MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 593   September, 1980

TOWARD A COMPUTATIONAL THEORY OF EARLY VISUAL PROCESSING IN READING

Mike Brady

ABSTRACT: This paper is the first of a series aimed at developing a theory of early visual processing in reading. We suggest that there has been a close parallel in the development of theories of reading and theories of vision in Artificial Intelligence. We propose to exploit and extend recent results in Computer Vision to develop an improved model of early processing in reading. This first paper considers the problem of isolating words in text based on the information which Marr and Hildreth's (1980) theory asserts is available in the parafovea. We show in particular that the findings of Fisher (1975) on reading transformed texts can be accounted for without postulating the need for complex interactions between early processing and downflowing information as he suggests. The paper concludes with a brief discussion of the problem of integrating information over successive saccades, and relates the earlier analysis to the empirical findings of Rayner.

1. Introduction

This paper presents computational and psychophysical evidence in support of a theory of one of the earliest stages of visual processing in reading, namely the isolation of words in text. As such it is the first step in the development of a computational theory of reading whose general direction is presented in the next section. A skeletal outline of the paper follows.

The goal of reading may be supposed to be the efficient extraction of meaning from imaged text. Realising this goal involves integrating "upward flowing" information uncovered by early visual processing with "downward flowing" cognitive interpretations. In this paper, we present an approach toward understanding the **visual** aspects of reading which we believe may contribute greatly to an understanding of the overall reading process.

Existing theories of reading have relied on a primitive model of early visual processing. We suggest that as a result they have typically accorded too much emphasis to the role of "downward flowing" cognitive information, in effect suggesting that its deployment is necessary for almost every aspect of reading. Indeed, over the past two decades there has been a close parallel between the development of theories of reading and theories of visual perception in Artificial Intelligence (AI). In particular, we note that a number of reading theorists have recently been attracted to complex processing models developed in AI. A major attraction of such models is that they seem to provide a mechanism supporting flexible behavior by which information available as a result of early visual processing could combine with downflowing information about the specific image domain to produce an interpretation or percept. Still more recently, AI has witnessed a fascination with relaxation style processing. This is not only claimed to support the interaction between low level and downflowing information, but to do so by local parallel interaction. A number of reading theorists have proposed similar mechanisms. For the most part, these theories have had limited success in explaining the empirical psychophysical data on reading. We argue that this is, in part, because they depend upon a primitive model of early visual processing. It is also partly because of an emphasis on the *mechanism* of integrating information from various sources, without addressing the issues of what purpose the information serves, what is the information which is passed, and how it is represented (see Marr, 1980, Marr and Nishihara, 1978).

Over the past few years there has been considerable progress in understanding early visual

processing. The achievements of Horn, Marr, Poggio, Ullman, and others in developing a computational theory of natural visual perception has little or no counterpart in theories of reading. For example, Frisby (1979, page 108) and Allport (1980, page 235) equate early processing with feature extraction as developed in optical character recognition systems (Duda and Hart, 1973). A fuller account of the relevant empirical findings is given in Cohen (1978, page 65), but her analysis falls considerably short of a being a precise and coherent theory. The computational theory of natural vision suggests that much richer information can be made available by early visual processing in reading, without the aid of downward flowing "higher level" knowledge of the domain being viewed. Reading has always attracted a great deal of attention from perceptual psychologists, in part because of the light it might shed on our understanding of human perception of the natural world. We claim that, temporarily at least, the boot is on the other foot, and that the recent developments in our understanding of real world perception can be gainfully applied to increase our understanding of reading.

Finally, we review some empirical findings about the earliest stages of visual processing in reading, and we settle upon the isolation of words as the first goal of the reader's perceptual processing. We note that eye movement studies show that a great deal of processing is carried out on text prior to foveation. It follows that it is reasonable to conjecture that word isolation is effected on the basis of information available in the parafovea. As part of an investigation of this conjecture, we suggest that Fisher's(1975) results on transformed text provide some insight into parafoveal word isolation, and so we analyze his results carefully. We argue that they can be explained on the basis of Marr and Hildreth's(1980) theory of edge detection without postulating the need for *"higher order visual processing"* as was claimed by Fisher. The explanation leads to a number of empirical predictions, which are confirmed using Fisher's own methods and materials. The concluding section sketches a theory of word isolation in the parafovea, and notes that the decision to activate the reading process in the first place is also not very mysterious.

2. Background to the study

2.1 Past approaches to theories of reading

From the earliest days of experimental psychology there has been a constant stream of research

findings about reading (see for example, Huey, 1908, Henderson 1977). All of the major schools of perception have considered reading to some extent, and have attempted to exploit various mathematical and computational insights to develop their theories. We are particularly concerned with the growth of interest over the past two decades, during which time a number of theories have developed, the majority being expressed in terms of *information processing.*

Relative to the Behaviorists' reliance on a simple mechanism, which bore many of the characteristics of early pattern recognition systems, and the extreme wordiness of the Gestalt and New Look theorists, information processing accounts of reading are refreshingly precise. They consist of individuated *stages*, at which some particular *functionally* defined 'process' is carried out (say to extract features or to consult a lexicon), together with interconnecting *arrows*, which represent the flow of information through the system under consideration. An important property of such models is that they describe the way in which a perceptual or cognitive process being studied unfolds over time. The particular class of individuated stage processes, and the topology of interconnecting arrows, are carefully chosen to account for relevant empirical findings. While the power of such formalisms is clearly sufficient to account for any given set of descriptions, in the absence of a wholly precise mathematical or computational account of reading, any particular model is inevitably vague in places. The extent to which it does or does not adequately explain the available empirical data (and the precision of the predictions which can be made from it) are limited. For example, Gough(1972) presents a flow diagram of "one second of reading" which embodies the theory that phonological recoding is obligatory. Marcel and Patterson(1979) present an alternative in which it is not. For further examples, see Estes(1977), Cohen(1978), and McClelland and Rumelhart(1980).

The box and arrow diagrams which feature in most information processing accounts of perception are highly reminiscent of the system flowcharts which used to be prepared by programmers in the early stages of developing a program. Flowcharts have fallen into disrepute in computer science as it has been realized that they provide an impoverished representation of such a key issue as the structure of a program. They are also wholly inadequate as a representation of process interaction and parallelism, being essentially restricted to the description of a single sequential process. Of course, they are merely the simplest first approximation to a model of processing, though one should be aware of the Computer Science experience that they unacceptably straitjacket thinking.

PAST APPROACHES TO THEORIES OF READING

Several authors have argued that it is not possible to develop a theory of an ability such as reading, in which the flow of information is wholly unidirectional, that is, a flow that proceeds from the processes which embody relatively general knowledge, and which make contact with the intensity levels of the image to the processes embodying knowledge about the specific objects and situations depicted in the image (see for example Allport(1979), Frisby(1979), Cohen(1978), Rumelhart(1977)). It is supposed that "downward flow" of knowledge about such objects and situations is also necessary to account for the remarkable abilities and flexibility of human perception.

The invocation of "downward flow" as an explanation for reading abilities has an interesting (perhaps not co-incidental) parallel with the history of computational theories of natural visual perception in the field of Artificial Intelligence (AI). The period 1963 to the early 1970's in the development of AI was most notable for extensive experimentation with edge detecting or region finding operators, designed *ad hoc* in accordance with the needs of some particular project. Authors time and again noted that the results of applying their operators to digitized images were essentially unpredictable; many concluded that it was simply not possible to develop a theory of early visual processing capable of generating predictably rich and useful descriptions that could then be used as the basis for computing the visible surfaces and objects in a scene. It was supposed therefore that, just as in the case of reading (although the AI workers involved would not have known of the parallel), "downward flow" of knowledge about the objects and situations imaged in the scene was essential to explain the remarkable abilities of human visual perception. The interaction between upward flowing information generated by relatively unknowledgeable early processing modules and downward flowing information was essentially dynamically determined and could not be completely defined in advance. It was conjectured by Minsky and Papert(1972) that among the tools developed in computer science, the best way to achieve this dynamically determined behavior was through process interactions, which, it was noted, need not be restricted to the simple patterns of (serial) activity provided in a language like Fortran or Algol. These were the considerations which lay behind the development of a rash of complex "heterarchical" programs to understand natural language, perceive utterances from a speech signal, and see in various narrowly defined domains. Programs such as Hearsay 2 (Lesser and Erman, 1977), Margie(Schank et. al., 1973), Barrow and Tenenbaum's (1976) Interpretation Guided Semantics, and the author's own program for "reading" Fortran code (Brady,1979; Brady and Wielinga, 1978) are typical of the genre.

The development of complex "heterarchical" programs such as Margie and Hearsay 2 is paralleled by the adoption of those computational models of processing by reading theorists eager to explain the use of downward and upward flow as determinants of a percept. Examples are Cohen's (1978) discussion of Speechlis (Nash-Webber, 1975), and Allport's (1979) detailed explanation of the operation of Margie.

In fact, a number of difficulties emerged in the dynamic processing account of perception as soon as vague theoretical notions like "process interaction" needed to be made precise (see Brady, 1979). There are two basic difficulties, one technical, the other more empirical in nature though reflecting a theoretical shortcoming. Technically, the potency of process interactions, and the stock of ideas about how to control and analyze them, remain very limited indeed. Secondly, and most notably, the presumed power of heterarchy never materialized. It repeatedly became evident that a small increase in the early processing capabilities of programs could have a far greater impact on the performance of a program as a whole than a vastly greater amount of "higher level reasoning".

Consider in particular the case of Hearsay 2 (Lesser and Erman, 1977). One of the main innovations of Hearsay 2 was the introduction of a centralized data structure called the "blackboard", on which the findings of a number of "knowledge sources" (which performed such tasks as isolating phonemes, syllables, words, or larger syntactic units) were presented. At any stage of the processing of a speech signal corresponding to an utterance, the contents of the blackboard represented the state of the system's interpretation. The addition of a piece of information by one knowledge source could enable the activity of several others. At any given stage, there were typically many runnable processes (up to two hundred), each of which was assigned a numerical priority value indicating its apparent importance. This design is illustrated in figure 1a, which shows the Hearsay 2 system as of January 1976. The authors note that "this implementation had poor performance (eg 10% of sentences correct in 85 million instructions per second of speech on a 250 word vocabulary" Lesser and Erman 1977, page 790). A second design, shown in figure 1b, was aimed at "making the lower levels of processing more sequential and bottom up" Lesser and Erman 1977, page 795. The authors reported that "this configuration performs substantially better (eg 85% correct in 60 million instructions per second of speech on a 1000 word vocabulary)" Lesser and Erman 1977, page 790.

