The Visual Processing of Text

Stephen J Emmott

Centre for Cognitive & Computational Neuroscience
Department of Psychology, University of Stirling.
Stirling, Scotland, UK.
FK9 4LA

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

1993

Contents

Acknowledgements

I am truly indebted and grateful to my supervisor, Roger Watt, for all his support throughout, and for giving me the encouragement and the freedom to learn.

I am immensely grateful for the good fortune I have had in spending the last thirty months in the company of Will Goodall-Lelong, Steven Dakin, and Patricia Carlin especially, but also everyone else in the CCCN. Their freindship kept me going in rough times, and kept me laughing in better times.

Steven Dakin and Peter Hancock in particular, but also Ian Patterson and Roland Baddeley kept me from computationally drowning. I thank them for their assistance and patience.

Rachel O' Sullivan, Trish Carlin, and Steven Dakin spent very long periods enduring unreasonable tasks as psychophysical subjects. I thank them for their patience.

This research was funded by the Science and Engineering Research Council and BT. I thank David Travis for his support, and for shielding me from the latter organisation's administrative abyss.

I finally thank Roger, Ben Craven, and David Travis for the unenviable task of commenting on my thesis. I can only hope that the effect it must have had on them is not permanent.

Abstract

The results of an investigation into the nature of the visual information obtained from pages of text and used in the visual processing of text during reading are reported.

An initial investigation into the visual processing of text by applying a computational model of early vision (MIRAGE: Watt & Morgan, 1985; Watt, 1988) to pages of text (Computational Analysis 1) is shown to extract a range of features from a text image in the representation it delivers, which are organised across a range of spatial scales similar to those spanning human vision. The features the model extracts are capable of supporting a structured set of text processing tasks of the type required in reading. From the findings of this analysis, a series of psychophysical and computational studies are reported which examine whether the type of information used in the human visual processing of text can be described by this modelled representation of information in text images.

Using a novel technique to measure the 'visibility' of the information in text images, a second stage of investigation (Experiments 1–3) shows that information used to perform different text processing tasks of the type performed in reading is contained at different spatial scales of visual analysis. A second computational analysis of the information in text demonstrates how the spatial scale dependency of these text processing tasks can be accounted for by the model of early vision.

In a third stage, two further experiments (Experiments 4–5) show how the pattern of text processing performance is determined by typographical parameters, and a third computational analysis of text demonstrates how changes in the pattern of text processing performance can be modelled by changes in the pattern of information represented by the model of vision.

A fourth stage (Experiments 6–7 and Computational Analysis 4) examines the time-course of the visual processing of text. The experiments show how the duration required to reach a level of visual text processing performance varies as a function of typographical parameters, and comparison of these data with the model shows that this is consistent with a time-course of visual analysis based on a coarse-to-fine spatial scale of visual processing.

A final experiment (Experiment 8) examines how reading performance varies with typographical parameters. It is shown how the pattern of reading performance and the pattern of visual text processing performance are related, and how the model of early vision might describe the visual processing of text in reading.

The implications of these findings for theories of reading and theories of vision are finally discussed.

1

Introduction

This thesis is concerned with understanding the nature of the early visual representation of text, the information delivered by such a representation, and how this visual information might be used in the processing of text.

Reading, like many other visual tasks, seems effortless. Yet despite this deceptive ease, the way in which early visual processing extracts the information in a page during reading, or even what information it extracts, remains poorly understood (see Henderson, 1987 for a review). This seems surprising since reading–and in particular the role of vision in reading–has been an important, and intensively studied, topic of research within psychology for over a century.

It is argued in this thesis that the failure of research to develop a proper understanding of the visual processing of text is due at least in part to an inadequate account of what vision actually does and the assumptions which have consequently had to be made about what the role of vision in reading might be. Furthermore, it is suggested, and indeed the aim is to show, that early vision may deliver a representation of text which is far different from that assumed by any existing account of the role of vision in reading.

