound gratings, *Invest. ophthal. Visual Science, Suppl. 20*, 124.

Korte, W. (1923) Über die Gestaltauffaung im Indirekten Sehen, *Zeitschrift für Psychologie, 93*, 17-82.

Le Grand, Y. (1964) *Optique Physiologique. Tome I: La dioptrique de l'oeil et sa correction*, Paris: Masson.

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

Marslen-Wilson, W.D. and Welsh, A. (1978) Processing interactions and lexical access during word recognition in continuous speech, *Cognitive Psychology, 10*, 29-63.

McClelland, J.L. and Rumelhart, D.E. (1981) An interactive activation model of context effects in letter perception: Part I. An account of basic findings, *Psychological Review, 88*, 375-407.

Morton, J. (1969) Interaction of information in word recognition, *Psychological Review, 76*, 165-178.

Nes, F.L. van and Jacobs, J.C. (1981) The effect of contrast on letter and word recognition, *IPO Annual Progress Report, 16*, 72-80.

O'Regan, J.K., Lévy-Schoen, A. and Jacobs, A.M. (1983) The effect of visibility on eye movement parameters in reading, *Perception and Psychophysics, 34*, 457-464.

O'Regan, J.K., Lévy-Schoen, A., Pynte, J. and Brugaillère, B. (1984) Convenient fixation location within isolated words of different length and structure, *Journal of Experimental Psychology, Human Perception and Performance, 10*, 250-257.

O'Regan, J.K. and Lévy-Schoen, A. (1987) Eye movement strategy and tactics in word recognition and reading. In: M. Coltheart (ed.), *Attention and Performance XII: The Psychology of Reading*, Hillsdale, NJ: Lawrence Erlbaum Associates, 363-383.

Osterberg, G. (1935) Topography of the layer of rods and cones in the human retina, *Acta Ophthalmologica, Suppl. 65*, Copenhagen, 1-102.

Paap, K.R., Newsome, S.L., McDonald, J.E., and Schvaneveldt, R.W. (1982) An activation-verification model for letter and word recognition: the word-superiority effect, *Psychological Review, 89*, 573-594.

Reicher, G.M. (1969) Perceptual recognition as a function of meaningfulness of stimulus material, *Journal of Experimental Psychology, 81*, 274-280.

Rolls, E.T. and Cowey, A. (1970) Topography of the retina and striate cortex and its relationship to visual acuity in rhesus monkeys and squirrel monkeys, *Experimental Brain Research, 10*, 298-310.

Rosenblatt, F. (1962) *Principles of neurodynamics*, New York: Spartan.

Rovamo, J., Virsu, V. and Nasanen, R. (1978) Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision, *Nature, 271*, 54-56.

Savin, H.B. (1963) Word frequency effects and errors in the perception of speech, *Journal of the Acoustical Society of America, 35*, 200-206.

Schwartz, E.L. (1980) Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding, *Vision Research, 20*, 645-669.

Selfridge, O.G., and Neisser, U. (1960) Pattern recognition by machine, *Scientific American, 203*, 60-68.

Tootell, R.B., Silverman, M.S., Switkes, E. and De Valois, R.L. (1982) Deoxyglucose analysis of retinotopic organization in primate striate cortex, *Science, 218*, 902-904.

Townsend, J.T., Taylor, S.G. and Brown, D.R. (1971) Lateral masking for letters with unlimited viewing time, *Perception and Psychophysics, 10*, 375-378.

Underwood, G., Hyona, J. and Niemi, P. (1987) Scanning patterns on individual words during the comprehension of sentences. In: J.K. O'Regan and A. Lévy-Schoen (eds) *Eye movements: from physiology to cognition*, Amsterdam: Elsevier.

Westheimer, G. (1982) The spatial grain of the perifoveal visual field, *Vision Research, 22*, 157-162.

Wilson, H.R. and Bergen, J.R. (1979) A four mechanism model for threshold spatial vision, *Vision Research, 19*, 19-32.

Woodworth, R.S. (1938) *Experimental psychology,* New York: Holt, Rinehart and Winston.

Convenient Viewing and Normal Reading

Martin M. Taylor[*]

1 Introduction

This paper relates O'Regan's presentation at this workshop to processes involved in normal reading. He presented studies on the Convenient Viewing Position (CVP), and has related the results to the fall-off in both visual and cortical acuity as distance from the fovea increases. His results and analysis may relate only to the less important of two kinds of process in normal reading. In normal reading, slow but accurate processes normally cooperate with fast but less reliable processes. The latter do not always require individual letters to be identified. Letters are harder to identify parafoveally than foveally in the neighbourhood of other letters, because of lateral interactions that increase with distance from the fovea. Although these interactions reduce the effectiveness of the slow processes, they may facilitate the fast processes. Lateral interactions may have evolved, long before reading, to permit rapid identification of potentially important patterns that may demand to be analysed at leisure using foveal vision; such a facility would be helpful in normal reading by rapidly suggesting a sequence of places worthy of slower detailed analysis. Only the slow processes use all the information about the letters and specifically about the order of the letters; these are the processes tapped by O'Regan's studies on CVP and Preferred Viewing Position (PVP). In normal reading, the PVP is at or near the informational centre of gravity of the word, but the CVP is well to the right of it. This discrepancy may accommodate both fast processes that are used for all words and slow processes hat might be needed for some words.

In this discussion I will try to relate the work on CVP described by O'Regan to the processes involved in normal reading. I take for granted that the experiments have been properly performed and described. O'Regan points out that both retinal and cortical acuity drop off nearly linearly as a function of distance from the fovea, but that cortical acuity drops off faster,

---

[*] Defence and Civil Institute of Environmental Medicine, P.O.Box 2000, Downsview, Ontario M3M 3B9, Canada