Some AI researchers (see for example Davis and Rosenfeld 1978, 1981, Barrow and Tenenbaum
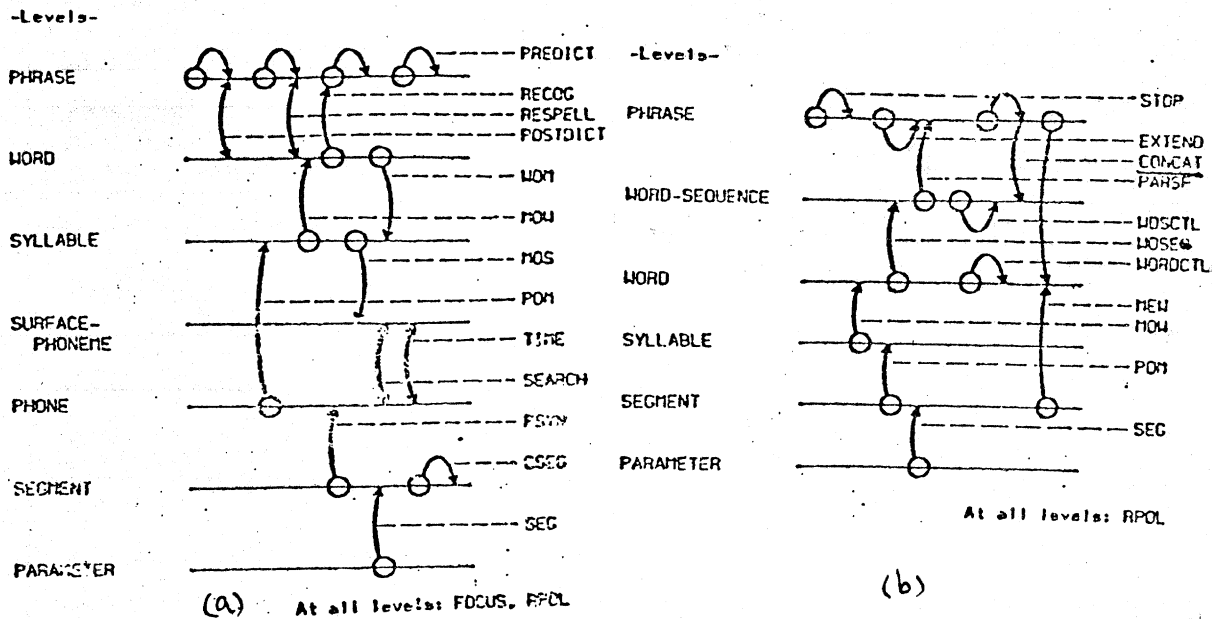
Figure 1. The structure of the blackboard state descriptor for the Hearsay 2 speech understanding system. Figure 1a: the system as of January 1976. Figure 1b: the second version as of September 1976 (Reproduced from Lesser and Erman 1977)

1978, Rosenfeld, Hummel, and Zucker, 1976, Waltz, 1978, Zucker 1978) concluded that the main drawback of the heterarchical process organisations discussed above was that they were essentially serial. They argue that much of their complexity arises because one is forced to choose a particular sequential order in which to carry out a number of processes. Since this order is inevitably often inappropriate (being unpredictable), one is then required to incorporate sufficient mechanism to facilitate recovery. Instead, such authors suggest the use of globally constrained local parallel processes, usually based on relaxation or other forms of nonlinear programming (see Luenberger, 1973). Note that in common with the heterarchy approach, the structure of the mechanism is developed and fixed in advance of the analysis of the particular perceptual problem being studied. The only issues which the theorist is left to settle in most accounts are parameter settings, such as the size of neighborhoods, thresholds, and the like (see Davis and Rosenfeld, 1981). We argued above that a major drawback with heterarchical accounts of perception was the difficulty in analysing and controlling them. It is important to realise that analogous problems arise with relaxation processes. It is usually extremely hard to guarantee that such a process settles down to a steady state ("converges"). As an example, consider the difficulty that Marr, Palm, and Poggio (1978) had in analysing the behavior of the Marr and Poggio(1976a,1976b) cooperative algorithm for computing stereo disparity. If this is difficult for

a single level of relaxation processing, it is considerably more so for the hierarchical or multi stage processes which have been advanced, though usually not implemented and tested, in the literature (eg McClelland and Rumelhart, 1980, Davis and Rosenfeld, 1978, Zucker, 1978). Few (if any) results are known regarding the convergence (including speed of convergence) of such relaxation processes (see Ullman, 1979, Zucker, Leclerc, and Mohammed, 1979). Without such results, the uncritical proposal of complex locally parallel processes is of questionable significance.

2.2 The computational approach to vision

Against this background of *ad hoc* experimentation and the construction of uncontrollable complex processing models in Artificial Intelligence, the computational theory of natural visual perception developed by Horn, Marr, Ullman, Poggio, Binford, and others is quite remarkable. A fuller account of the current state of computer vision can be found elsewhere (Marr, 1980, Brady, 1981, Horn, 1978, Marr and Poggio, 1979, Marr and Hildreth, 1980, Grimson, 1980). For the purposes of this article, it is sufficient to note that there now are mathematically precise theories and highly parallel, robust computer implementations of a variety of (human) visual processes. These include edge detection, stereopsis, shape from shading, shape from texture, early motion detection, and surface interpolation. In each case these theories concern processes which occur at an early stage of perception, and they embody knowledge about the world which is of considerable generality, for example that the world mostly consists of smooth surfaces. In short, the computational theory of vision is a compelling argument in support of the power of early visual processing. More significantly perhaps, it promotes a research methodology which defers consideration of knowledge rich, domain specific, downward flow of information until the considerable scope of early processing is more clearly understood. It also makes little sense to develop an understanding of the role of downward flow until we have a better appreciation of what information early processing can and does provide.

The computational theory of visual perception referred to above is also interesting for the research methodology which has developed from it. The first step is to isolate a perceptual ability for which there is empirical evidence for considerable competence on the basis of early processing. For example, Horn(1974) has studied the determination of lightness and the computation of and shape from shading (1978) from an image. Marr and his colleagues have considered edge detection (Marr and Hildreth, 1979), stereopsis (Marr and Poggio (1979), Grimson (1980)), and motion computation (Ullman (1978), Marr and Ullman (1979), Ullman and Richter (1980)). The particular problem is

7

then studied in three stages. First, we consider *what* information must be extracted from the scene, in order for the system to exhibit this competence, and *what constraints* on the world the system needs to assume in order to extract this information. The next step is to devise a representation which makes explicit the information required to explain the competence. Only then is it reasonable to devise algorithms to discover the appropriate representation instance for a scene. Finally, one can conduct experiments to discover the extent to which the algorithm explains human performance. Notice that in contrast to this methodolgy, the heterarchical and relaxation processes outlined above *start* with an algorithm (or commitment to a particular restricted kind of processing) and only then examine competence, devise representations, and analyze the basis of the competence.

2.3 Edge detection in the human visual system

As an example of the results of the computational approach to early visual processing, we take a brief look at Marr and Hildreth's (1980) theory of edge detection. The reason for this choice is quite simple. The theory addresses the very first stage of analysis of the visual input, and this is the stage which is most relevant to the study of parafoveal processing in reading which is presented in the balance of the paper.

Marr and Hildreth (1980, page 189) point out that "a major difficulty with natural images is that changes can and do occur over a wide range of scales, so it follows that one should seek a way of dealing with the changes occuring at different scales." One way to do this, which has been proposed several times in the image processing literature, is to pass the image through a number of band limited filters. Of course, the difficult issues concern the choice of filters (bar mask, Fourier, Gaussian), the number of them, and the exact band pass characteristics of each.

In fact, intensity changes are mostly localised in space, a fact which can be explained by their physical causes (see Horn (1977), Marr (1976), Marr and Hildreth(1980, page 189)). They are also localised in the frequency domain, since the world is mostly composed of visible surfaces of roughly uniform texture. Marr and Hildreth (1980, page 191) note that "unfortunately, these two localization requirements, the one in the spatial and the other in the frequency domain, are conflicting". They point out that the Gaussian optimises localisation in both domains simultaneously, and so it is chosen as the band limiting filter in the theory.

In order to locate edges, one can either find places where the first derivative of the intensity function reaches a maximum, or equivalently where the second derivative is zero. To locate edges at

arbitrary orientations with equal facility, we require a differential operator which is not directional. The Laplacian is the only first or second order differential operator with this property. Thus the Marr and Hildreth theory asserts that following Gaussian smoothing, the image is convolved with a Laplacian and zero crossings noted. In fact, by the so-called convolution theorem,

$$\nabla^2(G*Image) = (\nabla^2 G)*Image,$$

where G is a Gaussian operator, and * denotes convolution. Marr and Hildreth(1980, page 193) point out that the $\nabla^2 G$ operator closely resembles the difference of Gaussian (DOG) operators proposed by Wilson and Giese (1977) (see also Wilson and Bergen, 1979). Indeed they show that $\nabla^2 G$ is the limit of a DOG, and that the DOG closely approximates it. Wilson and Bergen's work suggests that there should be four bandpass channels at each retinal eccentricity, and that their characteristic sizes should scale linearly with eccentricity, being smallest in the fovea and doubling in size by about 4°. Recently, Marr, Hildreth, and Poggio (1979) have noted evidence for a fifth, smaller channel in the fovea, and Stevens (1980) has shown that the fifth, finest resolution channel plays the most important role in determining the information we compute foveally.

We can compute the width of the finest resolution channel at any eccentricity ϵ. If we digitise a text image, say at a resolution of 100 microns, we can compute the size of mask to use in a computer program which precisely models the information available in the finest resolution channel at eccentricity ϵ. Examples of the result of applying this process can be found in figure 6.

3. The isolation of words in text

3.1 Introduction

It is usual to equate early processing in reading with the extraction of character features, such as line endings, T-junctions, holes, and concavities. We are presently more concerned with an even earlier processing stage, namely the point at which the visual system first makes contact with (the gray level intensities forming the image of) a portion of text. Let us suppose for the moment that the "reading process" is already active. The work of Rayner (1975a, 1975b, 1977, 1978a, 1978b, 1979, Rayner and McConkie 1976, Rayner, McConkie, and Ehrlich, 1978, McConkie and Rayner, 1975) and others (see for example McConkie(1979), O'Regan(1979), Levy-Schoen and O'Regan(1979)) on eye movements demonstrates clearly that text is substantially processed before it is foveated. The extent to which

eye movement control is either (1) autonomous, being entirely determined by information computed by early processing from the gray level array; or (2) is capable of being explicitly controlled by downward flowing task specific information, say, by knowledge of the syntax and semantics of the text in question, is controversial. This is, of course, the invariance of the issue raised in section 2.1 about system organization.
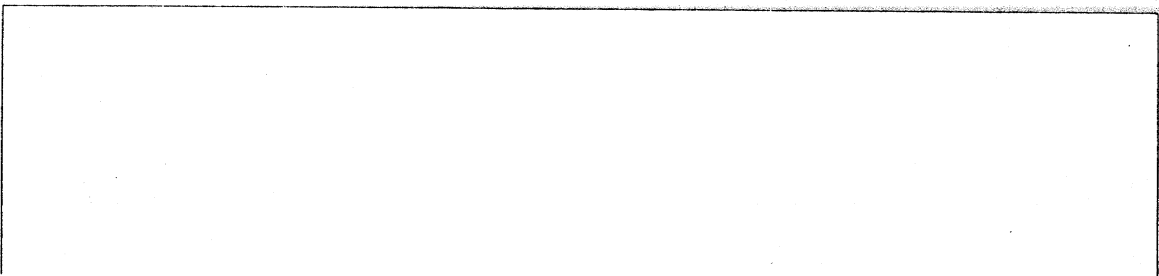
The goal of reading may be supposed to be the efficient extraction of meaning from imaged text. Given the nature of written language, particularly English, a presumably necessary primitive subgoal is the isolation of words. In normal text, words are clearly separated by spaces which are substantially wider than the spaces between individual letters. It would seem that the "program" controlling eye movements could be trivial given a reasonable theory of the separation of words from inter-word spaces such as that provided by the Marr Hildreth theory outlined in the previous section. Evidence in support of the contention that the control program is quite simple is easy to find. Firstly, it is well known that inter-word spaces, even when they are of varying width, are never foveated (Levy-Schoen, 1979). Conversely, if spaces corresponding to word boundaries are randomly introduced into previously elided text (as shown in figure 2), reading becomes exceptionally difficult. In this situation, the inconsistent information provided by a simple space finding algorithm and its utilisation by the processes which analyze the text, produce a complex pattern of foveations and a significant increase in the duration of any individual foveation. Intermediate behavior results when inter-letter spaces are made nearly equal to those between words.