1.1 Overview of Chapter 1

The work of this chapter begins with a discussion of the reading process itself, and in particular the text processing tasks performed as part of this process. Emphasis is placed on what is expected from the *visual* processing of text in reading. There then follows a selective review of the type of research conducted into the role of vision in reading (Section 1.3) in order to consider what is understood about the visual processing of text. The section concludes with the suggestion that, for the most part, the lack of understanding of the visual processing of text in reading has arisen because both research and theory in reading is often driven by inadequate assumptions made about what vision does, and indeed what the

reading process itself involves. It is argued that little progress in this area can be made without a more thorough consideration of both early visual processing (by applying a model of visual processing to text) and the visual tasks involved in reading in a new approach to an examination of the visual processing of text in reading.

Sections 1.4 and 1.5 therefore consider the operations which might be performed by early visual processing. This discussion extends to how the information needed to process text might be extracted from a page and represented by the operations of early visual mechanisms in the brain. The utility of a computational approach to vision is outlined.

Section 1.6 examines the relationship between typography and vision, and discusses how typography must have evolved around how vision operates. In this respect, the value of using typographical parameters for studying the relationship between vision and text is considered.

The final Sections (1.7 and 1.8) summarise the preceding discussion of this introductory chapter. It concludes with a proposal for a way forward in understanding the visual processing of text in reading, and the methodological considerations involved.

1.2 Reading

Reading a page of text is certain to require a number of visual processing tasks to be performed to allow the reader to extract the information contained in the marks on the page. Presented with a page of text, the reader first needs to know something about the layout of the page. In particular, visual information is needed which tells the reader which way the text is oriented and what the scale or magnification of the text is, before reading can begin. Because readers know that text is printed as rows of lines, and each line tells the reader approximately what size the letters are, all this information can be determined by extracting information about the orientation, size, and number of lines.

Having obtained this information, the reader may need, or wish, to find the start of a sentence. One, perhaps obvious, strategy for doing this might be to look for capital letters, but this would also locate all parts of the text which have uppercase letters, which would be expected to make finding sentence boundaries an unreliable and therefore slow and error prone process, and it is known that readers can locate sentence boundaries quickly and reliably (Emmott & Watt, 1992). It is interesting to note in this respect, that typographical practice includes an additional space at sentence boundaries, and locating this additional space might be a more efficient strategy. Indeed, this inclusion of an additional space may well serve as some kind of visual marker to enable this task to be performed. If so, it would be interesting to discover how vision extracts and represents this marker.

Before any word can be read, it is necessary first to know where the word begins and ends (each word needs to be 'segmented'). This provides information about the length of the word so that if it needs to be fixated, the eyes know where to move to in order to fixate on the informative part of it. Evidence from eye movement studies suggests that the information necessary to represent this information is made available parafoveally and/or in the visual periphery; that is, at coarse resolutions (*e.g.*, McConkie & Rayner, 1975; Rayner & McConkie, 1976; Rayner & Betera, 1979). Indeed, evidence also suggests that word length information may serve as an important cue in reading as a guide to where to fixate (*e.g.*, Underwood *et al.*, 1988; Vitu, 1991) and in appropriate visual and narrative context, information about word length and other aspects of word shape may be sufficient to identify that word (Monk & Hulme, 1983; Haber & Haber, 1981).

Once the eyes have moved to fixate a word (to align the word foveally), the fine scale resolution at the fovea could then be used to extract visual information about the letters and their order in the word, which would uniquely specify the orthography of that word.

This outline of the reading process is speculative, but plausible. It is based on a number of assumptions about the ability of visual processing to deliver the type of information required by such a process. A number of hypotheses generated from these assumptions are tested in this thesis. However, reading undoubtedly requires a number of text processing tasks of the sort outlined here to be performed in order to extract the information contained in the marks on the page.

This view of reading raises a number of issues. It defines reading as the product of several text processing tasks, and not a simple or single event in which the role of vision is confined to delivering a representation of letter features or letters to provide an orthographic representation of any single word from a whole page of text. Furthermore, it suggests that the level of detail at which the information required to perform particular text processing tasks might be found at different levels of visual processing. This suggests that if the visual information required for different text processing tasks is specific to that task, a proper study of the visual processing of text requires an examination of specific text processing tasks rather than the product of these tasks–reading. This description has important consequences for any study attempting to discover something about the visual processing of text. In particular, it provides a description of the reading process which is far more structured than that often considered in reading research, and suggests that it is a process upon which vision would be expected to have an important and fundamental impact.