However, as is equally well known, spaces are not unique in avoiding foveation. In particular, function words such as "and" and "the" are rarely foveated. This partly explains the difficulty difficulty we have in proof reading "Paris in the the spring" relative to this sentence as a whole. This raises the ever present question: how "intelligent" does the eye movement controller need to be? Is the word "the" omitted on the basis of information available in the parafovea, where individual letter recognition is poor (Bouma, 1971), or alternatively does it rely on knowledge about the linguistic context?

3.2 Fisher's results on reading transformed text

In fact, the trivial word isolation process sketched above does not work in every circumstance in which people can read quite easily. This was demonstrated in an elegant experiment performed by

10

```
Itn owb eca mee vid ent tha tth eci tym ust bea ban don eda ton ceT her
awa sad iff ere nce ofo pin ion inr esp ect tot heh our ofd epa rtu reT hec
ayt ime itw asa rgu edb yso mew oul dbe pre fer abl esi nce itw oul den abl
```
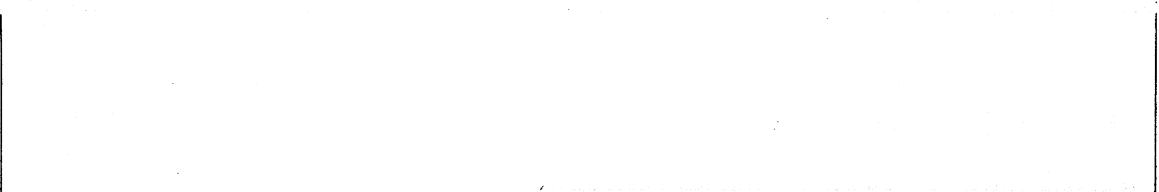
Figure 2. Text into which spaces have been randomly introduced after elision

Fisher, 1975. Building upon the earlier work of Smith, 1969 and Hochberg, 1970, Fisher used the transformed texts illustrated in figure 3 to investigate the effect of manipulations of word shape and word boundary on reading. Word shape was "manipulated" via three *type variations*: normal, all upper case, and alternating upper and lower case letters. These are illustrated in samples one to three of figure 3. *Word boundaries* were also "manipulated" in three ways: normal spacing, replacing an inter-word space by the filler character "+" or "@", and elision to remove inter-word spaces. These manipulations are illustrated for the upper case type variation in samples two, five, and eight of figure 3. In all, there are nine possible type and word boundary combinations, and they are shown in figure 3.

(Fisher 1975, page 189) recorded the length of time taken by subjects to read nine paragraphs of approximately equal length and complexity, whose texts had been randomly manipulated in the ways described above. As a safeguard against skim reading without understanding, a subject was required to answer a number of questions (typically four) about the passage just read, and was required to get a certain number correct for the data point to be recorded. The results are presented in figure 4.

Fisher 1975, page 189 noted that the "interdependence of cues causes a reduction in reading speed to nearly one third of the speed of the separate cue manipulations", and he suggested that this "*interdependence of word shape and word boundary cues tends to implicate higher order visual*
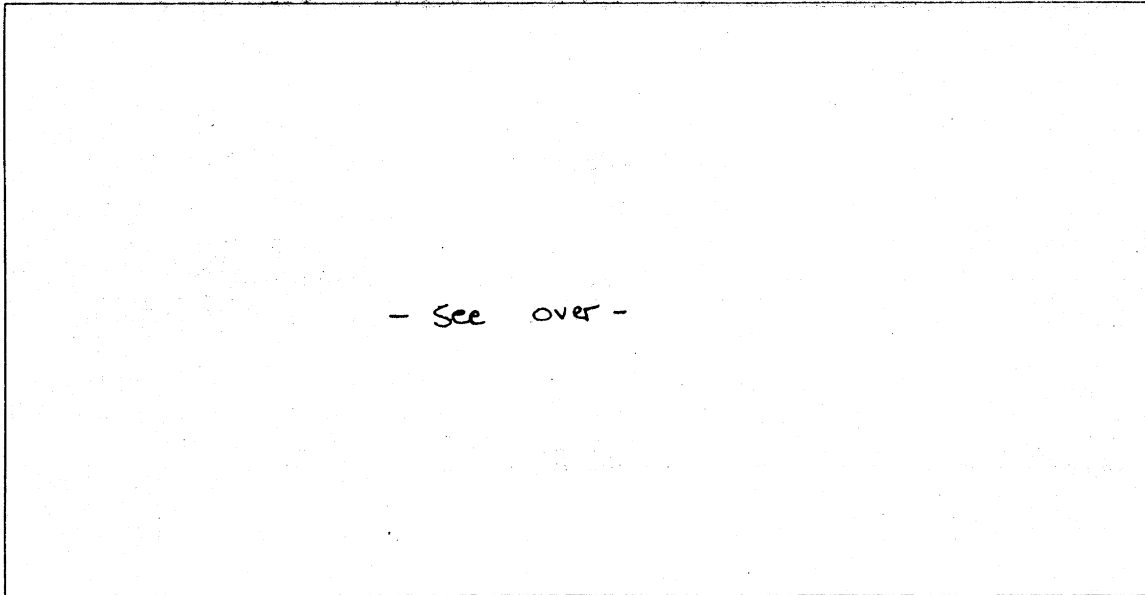
11

— See over —

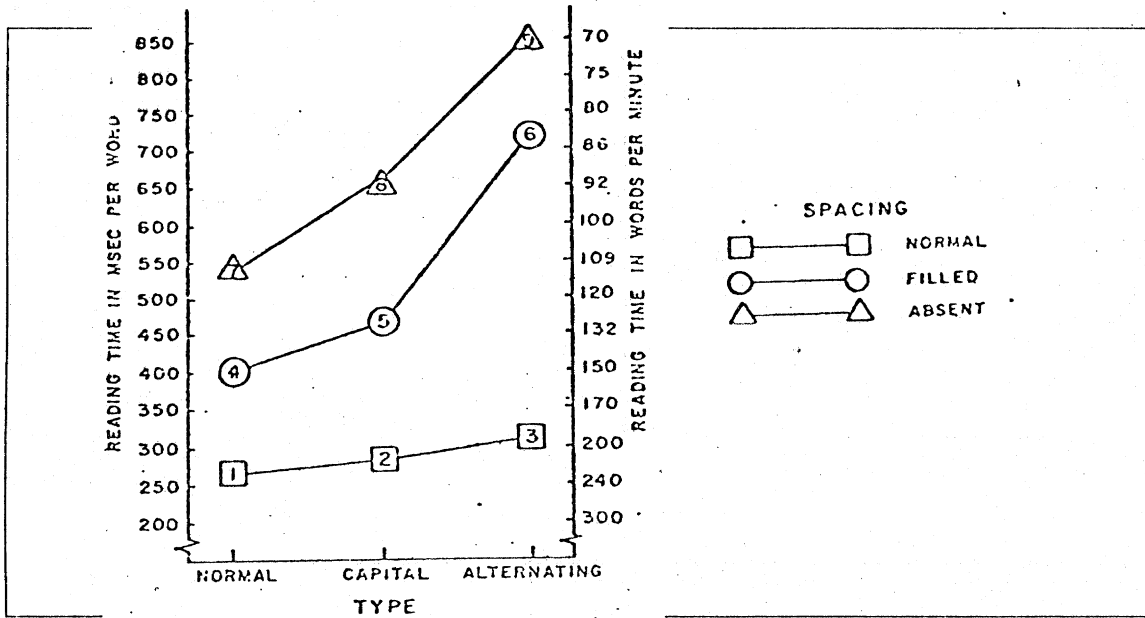Figure 3. The nine type and boundary variations used by Fisher

Figure 4. Fisher's results, reproduced from Fisher 1975, page 189

processing than might be required simply for word identification" Fisher 1975, page 190.

3.3 The role of early visual processing in the isolation of words in text

In the Introduction, we commented on the difficulty of devising and controlling processes which

Figure **3** *Fisher's mutilated texts*
Source: *Fisher* (1975)

1

The government of Henry the Seventh, of his son, and
of his grandchildren was, on the whole, more arbitrary
than that of the Plantagenets. Personal character may

4

The+government+of+Henry+the+Seventh,+of+his+son,+and++++
of+his+grandchildren+was,+on+the+whole,+more+arbitrary++
than+that+of+the+Plantagenets.++Personal+character+may++

7

Thegovernmentof Henry theSeventh,ofhisson,andofhis
grandchildrenwas,onthewhole,morearbitrarythanthat
of thePlantagenets.Personalcharactermayinsomedegree

2

THE GOVERNMENT OF HENRY THE SEVENTH, OF HIS SON, AND
OF HIS GRANDCHILDREN WAS, ON THE WHOLE, MORE ARBITRARY
THAN THAT OF THE PLANTAGENETS. PERSONAL CHARACTER MAY

5

THE@GOVERNMENT@OF@HENRY@THE@SEVENTH,@OF@HIS@SON,@AND@@@@
OF@HIS@GRANDCHILDREN@WAS,@ON@THE@WHOLE,@MORE@ARBITRARY@@
THAN@THAT@OF@THE@PLANTAGENETS.@@PERSONAL@CHARACTER@MAY@@

8

THEGOVERNMENTOFHENRYTHESEVENTH,OFHISSON,ANDOFHIS
GRANDCHILDRENWAS,ONTHEWHOLE,MOREARBITRARYTHANTHAT
OFTHEPLANTAGENETS.PERSONALCHARACTERMAYINSOMEDEGREE

3

ThE GoVeRnMeNt oF HeNrY ThE SeVeNtH, oF HiS SoN, aNd
Of hIs gRaNdChIlDrEn wAs, On tHe wHoLe, MoRe aRbItRaRy
ThAn tHaT Of tHe pLaNtAgEnEtS. PeRsOnAl cHaRaCtEr mAy

6

ThE@GoVeRnMeNt@oF@HeNrY@ThE@SeVeNtH,@oF@HiS@SoN,@aNd@@@@
Of@hIs@gRaNdChIlDrEn@wAs,@On@tHe@wHoLe,@MoRe@aRbItRaRy@@
ThAn@tHaT@Of@tHe@pLaNtAgEnEtS.@@PeRsOnAl@cHaRaCtEr@mAy@@

9

ThEgOvErNmEnToFhEnRyThEsEvEnTh,oFhIsSoN,AnDoFhIs
GrAnDcHiLdReNwAs,oNtHeWhOlE,MoReArBiTrArYtHaNtHaT
oFtHePlAnTaGeNeTs.pErSoNaLcHaRaCtErMaYiNsOmEdEgReE

embody an interaction between upward flowing and downward flowing information, and argued for a model where early visual processing plays a bigger role. Since word isolation is clearly one of the first steps in reading, we start by examining Fisher's results more closely, in the hope of discovering an explanation of his findings without resorting to higher level cues. Firstly, the reading time per word in sample seven is significantly lower than that in sample eight. This might be explained on the grounds of the latter's lesser shape variability. However, sample nine has *greater* variability in shape than sample eight, and yet the time to read eight is significantly lower than that for nine. Similarly, there is greater variability in the shape of sample three than sample two, and yet the time to read three is significantly greater. Clearly, one possible explanation is that in the absence of spaces, capital letters can be used to signal word boundaries. According to this explanation, samples three and nine provide information (random capitals) about word boundaries inconsistent with that discovered by the processes which analyze the text. (Compare figure 2 and its discussion in the text). It would then follow that the eye guidance system could make the distinction between upper and lower case characters and makes use of that information in isolating words.