1.3 Vision and text: a review

1.3.1 Introduction

Cognitive psychology's history has been concerned to a large extent with reading, and much of the research has been concerned with the role of vision in this process. In particular, the research and debate has been focused almost exclusively on the role of vision in providing an orthographic representation (the list of letters and their order) of a word. Within this debate, research has provided three possibilities. The first possibility is that visual processing provides a description of whole word information, such as word length and the positions of ascending, descending, and small letters, which may be sufficient to specify the orthography of the word (Cattell, 1886; Haber & Haber, 1981; Monk & Hulme, 1983; Rayner & Pollatsek, 1989).

The second possibility is that visual processing delivers a representation only of letter features, on which some form of 'higher–level' processing or application of 'top-down' knowledge is required to enable stored letter-level and word-level representations to be 'contacted' (*e.g.*, the interactive-activation model: McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; PABLO: McClelland, 1986; McClelland, 1987).

The third possibility, and a more recently emerging issue, is that an "attentional" mechanism (*i.e.* a cognitive–as distinct from a visual–process) is "applied" or "directed" to either whole words (*e.g.*, Duncan, 1987; Bock, Monk & Hulme, 1993), letter features and letters (Treisman & Souther, 1986), or to both (McConkie & Zola, 1987) in order that vision is then able to provide a representation of the orthography of a word which will identify it.

It is argued in this and subsequent sections of the thesis that this debate has been influenced strongly by past and current assumptions about what information visual processing is capable of (or limited to) delivering in reading. It is a specific aim of the following selective review of research into reading to make a case for this argument. A further aim is to establish what has actually been learned from the research to date about the visual processing of text. The review also serves the purpose of the final aim which is to illustrate the need for a new approach to the study of the visual processing of text in reading.

1.3.2 Single word studies

The debate concerning whether the early visual representation of any word is only of its letter features or also of its whole shape owes its existence to the experimental technique which has helped to fuel it: the presentation of letters, single words or nonsense letter strings

1.3 Vision and text: a review

1.3.1 Introduction

Cognitive psychology's history has been concerned to a large extent with reading, and much of the research has been concerned with the role of vision in this process. In particular, the research and debate has been focused almost exclusively on the role of vision in providing an orthographic representation (the list of letters and their order) of a word. Within this debate, research has provided three possibilities. The first possibility is that visual processing provides a description of whole word information, such as word length and the positions of ascending, descending, and small letters, which may be sufficient to specify the orthography of the word (Cattell, 1886; Haber & Haber, 1981; Monk & Hulme, 1983; Rayner & Pollatsek, 1989).

The second possibility is that visual processing delivers a representation only of letter features, on which some form of 'higher–level' processing or application of 'top-down' knowledge is required to enable stored letter-level and word-level representations to be 'contacted' (*e.g.*, the interactive-activation model: McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; PABLO: McClelland, 1986; McClelland, 1987).

The third possibility, and a more recently emerging issue, is that an "attentional" mechanism (*i.e.* a cognitive–as distinct from a visual–process) is "applied" or "directed" to either whole words (*e.g.*, Duncan, 1987; Bock, Monk & Hulme, 1993), letter features and letters (Treisman & Souther, 1986), or to both (McConkie & Zola, 1987) in order that vision is then able to provide a representation of the orthography of a word which will identify it.

It is argued in this and subsequent sections of the thesis that this debate has been influenced strongly by past and current assumptions about what information visual processing is capable of (or limited to) delivering in reading. It is a specific aim of the following selective review of research into reading to make a case for this argument. A further aim is to establish what has actually been learned from the research to date about the visual processing of text. The review also serves the purpose of the final aim which is to illustrate the need for a new approach to the study of the visual processing of text in reading.

1.3.2 Single word studies

The debate concerning whether the early visual representation of any word is only of its letter features or also of its whole shape owes its existence to the experimental technique which has helped to fuel it: the presentation of letters, single words or nonsense letter strings

using a tachistoscopic (extremely brief: as short as 10msec) display (*e.g.*, Cattell, 1886; Reicher, 1969; Wheeler, 1970; Paap, Newsome & Noel, 1984). The general finding from such experiments is that the probability of a target letter being detected is greater if it appears in words than in non-words or in isolation. Two interpretations are typically offered for this finding, which has been termed the 'word superiority effect' (WSE). The first is that visual processing makes available whole word information, that is, information about word shape (by which is taken to mean the pattern of ascenders, descenders and small letters making up the word envelope). These 'supraletter' features might then used to facilitate the recognition of letters in the word which allow the word itself to be recognised (*e.g.*, Cattell, 1886).