This leads to our first **empirical prediction**: if the paragraphs used by Fisher are transformed by first capitalizing the initial letter of each word and then eliding, so as to appear as in figure 5a, the resulting text should be significantly easier to read than the elided text sample shown in figure 5b (compare sample seven in figure 4). This prediction forms experiment one. The experimental details can be found in the next section. For the purposes of this section, it suffices to note that the experiments were designed strictly in accordance with the method devised by Fisher(1975) to maximise comparability with his results. Subjects were required to read texts which had been transformed in various ways similar to those shown in figure 4. The average reading time per transformed word was compared for significance between two variations. According to this metric, the phrase "significantly easier to read" means that the reading time per word was significantly shorter.

It turns out that the **capitalized elided text shown in figure 5a is indeed significantly easier (p < 0.01) to read than the elided normal text shown in figure 5b**. This supports the hypothesis that we are capable of distinguishing between upper and lower case characters on the basis of information available in the parafovea. Significantly, however, it leaves open the precise details of the way in which that distinction is made.

ItNowBecameEvidentThatTheCityMustBeAbandonedAtOnceThereWasADifferenceOf
OpinionInRespectToTheHourOfDepartureTheDaytimeItWasArguedBySomeWouldBe
PreferableSinceItWouldEnableThemToSeeTheNatureAndExtentOfTheirDangerAndTo

(a)

ItnowbecameevidentthatthecitymustbeabandonedatonceTherewasadifferenceof
opinioninrespecttothehourofdepartureThedaytimeitwasarguedbysomewouldbe
preferablesinceitwouldenablethemtoseethenatureandextentoftheirdangerandto

(b)

Figure 5. Typical data for Experiment one. Figure 5a: text which has been elided after capitalizing the initial letter of each word. Figure 5b: elided normal text like that in sample seven of figure 4.

Some evidence bearing upon this distinction can be gleaned from the results for samples five, six, eight, and nine in figure 4. Whereas sample five is significantly easier to read than sample eight, there is insignificant difference between the ease of reading samples six and nine. This is a puzzle. The advantage of sample five over sample eight suggests that we are capable of dynamically modifying our eye movement control system to exploit the delimiter "@", and this contention is supported by the significant advantage of sample four over sample seven. However, if we are capable of distinguishing upper case characters and the character "@" in the parafovea in a way which is entirely robust and reliable, we would expect to find a similar significant advantage for sample six over sample nine; but we do not. One possible resolution of this puzzle would be to show that it is often difficult to distinguish "@" and upper case characters when they are viewed in the parafovea. If that were so, the use of "@" as a filler would give some advantage in sample five relative to sample eight, but the advantage would be offset by the inconsistent information provided by fillers and text in sample six.

To investigate this question precisely, we need a detailed representation of the information which is actually available in the parafovea. Fortunately, such a representation is now available, having recently been developed by Marr and Hildreth (1980), and it was sketched in the previous section. Figure 6 shows the result of applying the digitisation process described in that section to
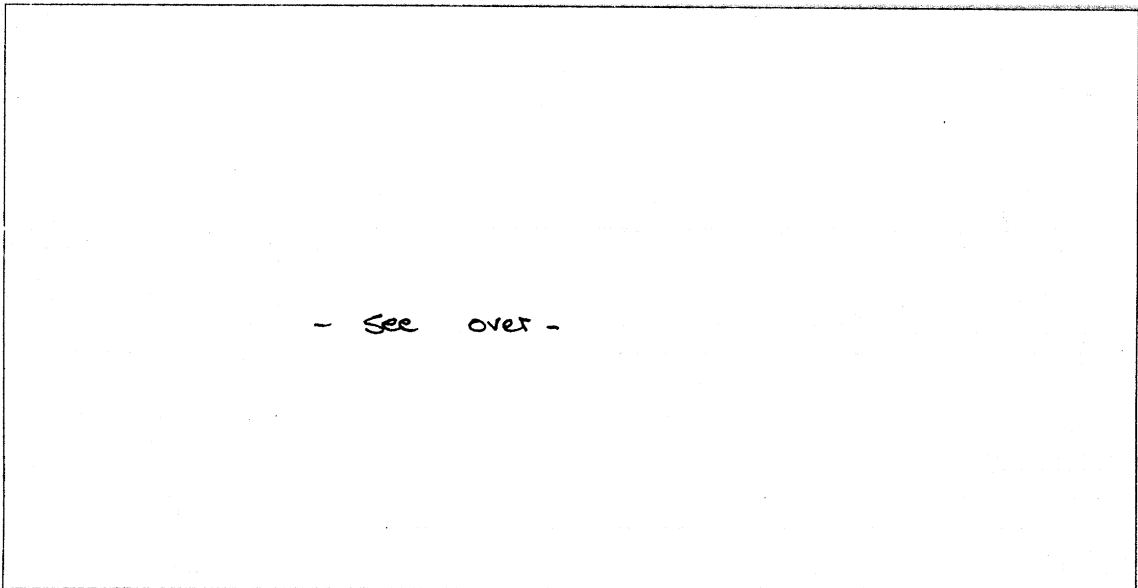
- See over -

Figure 6. The result of convolving sample five of Fisher's data to show the information available at 4°. Figure 6b: all instances of the character "@". Figure 6c: instances of the character "@" which are difficult to distinguish on the basis of shape.

sample five of Fisher's data (figure 4) at an eccentricity of four degrees. Figure 6b explicitly marks the convolved "@" characters. It can be seen quite clearly that while some of them are relatively easy to distinguish on the basis of shape, others (for example those marked in figure 6c) are not.

This evidence does indeed seem to show that it is often difficult to distinguish "@" and upper case characters when they are viewed in the parafovea. We suggest that this resolves the puzzle of Fisher's results discussed above without the need to postulate any downward flow of high level information. It further suggests that while upper and lower case characters can be clearly and reliably distinguished (in most fonts), the model of "upper case character" used by the early visual system in guiding eye movements is actually quite crude. Tentatively we may suppose that the model of an upper case character amounts to an assertion that they are relatively large compared to those in lower case and have relatively lower curvature. This simple model normally serves the reader well, since written text consists mostly of upper and lower case characters. However, being a simple model, it is easily confused, and is particularly unreliable at making the distinction between upper case characters and "@".

A number of predictions follow from this analysis. Firstly, it suggests that a font in which the

6a

THE COVENANT OF HENRY THE SEVENTH, OF MIS SON, OAS
HIS GRANDCHILDREN HAS, POS THE MIDLE, ONE MORE ARBITR

6b

THE COVENANT OF HENRY THE SEVENTH, OF MIS SON, OAS
HIS GRANDCHILDREN HAS, POS THE MIDLE, ONE MORE ARBITR

6c

THE COVENANT OF HENRY THE SEVENTH, OF MIS SON, OAS
HIS GRANDCHILDREN HAS, POS THE MIDLE, ONE MORE ARBITR

An inquiry which has just been
held at Brighton once more il-
lustrates the kind of leading
strings in which local munici-
palities are kept. An inspector
of the local government Board
has been holding a kind of pub
-lic inquest on the proposal of
the Brighton corporation to bor-
row 55,000*l*. This enterprising
public body in its desire to in-
crease the attractions of the
great sussex watering-place
has resolved to buy an estate

Figure 7. A font in which the distinction between upper and lower case would be difficult to make. It is reproduced from Spencer(1968, page 16), from which we quote: "A new kind of type proposed in the 1880's by Andrew Tuer in which 'the tailed letters projecting above or below the line, have been docked' to provide maximum type size 'where economy of space is an object - as in the crowded columns of a newspaper' ".

distinction between upper and lower case is difficult to make on the basis of size and shape would be quite hard to read. Figure 7 shows such a font. Indeed, as we point out in the Conclusion, the analysis here can be viewed as a first step towards making font design less subjective than it has been in the past (see for example Spencer(1968)). Secondly, the analysis suggests that on the basis of the information available in the parafovea, it would be difficult for the visual system to distinguish the capitalized elided text shown in figure 8a and the text filled with "@" shown in figure 8b. This translates into a prediction that there should be insignificant difference in the case, that is to say speed per word, of reading the samples in figure 8. Experiment 2 confirms this prediction; the relative advantage of one sample over the other failing to reach significance at the 10% level.

The same computational argument can be turned around, in which case it leads to the prediction that using a "visually striking" character as a filler would produce text that is significantly easier to read than when "@" is used. Indeed, insofar as this can be shown empirically, it *essentially enables us to frame a precise definition of "visually striking"*. In Experiment 3, we compare the effect of using "\" and "@" as fillers. The choice of "\" was quite deliberate. Figure 9 shows a sample of text which has

ItNowBecameEvidentThatTheCityMustBeAbandonedAtOnceThereWasADifferenceOf
OpinionInRespectToTheHourOfDepartureTheDaytimeItWasArguedBySomeWouldBe
PreferableSinceItWouldEnableThemToSeeTheNatureAndExtentOfTheirDangerAndTo

(a)

It@now@became@evident@that@the@city@must@be@abandoned@at@once@There@was@a
difference@of@opinion@in@respect@to@the@hour@of@departure@The@daytime@@it
was@argued@by@some@would@be@preferable@since@it@would@enable@them@to@see

(b)

Figure 8. a. Text sample in which words have been elided following capitalizing each initial letter.
b. Text in which spaces have been filled by "@" (compare Figure 4, sample 5)

been digitised and convolved according to the Marr Hildreth theory at a number of eccentricities in the manner sketched earlier. Figure 9b shows the information available way out at $9°$ (corresponding to about 36 letter spaces), and figure 9c shows the instances which every one of a group of five subjects chose when they were instructed to simulate an unintelligent program to extract "\" from figure 9b. Figure 9d illustrates the information available at $7°$, and shows that the subjects correctly isolated each and every instance of "\". Finally, figure 9e shows the information available at $4°$. It is clear that the early visual system could more easily and reliably find instances of "\" than "@", and so we are led to predict that the Fisher like sample of text shown in figure 9a would be significantly easier to read than the same thing with "\" replaced by "@". Experiment 3 confirms this prediction. Indeed, in Experiment 4, we compared the visually striking filler "\" and normal spacing (sample 1 of figure 4), and we find that the relative advantage of normal spacing fails to reach significance even at the 10% level.