The second interpretation is based on the finding that the type of post- stimulus mask influences this WSE (*e.g.*, Johnson & McClelland, 1973). Specifically, a post-stimulus luminance mask tends to abolish the WSE effect, but a pattern mask does not. The prevailing interpretation of this finding is that words 'activate' a stored (lexical) representation of their identity which, because it is not a visual representation, is less susceptible to disruption by pattern-specific mask (which looks like 'chopped-up' letters) than nonsense words. Nonsense words will not have a lexical representation, but will have a visual representation, and will therefore be severely disrupted by a pattern mask. The conclusion reached from this, essentially as a corollary, and one which Johnson & McClelland (1980), McClelland & Rumelhart (1981) and Rumelhart & McClelland (1982) went on to develop, was that the visual processing of words is confined to a representation of letter features.

From this latter interpretation, McClelland and colleagues proposed a model of word recognition in reading, which was termed the "interactive-activation" model (Johnson & McClelland, 1980; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). It serves little purpose for the present argument to discuss the details of the model, except to say something about two key pertinent features of it, which are illustrated the architecture of the basic model, shown in Figure 1.1.

The first is that the role of vision in this model is confined to delivering a representation only of letter features and letters. The second is that visual processing must always provide a representation of letter features before letters[1].

The crucial point to note here is not the WSE itself, but the view of visual processing in reading which it has provoked. The popularity and influence of the interactive-activation

---

1 Although, in fairness to the original model, it only ever dealt with the recognition of single words, extensions to the model (see McClelland, 1986; McClelland, 1987) to encompass reading, rather than single word recognition, still include these key components.

model (despite its inability to account for a number of findings; see *e.g.*, Monk, 1985) has further established the role of vision in reading as one of letter recognition.

What is significant for the present purposes about most of this research is that it has been devoted almost exclusively to examining the same single issue: whether visual processing in reading is confined to representing letter features only, or extends to representing whole word shape, which is used for word recognition. It is also significant how the almost singular devotion to this issue has influenced the methodologies used to examine visual processing in reading, the findings necessarily from such methods and their subsequent interpretation. A series of experiments which illustrates neatly what is meant by this, which uses common and popular methodologies, and which has been influential in current debate is that of Paap, Newsome & Noel (1984). A critical summary of two of the experiments reported by Paap and colleagues is given here to make this point.

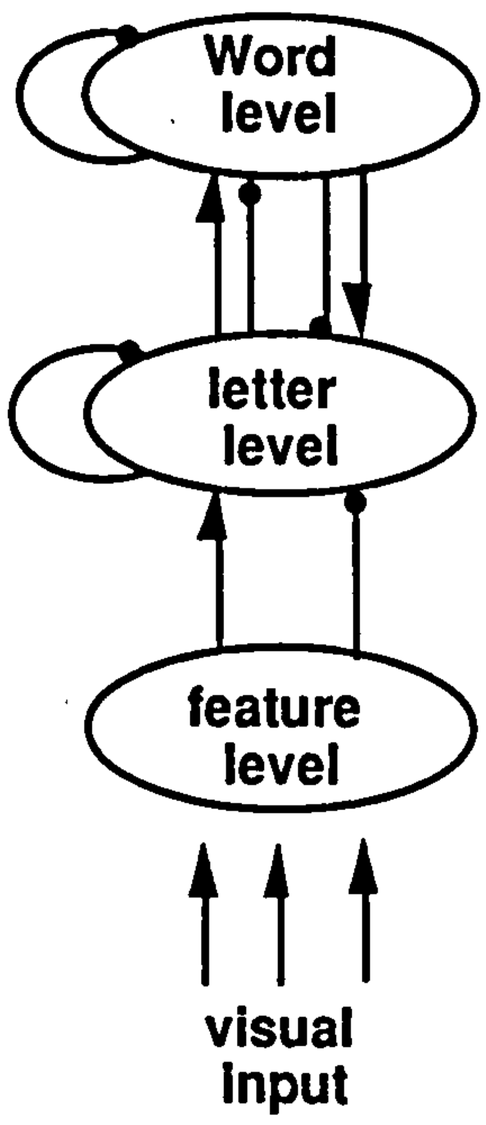


Figure 1.1. The interactive activation model of word recognition. Note how visual processing is only ever able to deliver a representation of letter features. 'Higher-level' stored representation of words can interact in an excitatory (arrows) or inhibitory manner (circles) to determine letter representations. Adapted from McClelland & Rumelhart, 1981.