The final experiment 5 is a tribute to the versatility of the computing facilities available for this research. Consider the text sample in figure 10a, in which the forward slash character is used as a delimiter. Since the downstrokes of ascender characters such as "b" and "f" slope slightly forwards but not nearly so much as the slope of "/", we would expect a similar significant advantage for "/"

17

It\now\became\evident\that\the\city\must\be\abandoned\at\once\There\was\a
difference\of\opinion\in\respect\to\the\hour\of\departure\The\daytime\\it
was\argued\by\some\would\be\preferable\since\it\would\enable\them\to\see

(a)

(b)—(d) see over

Figure 9. a. text sample in which "\" is used as a filler between words. b. resulting of convolving the sample in (a) to show the information available at 9°. c. Instances of "\" found in (b) by a group of subjects simulating an unintelligent program. d. Information available in the convolved image at 7° eccentricity. e. Information available to early visual processing at 4°.

over "@". It turns out that this is the case. More interestingly, we were able to design a font in which the only change compared to that of characters in figure 10a is that the forward slash character had *precisely the same slope as the downstroke of an ascender (see figure 10b)*. Figures 10c and 10d show the convolved images of the samples in figures 10a and 10b respectively. The analysis developed above leads us to predict that text samples of the form shown in figure 10a will be significantly easier to read than those in the special font shown in figure 10b, though we might expect that there will be a reduced advantage compared to that shown by "/" or "\" over "@". Experiment 5 confirms this prediction, the significance being only at the 5% level.

4. Experimental details

The experiments were designed strictly in accordance with the method devised by Fisher (1975) to maximize comparability with his results.

Method. Twelve members of the Artificial Intelligence Laboratory who were naive with regard to the purpose of the experiment took part.
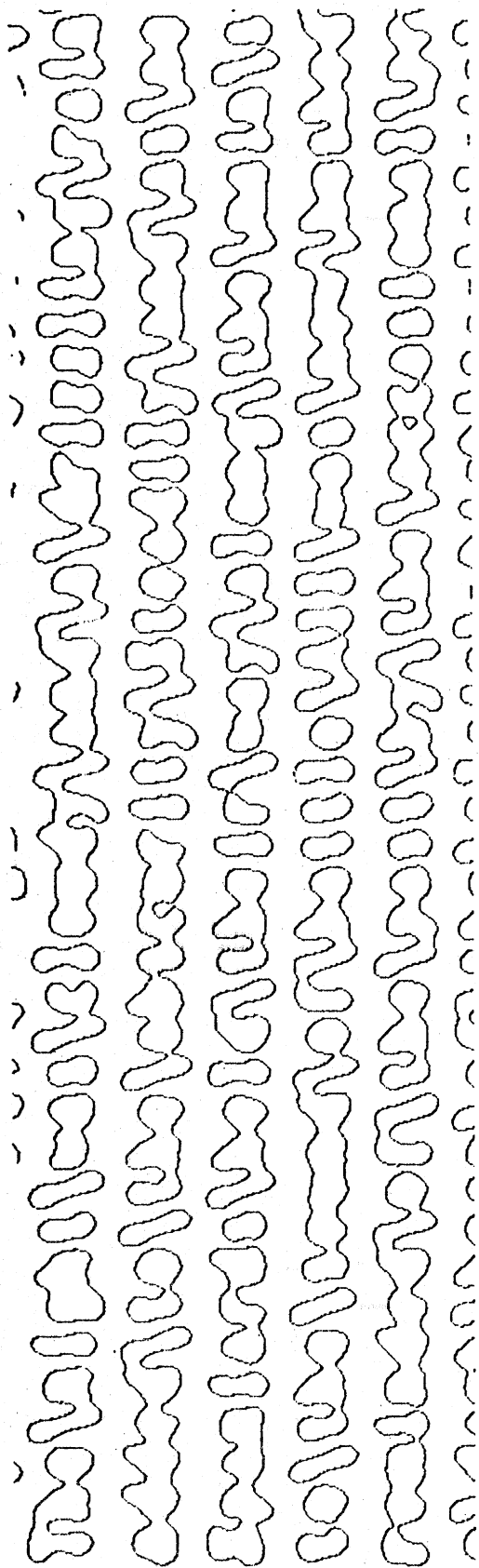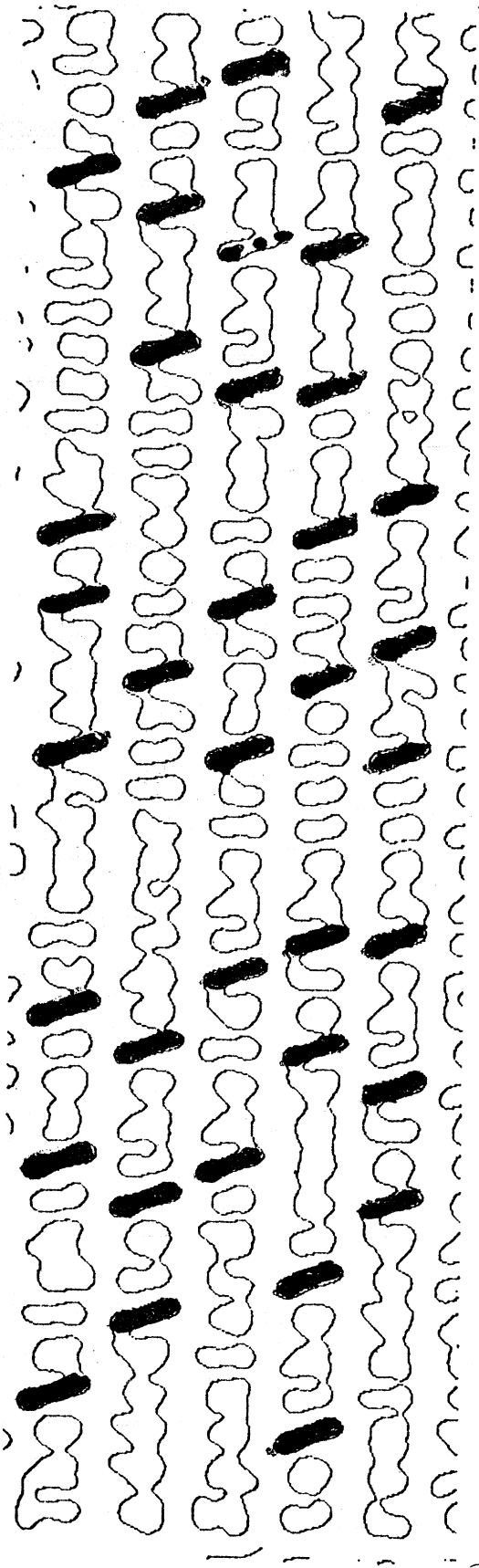
```
(diszc 'sc 35)
#<ART-Q-300.-1600. 2621222.>
;
;
;;   The second paragraph from the Nelson-Denny
;;   test used by Fisher.   Spaces between words
;;   have been replaced by a single \


;;   This is the result of convolving the image
;;   with a mask of central panel width 16.
```

corresponding to 9° eccentricity

= figure 9b =

LISP-LISTENER-G0001

12:05   TYI

```
(diszc 'sc 35)
#<ART-Q-300.-1600. 2621222.>
;
;
;;    The second paragraph from the Nelson-Denny
;;    test used by Fisher.   Spaces between words
;;    have been replaced by a single \


;;    This is the result of convolving the image
;;    with a mask of central panel width 16.


          figure 9c.




LISP-LISTENER-G0001
```

10:40   TYI   ___

Est..run time 872 sec.2 1

512-ZC

The night was cloudy and (a) drizzling ra
added to the obscurity Stoadily and
Spaniards held the fairway along the bull in
to the sound of bullfi All was now hush
resounded of the pass by the occasioooo

Figure 9.d

to package USER

```
;; this is the third paragraph of the
;; Nelson-Denny test convolved at w=8
;; there are inserted slashes.



        figure 9e


LISP-LISTENER-G0001                          ×
```

7:43   TYI   ___

It/now/became/evident/that/the/city/must/be/abandoned/at/once./There/was/a difference/of/opinion/in/respect/to/the/hour/of/departure./The/daytime.//it/was argued/by/some./would/be/preferable./since/it/would/enable/them/to/see/the

(a)

Itlnowlbecamelevidentllthatlthelcityhmustlbelabandonedlatlonce.lTherelwasla differenceloflopinionlini;espectltolthelhourlofldeparture.lTheldaytime.llitlwaslargued bylsome.lwouldlbelpreferable.lsincelitlwouldlenablelthemltolseelthelnaturelandlextent

(b)

(c-d) see over

Figure 10. a. text sample filled with "/". b. text sample in the special font in which the forward slash character has precisely the same slope as the ascender of "d". c. convolved image of (a) at $4°$. d. convolved image of (b) at $4°$.

Materials. The nine paragraphs of the 1960 revised Nelson Denny Reading Test (Denny 1960) were used, together with three paragraphs of similar length (about 200 words) and complexity. The Nelson Denny texts were used by Fisher because they "had a very high degree of standardization from high school through college aged adults" (Fisher 1975, page 189). A Times Roman 10 point font was used throughout the experiments. There were several variations to the basic font:

(i) regular spacing between words ("normal").

(ii) all words elided together, that is, inter-word spacing removed.

(iii) words elided together after the initial letter of each word had been capitalised.

(iv) inter-word spaces filled by "@".

(v) inter-word spaces filled by "\".

(vi) inter-word spaces filled by "/".

(vii) inter-word spaces filled by a special character of the same slope as the descenders in the font.

The experiments (1-5) described in the previous section were designed to compare the relative ease of

figure 10c

```
(diszc 'sc8 80)
#<ART-0-250.-1250. 2392358.>
;; the second paragraph of the Nelson-
;; Denny test with forward slashes
;; implanted in every space. This is the
;; regular space character in the italic
;; font which is times 12ital.
;; Notice that its slope is percepibly
;; greater than that of down slopes of
;; characters such as l,h.

;; In this image it is convolved at w=8.


LISP-LISTENER-G0002                    ✕
```

figure 10d

X

```
;; to test the hyothesis that it is the
;; ease of computing the beginnings of
;; words that is crucial and not the slope of
;; the slash which fills the inter-word
;; spaces, we designed a new character for the
;; font times 12ital by trimming the
;; downstroke of I using R. Sjoberg's font
;; editor.  It has exactly the same slope as
;;  a regular downstroke.

;; the current image is convolved with a mask
;; whose central panel width is 8.


(diszc 'sc8 40)
#<ART-Q-230.-1250. 2361126.>



LISP-LISTENER-G0002
```

reading several pairs of the variations listed above. Specifically, the following hypotheses were tested:

(1) (ii) vs (iii): It was hypothesized that it would be significantly easier to read variation (iii) than variation (ii).

(2) (iii) vs (iv): It was hypothesized that there would be insignificant difference between the ease of reading variations (iii) and (iv).

(3) (iv) vs (v): It was hypothesized that it would be significantly easier to read variation (v) than variation (iv). A similar hypothesis was that variation (vi) would show significant advantage over (iv).

(4) (i) vs (v): It was hypothesized that there would be insignificant difference between the ease of reading variations (i) and (v).

(5) (vi) vs (vii): It was hypothesized that it would be significantly easier to read variation (vi) than variation (vii).

The variations (i) to (vii) were divided into two overlapping sets (i), (ii), (iii), (iv), (v) and (ii), (iii), (iv), (vi), (vii). The subjects were divided into two groups of six and each group was associated with one of the two sets of variations. Each subject had an individually prepared booklet consisting of the twelve paragraphs. The booklets comprised two instances of paragraphs in three of the variations and three instances of two of the variations. The choices of variations and the order of presentation of the variations was counterbalanced over all subjects. "After each paragraph, a set of four multiple choice questions was presented which had to be answered. The questions were taken from the Nelson Denny Reading Test. A digital clock graduated in [steps of 0.1 second] provided a display of the time to read and was clearly visible to all subjects"(Fisher 1975, page 189).

Procedure. Each subject was given a page of instructions containing the variations of text which would appear, the individually prepared booklet of twelve paragraphs, and a question and answer sheet. "When subjects finished reading, they were to look at the time ... they were then to turn the page, answer the questions, and wait for instructions to go on to the next paragraph."(Fisher 1975, page 189).

Results. As there was a substantial spread in the reading speed of the subjects, averaging the data points over all subjects for a particular text produces an unacceptably large standard deviation.

As we are in fact most interested in the **relative** ease of reading two variations, the relevant hypothesis for comparing one text variation α against another β is the null hypothesis:

$$H_0 : \mu[\frac{\alpha}{\beta}] = 1.$$

We can use the simple t statistic defined by

$$t = \frac{r - 1}{s\sqrt{\frac{1}{\nu}}},$$

where there are $\nu + 1$ subjects, r is the mean of the individual values of $\frac{t_\alpha}{t_\beta}$, where t_α is the time taken per word to read the paragraphs in variation α, and s is the standard deviation of that measure from r. The actual results were given in the previous section.

5. Conclusion

This paper began by sketching the background against which this investigation of word isolation in the parafovea has been conducted. Our aim has been to show how published empirical data, especially that of Fisher (1975), could be accounted for using the rich theories of early visual processing of the natural world which have recently been developed in Artificical Intelligence. On the basis of a precise representation of the information available in the parafovea, we proposed an explanation of Fisher's results by postulating a crude, though mostly reliable, model of upper versus lower case characters. The same computational evidence led us to frame a number of predictions, each of which was then confirmed by psychophysical experimentation. As a side effect, we were required to consider how the idea of a character being "visually striking" might be made precise. This approach provides a method for the study of legibility to add to those listed by Spencer(1968, page 21).

As we pointed out in the Introduction, this study is the merely the first step on the long haul towards understanding through computation the exquisite human skill of reading. The results reported here have encouraged us to proceed to consider the next step in the process of acquiring meaning from the sea of gray level intensities which form the image. We consider the next step to be the problem of *integrating information over successive saccades.* Rayner's (1975a, 1975b, 1977, 1978a, 1978b, 1979, Rayner and McConkie 1976, Rayner, McConkie, and Ehrlich, 1978, McConkie and Rayner, 1975) work provides a rich background of empirical data for our study, which is intended

CONCLUSION

to exploit detailed computational models of natural vision in the manner of this paper. It is clear for example that the notion of "word shape" needs to be made more precise by defining an appropriate representation of the information available when a word is convolved at $2°$. Rayner's (1975, page 76) finding that the first and last letters of a word (his NS condition) cause a significant increase in foveation duration is entirely consistent with the approach pursued here. When two nearby lines are convolved, they produce a smeared blob. This occurs not only for strokes **within** a character, but for nearby strokes of two **adjacent** characters (see figure 11). Such inter-character smearing confounds any process whose goal is to elicit structure within a word, and in particular to discover the precise locations of its individual characters. The extremal characters are relatively unaffected by such inter-character smearing, and hence the information gleaned at $4°$ will closely match that computed on a subsequent (foveal) saccade. A similar argument applies to ascenders and descenders, so long as they are relatively isolated. It is not inconceivable that we have learned that such shape information at the extremities of words and from isolated ascenders and descenders within a word are preserved over a typical $2°$ saccade, and have based our word representation scheme, which develops over several such saccades, and the corresponding processes for eliciting substructure, upon it. Further study is needed to make the representation and matching process precise.

For the moment at least, we are left with a reasonably detailed model of eye movement control whose goal is the isolation of words in text on the basis of the information which is available in the parafovea.

1. We can reliably isolate spaces above a size which is yet to be determined, but is about one character space in normal text. We assume that such spaces delimit words, and mostly this inference serves us well. We are confused (and our reading is inhibited) when they do not.

If a space is located on either side of a blob which subtends a visual angle of roughly the same size as an individual saccade, we initiate an eye movement to the beginning of the as yet unprocessed blob. O'Regan's (1979) data gives us some evidence on which to develop the details of this process. In particular, the control may involve a crude representation of the sort discussed earlier for upper case characters, in which case it would presumably be easy to confuse. Again, this requires detailed investigation.

2. If spaces are not available, but words are delimited by some filler character, we dynamically
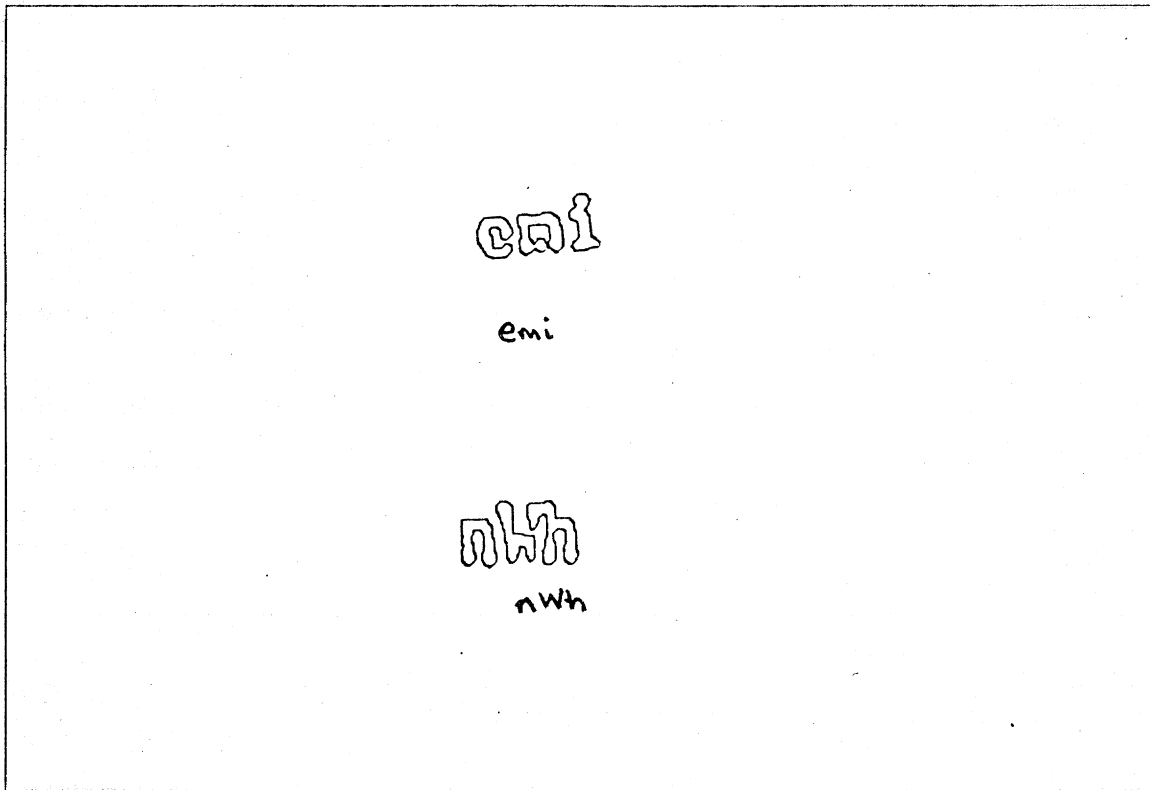
Figure 11. The smearing of nearby lines by convolution is illustrated for strokes within a character (marked "a"), and between two characters ("b").

adjust our scanning strategy to locate instances of that filler. This requires that we first compute a description of the appearance of the filler in the parafovea, and secondly that we search for instances of the description in the convolved parafoveal image. This strategy is reliable to the extent that the filler is "visually striking", that is to say, its instances can reliably be extracted from the available information. The backwards and forwards slash characters are visually striking in this sense, the "@"sign is less so. It is to be expected that the first foveation of text in which spaces are routinely filled in this way would be considerably longer than subsequent ones (there is some evidence that this is generally true, see Levy-Schoen 1979, page 12). It may be conjectured that this can be explained on the basis of the considerations discussed in this paper.

In particular, our model leads to the following prediction. Consider a text sample which consists of a sequence of "segments", each of which can be several words long and is associated with a particular filler character. For example, a segment filled with "\" might be followed by a segment filled with "/" and so on. We would expect that there would be a significant increase in the duration of

23

CONCLUSION

foveations at the boundary between two segments as the parafoveal processing fails to discover an instance of its currently "loaded" filler, and has to locate and load the description of the filler for the next segment.

3. We distinguish between upper and lower case characters on the basis of size and lower curvature only. Capital letters mark important linguistic events in English, such as proper names and the beginnings of sentences. As before, we assume that this importance has been translated into a coarse description which often can be reliably computed in the parafovea. While it often serves us well in isolating upper case characters and drawing our attention to the corresponding linguistic event, it is a coarse description and is easily confused.

Other work, not reported in detail here, shows a slight though not statistically significant advantage over sample seven in figure 4 for a word sequence in which words are alternately printed in a roman font and in italics. This effect is less than that which occurs when bold font is alternated with regular roman. This is consistent with the findings of legibility research. Various researchers, including Tinker(1955), have found that italics actually retard reading, and that readers mostly do not like italics. Tinker(1955) found that 96% of his adult subjects were of the opinion that they could read lower case roman more easily than italics.

This study assumes that the word isolation process is already activated at the time when the text is initially encountered, and it might be thought that high level knowledge would be required to effect this activation. Figure 12c shows a sample of text (figure 12a) convolved with a mask size which corresponds to foveation at a distance of 5.83 metres. The regular texture of lines of blobs is quite clear, even though it is impossible to make any sense of the text. In short, the image *looks like text even at a distance*, as does the image in figure 12g, although in this case it is in fact the convolution of the image shown in figure 12d. Once again, the theory being advanced here is that we **interpret** a particular image as a piece of text on the basis of quite a crude representation, which, however, mostly serves us well.