Paap, Newsome and Noel (1984) began by suggesting that if word shape is represented visually and used to recognise a word, words differing in word shape frequency should be processed differently. Specifically, words with rare shapes should be processed more

efficiently (*i.e.* faster, according to Paap, Newsome and Noel) than words with common shapes. A probed-letter task (which is the same tachistoscopic technique to that described earlier) found no evidence for an effect of word shape on target decision accuracy. The conclusion that Paap, Newsome & Noel reached from this was essentially that visual processing does not provide a representation of whole word shape which is used for the recognition of words. However, Paap, Newsome and Noel's (1984) claim that words with a rare shape should 'speed-up' processing is questionable. For example, there is evidence that words having a common shape, particularly the shape cue of length, may, when combined with contextual information, be processed visually very quickly, often even without fixation (Fisher & Shebilske, 1985; Haber & Haber, 1981).

In any case, this first finding was contrary to a second finding from a lexical-decision task. This is an experiment in which a tachistoscopic technique is still used but in which the subject now has to decide whether what was presented was a word or a non-word. The lexical-decision task *did* find evidence for an effect of word shape. Words with rare shapes were found to have shorter lexical decision times than words having common shapes. But, because this effect was found to occur also for uppercase words, this was given as further evidence that early visual processing must deliver only a representation of letter features.

This latter interpretation and conclusion is important. The conclusions of the study rely on the ease with which the contrary findings of the lexical-decision task and the probed-letter task were 'reconciled' by assuming that uppercase words do not have any distinctive shape. This claim makes a telling observation about the sort of assumptions which are made about what information vision is capable of delivering. It is an interpretation which is based on unfounded assumptions about the nature of the early visual representation of text.

The Paap, Newsome & Noel (1984) studies have been outlined because they are typical of the type of approaches commonly adopted in studying the role of vision in reading. Indeed, they are approaches which, by their nature, at least makes the tacit assumption that the visual context in which a word appears is unlikely to have any effect on the visual processing it receives. Certainly, it is an approach which is unable to examine whether the visual context in which words appear has important consequences for the visual processing they receive.

1.3.3 Proofreading

Studies which involve reading pages of text have been claimed as being better able to identify the information normally used in the visual processing of text in reading than single word studies. Proofreading tasks fall into this category. The rationale behind such studies is that the reading task requires the subject to perform visual processing tasks normally

involved in reading. Typical of this technique is a study by Monk & Hulme (1983). In this study, readers were required to proofread a page of text and detect misspellings. The misspellings to be detected either did or did not alter word shape. The argument went that if a visual representation of word shape is used in word recognition, then misspellings which did not alter shape should be harder to detect, and indeed, this was found to be the case. However, using a very similar technique, a third in the series of studies by Paap, Newsome & Noel (1984), of which the other two were described earlier, failed to find this effect.

Contrary findings such as these suggest that proofreading may be a technique which is unable to disambiguate the ambiguous evidence concerning visual processing in reading. At least partly the reason for this may be because proofreading is a specific, and in many respects atypical, reading situation in which the information required to perform the task is likely to be found at a different, finer, level of detail from that required for other, more commonly occurring, reading situations.

1.3.4 Eye movement studies

Like proofreading, eye movement studies have also been proposed as a natural method of establishing something about what visual processing of text occurs in reading–a claim based on the assumption that they allow 'normal' reading while eye movements are monitored.

Two methodologies have been used. The first is where eye movements are simply recorded, the rationale being that a recording of where the eyes do and don't fixate provides information from which it can be inferred what visual processing might be occurring. Using this technique, Rayner & McConkie (1976) found that the probability of fixating any word is determined by its length: the longer a word is, the more likely it will be fixated. This suggests that the visual system provides a representation of whole word shape, particularly word length, which seems to be used for some aspects of text processing in reading. Also using this technique, Vitu (1991) has shown that readers tend to make a saccade to the centroid or "centre-of-gravity" of words, except very long words. Eye movement data such as this is therefore consistent with the view of text processing requirements during reading expounded in Section 1.2: that visual processing at coarse spatial scales of resolution might provide a representation of word length at the point of word segmentation. This is an important point, and one which is returned to in subsequent chapters.