We conclude with one final remark on the notion that the ease with which a text can be read is directly related to the ease with which information can be reliably computed from its convolved image, and it concerns font design. A great deal of research on font design (see for example Spencer 1968) is depressingly subjective. Recently however, Julesz(1980) and his colleagues have begun a

I scanned the text at roughly the resolution of the human system, viewing the text at a distance of 30 centimeters.
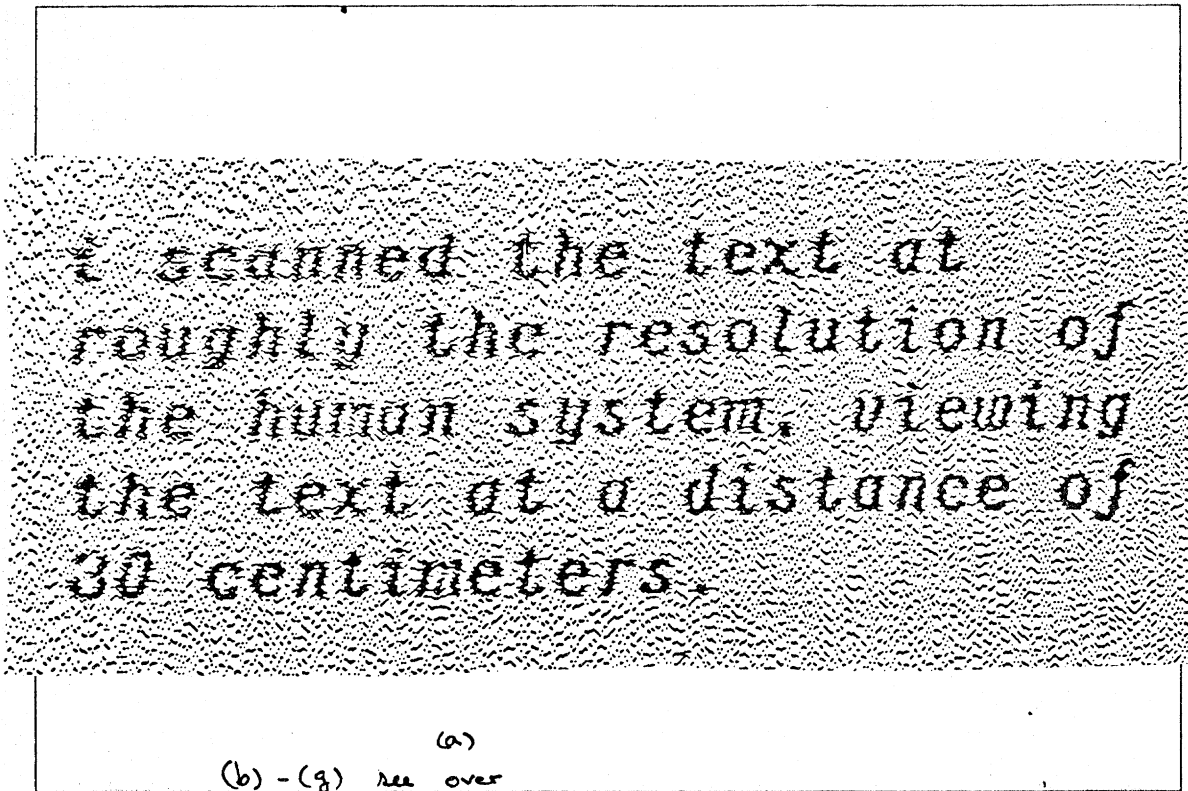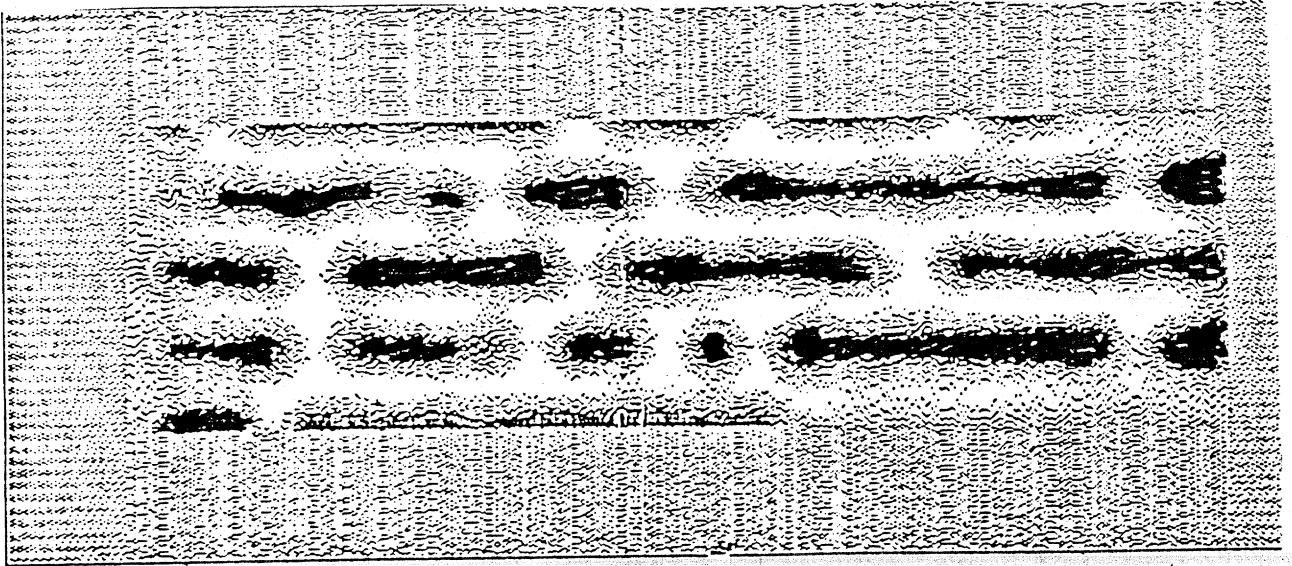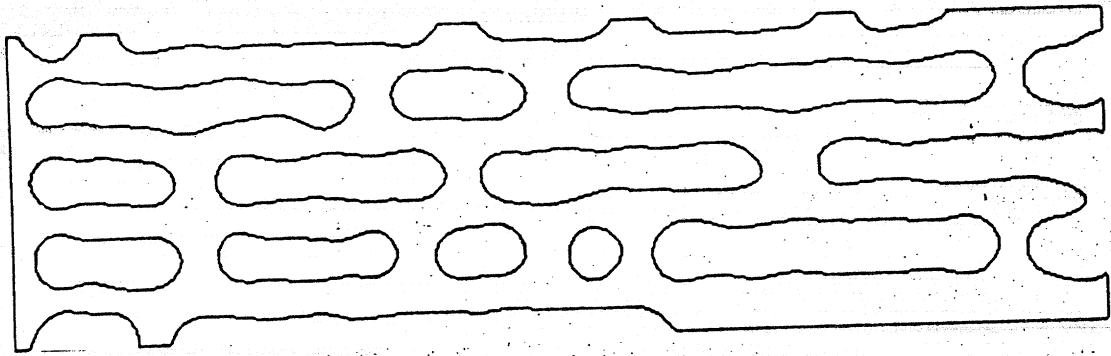
(a)

(b) - (g) see over

Figure 12. a. A sample of text displayed, after photoscanning at a resolution of 100 microns, using a pseudo grey level system devised and constructed by Berthold Horn. b. The result of convolving the text in figure 12a with a mask whose central panel width is 36. This corresponds to foveating the text at a distance of 5.83 metres. c. Zero crossings of the convolution shown in figure 12b. The pattern of blobs corresponding to words is evident. d. A set of random marks produced by filing in the regions which arise from tracing round the text sample given in figure 12a. e. A number of cross sections of the intensity profile shown in figure 12d in the x and y directions. f. the result of convolving the image shown in figure 12e in the same way as figure 12b. g. The zero crossings of the convolution in figure 12f. The result is quite similar to figure 12c.

study which is analogous to that pursued here. They apply their ideas about texture discrimination to define a set of so-called "textons" and then advocate the design of fonts based on the discriminability of textons. Our approach also relates the legibility of a font to the processes of natural perception, but we are currently more concerned with understanding the perceptual basis of the efficacy of using serifs and so forth than with the aesthetics of font design. There is nevertheless a good deal of similarity between our goals. Much more work is necessary to develop the ideas sketched in this section into a coherent and precise theory.
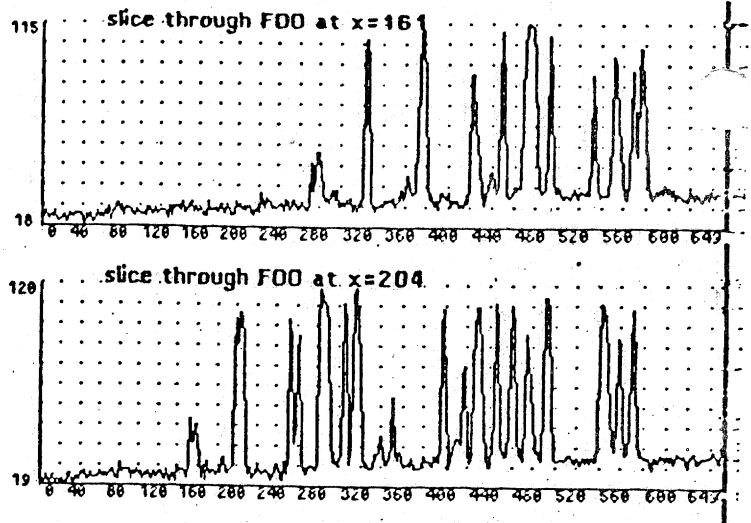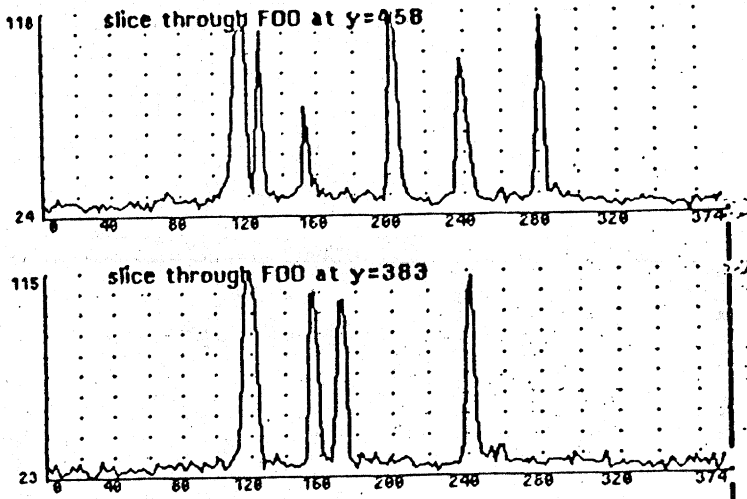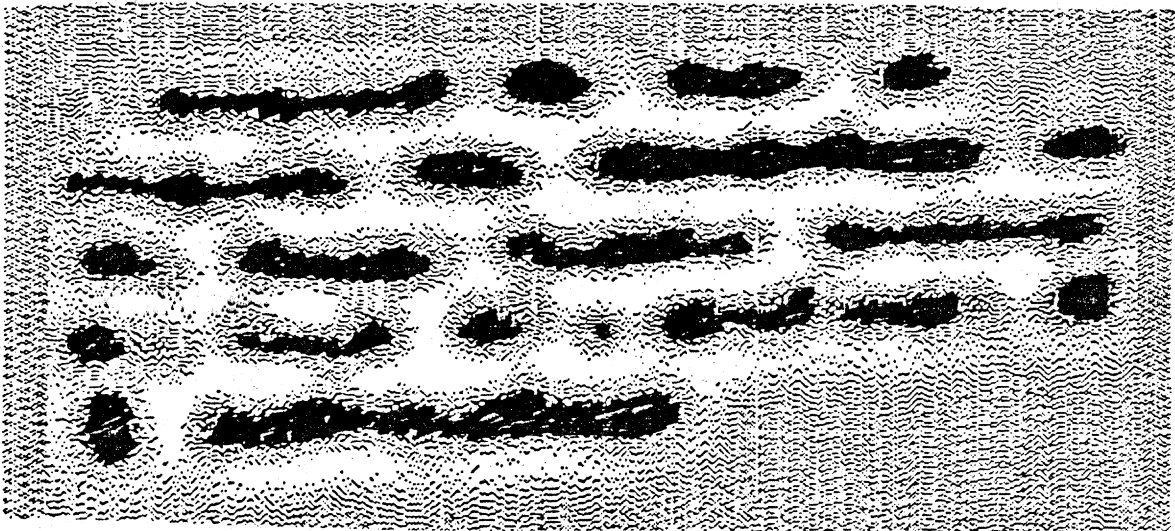
(b)

(c)

(d)

slice through FOO at y=458

slice through FOO at x=161

slice through FOO at y=383

slice through FOO at x=204

(e)

(f)

(g)

6. Acknowledgements

7. References

Allport A. Word recognition in reading, *Processing of Visible Language* eds. Kolers P. A.,Wrolstad M. E., and Bouma H., Plenum Press, 1979.

Barrow H. G. and Tenenbaum J. M. Experiments in Interpretation Guided Semantics, Technical Note, 123, SRI International, 1976.

Barrow H. G. and Tenenbaum J. M. Recovering intrinsic scene characteristics from images, *Computer Vision Systems* eds. Hanson and Riseman, Academic Press, 1978.

Bouma H. "Visual recognition of isolated lower case letters," *Vision Research* 11 (1971), 459-474.

Brady J. M. "The development of a computer vision system," *Recherche Psicologica* (1979), .

Brady J. Mike "Special issue on computer vision (ed.)," *Artificial Intelligence* 11 (1981), .

Brady J. M. and Wielinga B. J. Reading the writing on the wall, *Computer Vision Systems* eds. Hanson and Riseman, Academic Press, 1978.

Cohen Gillian Speech perception and reading, *Cognitive Psychology (part 2)*, Open University Press, 1978.

Davis Larry S. and Henderson T. C. Hierarchical constraint processes for shape analysis, Computer Sciences Dept., University of Texas, Austin, TR-115, 1979.

Davis Larry S. and Rosenfeld Azriel Hierarchical relaxation for waveform parsing, *Computer Vision Systems*, eds. Hanson and Riseman, Academic, 1978.

Davis Larry S. and Rosenfeld Azriel "Cooperating processes for low-level vision: a survey," *Artificial Intelligence: Special issue on Computer Vision* 11 (1981), .

Duda R. and Hart P. *Pattern Classification and Scene Analysis*, John Wiley, New York, 1973.

Estes W. K. Interaction of perception and memory in reading, *Basic processes in reading: perception and comprehension,* Laberge and Samuels (eds.), John Wiley, 1977.

Fisher D. F. "Reading and Visual Search," *Memory and Cognition* 3 (1975), 188-196.

Frisby J. P. *Seeing*, Oxford University Press, Oxford U.K., 1979.

Gough P. B. One second of reading, *Language by Ear and Eye,* eds. Kavanagh J. F. and Mattingly I. G., MIT Press, 1972.

Grimson W. E. L. Computing shape using a theory of human stereo vision (PhD thesis), forthcoming book published by MIT press, MIT, 1980.

Hildreth E. C. Implementation of a theory of edge detection (MS dissertation), also AI-TR 579, MIT, 1980.

Henderson L. Word Recognition, *Tutorial essays in Experimental Psychology,* ed. N. S. Sutherland, vol. 1, Erlbaum Press, 1977.

Hochberg J. Components of literacy: speculation and exploratory research, *Basic studies in reading*

REFERENCES

eds. Levin H. and Williams J. P., Basic Books, 1970.

Horn B. K. P. "Determining lightness from an image," *Computer Graphics and Image Processing* 3 (1974), 277-299.

Horn B. K. P. "Understanding image intensities," *Artificial Intelligence* 8 (1977), 201-231.

Huey E. B. *The Psychology and Pedagogy of Reading* , Macmillan, New York, 1908.

Julesz B. "Spatial nonlinearities in the instantaneous perception of textures with identical power spectra," *Phil. Trans. R. Soc. Lond. B* 290 (1980), 83-94.

Lesser Victor R. and Erman Lee D. "A retrospective view of the Hearsay-II architecture," *Proc. Int. Jt. Conf. Artificial Intelligence* 2 (1977), 790-800.

Levy-Schoen A. and O'Regan K. The control of eye movements in reading, *Processing of Visible Language,* eds. Kolers P. A.,Wrolstad M. E., and Bouma H., Plenum Press, 1979.

Luenberger D. G. *Introduction to linear and nonlinear programming* , Addison-Wesley, New York, 1973.

Marcel T. and Patterson Karalyn Word recognition and production: reciprocity in clinical and normal studies, *Attention and Performance* ed. Requin J., vol. 7, Erlbaum, 1979.

Marr D. "Early processing of visual information," *Phil. Trans. R. Soc. Lond. B* 275 (1976), 483-524.

Marr D. *Vision* , Freeman, San Francisco, 1981.

Marr D. and Hildreth E. "Theory of edge detection," *Proc. R. Soc. Lond. B* 207 (1980), 187-217.

Marr D., Hildreth E., and Poggio T. Evidence for a fifth, smaller channel in early human vision, MIT, AI Memo 541, 1979.

Marr D., Palm G., and Poggio T. "Analysis of a cooperative stereo algorithm," *Biol. Cybernetics* 28 (1978), 223-229.

Marr D. and Nishihara H. K. "Representation and recognition of the spatial organisation of three dimensional structure," *Proc. R. Soc. London B* 200 (1978), 269-294.

Marr D. and Poggio T. "Cooperative computation of stereo disparity ," *Science* **194** (1976a), 283-287.

Marr D. and Poggio T. Cooperative computation of stereo disparity, MIT, AI Memo 364, 1976b.

Marr D. and Poggio T. "A theory of human stereo vision," *Proc. R. Soc. Lond. B* **204** (1979), 301-328.

Marr D. and Ullman S. "Directional selectivity and its use in early visual processing," *Proc. R. Soc. Lond. B* (1981), .

McClelland James L. and Rumelhart David E. "An interactive activation model of the effect of context in perception, part 1", center for human information processing, University of California, San Diego, 8002, 1980.

McConkie G. W. On the role and control of eye movements in reading, *Processing of Visible Language* eds. Kolers P. A.,Wrolstad M. E., and Bouma H., Plenum Press, 1979.

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

Minsky M. and Papert S. Artificial Intelligence Progress Report, MIT, AI Memo 252, 1972.

Nash-Webber B. L. The role of semantics in automatic speech understanding, *Representing and Understanding: studies in cognitive science* eds. Bobrow D. and Collins A., Academic Press, 1975.

Nelson *The Nelson Denny Reading test* , Houghton-Mifflin Company, Boston Mass., 1960.

O'Regan Kevin Moment to moment control of eye saccades as a function of textual parameters in reading , *Processing of Visible Language* eds. Kolers P. A.,Wrolstad M. E., and Bouma H., Plenum Press, 1979.

Rayner K. "The perceptual span and peripheral cues in reading," *Cognitive Psychology* **7** (1975a), 65-81.

Rayner K. "Parafoveal identification during a fixation in reading," *Acta Psychologica* **39** (1975b), 271-282.

Rayner K. "Visual attention in reading: eye movements reflect cognitive processes," *Memory and*

REFERENCES

Cognition 5 (1977), 443-448.

Rayner K. "Eye movements in reading and information processing," *Psychological Bulletin* **85** (1978a), 618-660.

Rayner K. Foveal and parafoveal cues in reading, *Attention and Performance VII* ed. Requin J., Erlbaum, 1978b.

Rayner K. Eye movements in reading: eye guidance and integration, *Processing of Visible Language* eds. Kolers P. A., Wrolstad M. E., and Bouma H., Plenum Press, 1979.

Rayner K. and McConkie G. W. "What guides a reader's eye movements?," *Vision Research* **16** (1976), 829-837.

Rayner K., McConkie G. W., and Ehrlich S. "Eye movements and integrating information across fixations," *Jour. Exp. Psychology: Human Perception and Performance* 4 (1978), 529-544.

Richter J. and Ullman S. A model for the spatio-temporal organization of X and Y-type Ganglion cells in the primate retina, MIT, AI Memo 573, 1980.

Rosenfeld Azriel, Hummel R. A., and Zucker S. W. "Scene labelling by relaxation operations," *IEEE Trans. Systems, Man, Cybernetics* SMC-6 (1976), 420-433.

Rumelhart David E. Toward an interactive model of reading, *Attention and Performance VI* ed. Dornic S., Erlbaum Press, 1977.

Schank R., Goldman N., Rieger C., and Riesbeck C. "MARGIE: Memory Analysis Response Generation and Inference on English," *Proc. Int. Jt. Conf. Artificial Intelligence* 3 (1973), 255-261.

Smith F. "Familiarity of configuration vs. discriminability of features in the visual identification of words," *Psychonomic Science* 14 (1969), 261-262.

Spencer *The Visible Word*, The Royal Academy, London, 1968.

Stevens K. A. On the visual detection of fine detail, MIT, in preparation, 1980.

Ullman S. *The interpretation of visual motion*, MIT Press, Cambridge, Mass., 1978.

Ullman S. "Relaxation and constrained optimisation by local processes," *Computer Graphics and Image Processing* **10** (1979), 115-125.

Waltz D. L. A parallel model for low level vision, *Computer Vision Systems* eds. Hanson and Riseman, Academic Press, 1978.

Wilson H. R. and Giese S. C. "Threshold visibility of frequency gradient patterns," *Vision Research* **17** (1977), 1177-1190.

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

Zucker S. W. Vertical and horizontal processes in low level vision, *Computer Vision Systems* eds. Hanson and Riseman, Academic Press, 1978.

Zucker S. W., Leclerc Y. G., and Mohammed J. L. "Continuous relaxation and local maxima selection -- conditions for equivalence," *Proc. Int. Jt. Conf. Artificial Intelligence* **6** (1979), 1014-1016.