In the second type of eye movement study the text displayed is contingent on fixation position. This method allows changes to be made to the text at various parts of the visual field which are contingent upon eye movements, and the ability to detect such changes are then measured. The rationale here is that any changes which cannot be detected do not

receive visual processing. Using this sort of technique, McConkie & Zola (1979) found that changing the case of case alternated text (WhIcH lOoKs LiKe ThIs) in the saccade between peripheral exposure and fixation produced no significant change in naming latency. Since both letter features and word shape are changed in this procedure, these data present a confusing picture, given that most of the research conducted argues that early visual processing in reading provides either a representation of letter features and/or word shape for the purposes of word recognition.

A key claim for the usefulness of eye movement recording techniques: that they allow an examination of the visual processing of text under 'natural' conditions needs examination. In many studies, only a single line of text is presented (*e.g.*, McConkie & Zola, 1979), in others (*e.g.*, Rayner, McConkie & Ehrlich, 1978) only a single target word is presented. The text processing requirements during reading a single target word, or from an isolated line, may well be different from those when reading from a whole page of text, according to the description or the reading process outlined in Section 1.2. Therefore, the visual processing performed may also be different. In other words, the *visual context* in which words occur may have an important effect on the visual processing they receive. It is perhaps not surprising to learn then, that there is considerable debate about how best to interpret eye movement data (see Rayner & Pollatsek, 1987 for a discussion of problems interpreting eye movement data).

1.3.5 The visual processing of text, visual attention and reading

The conflicting and confusing data concerning the nature of the visual information used in reading and the consequent inability of the research to provide an adequate account of the visual processing of text has led to the conclusion by an increasing number of psychologists that early vision is unable to provide an appropriate or suitable representation of the information in text. A separate, cognitive, mechanism: "attention" as an additional step in the reading process has been proposed as a possible means of providing a satisfactory resolution and explanation of the inability to reconcile the conflicting data (*e.g.*, Treisman & Souther, 1986; Duncan, 1987; McConkie & Zola, 1987; Bock, Monk & Hulme, 1993).

Theories of visual attention broadly fall into two types. One is 'spotlight' based accounts (*e.g.*, Eriksen & Yeh, 1975; Posner, 1980). The other is object-based approaches (*e.g.*, Duncan, 1984; Pinker, 1984). Object-based theories are currently popular as central to accounts of the processing of text in reading (*e.g.*, Treisman & Souther, 1986; Duncan, 1987; McConkie & Zola, 1987). A particularly influential example of this approach is that of McConkie & Zola (1987).

McConkie & Zola (1987) used an eye movement contingent display, in which a passage was presented on a single line of text in which one critical letter in a word was changed during saccades. The ability to detect this critical letter in a word *n* letters from fixation was estimated. From this data, McConkie & Zola estimated the probability of attending to any letter in a word by constructing a letter perception index (LPI). On the basis of the distribution of this LPI as a function of letter position in a word, McConkie & Zola concluded that "apparently all letter positions of that word are attended during a single fixation". They proposed from this that readers attended to "word-level objects" rather than "some other type of region" (p. 393).

Although not based directly on the empirical evidence from their study, McConkie & Zola (1987) concluded by suggesting that a page of text can be considered as a "four-level object hierarchy." The visual system provides a representation of the entire page and attention can then be applied to any of these levels: viz. to line-, word- and letter-level objects.

In a very recent study, Bock, Monk & Hulme (1993) examined what visual information determines how attention can be applied to words. They conducted a number of experiments in which subjects were presented with pages of text in which letter size was systematically transformed between 9 and 18 points, as was the number of letters in which size was changed in this way[2]. Time to read each page of text was then measured. An example of the conditions used is shown in Figure 1.2.

The first experiment showed that changing the size of letters in a word significantly slowed reading. It was suggested that this result was due to disruptions at a 'word-level' because the effect of size-changing the number of letters in a word (conditions 3–4 vs. 1–2) was found to have a greater effect than size-changing the number of letters on the page (conditions 1–3 vs. 2–4). The second experiment found that this effect remained using uppercase text. Bock, Monk & Hulme concluded from this result that the effect of changing letter size could not therefore be due to a disruption of the visual processing of word shape (making the assumption that uppercase words do not have any distinctive shape). Bock, Monk & Hulme therefore suggested that the effects of font size transformations might be due to disruptions in some word grouping mechanism.

---

[2] Text transformation studies in which case and size has been alternated within a word (e.g. Smith, 1969; Fisher, 1975; Rudnicky & Kolers, 1984) have usually been taken as evidence that the visual processing makes available information only about letter features which is used for reading.

Drawing from psychophysical evidence (Wilson & Bergen, 1979), Bock, Monk & Hulme noted that vision operates at a range of spatial scales, differing by approximately one octave. They suggested that, as their experimental conditions contained words having letters differing in size by a factor of 2, it might be the case therefore that "separate attention to, and processing of, those changed letters [is required]" (p. 86). On this basis, they measured reading speed when adjacent letters in a word were similar (but not the same) in size but the overall range of letter size change remained the same. This was done by having a 9–point letter next to a 14–point letter, which was next to a 18–point letter, which was next to a 14–point letter, and so on. This produced an unexpected finding. Although reading rate was still affected, as with all text transformations, it was relatively less slow in this condition. Bock, Monk & Hulme (1993) concluded by suggesting that this result was because words in this condition, which had greater neighbouring similarity, were easier to group and direct visual attention to.

Control
parasite called "Plasmodium". Although it had nothing to do with

Condition 1: 1 letter/word, 125 words
parasite called "Plasmodium". Although it had nothing to do with

Condition 2: 1 letter/word, 250 words
parasite called "Plasmodium". Although it had nothing to do with

Condition 3: 2 letters/word, 125 words
parasite called "Plasmodium". Although it had nothing to do with

Condition 4: 2 letters/word, 250 words
parasite called "Plasmodium". Although it had nothing to do with

Figure 1.2. Example of text used in Bock, Monk & Hulme's (1993) study. Letter size varied between 9 and 18 points. This manipulation occurred as a function of the number of letters in a word, between 1 (conditions 1-2) and 2 (conditions 3-4) and the number of letters changed in a page, between 125 (conditions 1-3) and 250 (conditions 2-4). Text presented to subjects was whole pages. Time to read each page (mean word length =411 words) was the dependent variable.

The findings of both McConkie & Zola (1987) and Bock, Monk & Hulme (1993) are interesting. However, the conclusions reached from them are open to debate, for a number of reasons. The first reason is that the explanation of these findings in terms of an object-based visual attentional mechanism rests, it is argued, on questionable assumptions. In order for McConkie & Zola's account of attention to be applied to any 'object-levels' on a page of text, the objects themselves need to be defined. This is where the problem with

object-based attentional accounts lies. It is suggested (and a case made for the argument in Section 1.4) that "objectness" is defined by visual processing: what objects are perceived depends not only on the structure of the image but also, and importantly, on the state of the visual system at a particular time.

The second reason to question the explanation of the data in terms of an attentional mechanism is based on the conclusion by Bock, Monk & Hulme that an observed slower reading rate for disruptions to uppercase texts could not be due to a change in the visual representation. They offered an explanation in terms of attention (or failure of it in this case). This interpretation is based at least partly on the assumption that uppercase words do not have a distinctive shape. As outlined above, this assumption is not based on any knowledge of what the visual representation of text is, and importantly, what vision is capable of delivering.

Finally, there is a methodological consideration to be made. Not only the Bock, Monk & Hulme study, but also other studies have used silent reading rate as a measure of the effects of changes to the visual processing of text. It is suggested that it cannot be determined what, out of the many linguistic and visual cues existing in text is affected by transformations of it from a measure of reading rate alone. Changes in reading rate which are interpreted as an index of the visual processing affected or occurring must inevitably rest on a number of assumptions which may well be difficult to justify.

1.3.6 Psychophysical studies

In comparison to vast amount of cognitive studies into the role of visual processing in reading, an extremely small number of studies have been conducted into this subject by the vision community. The most notable studies examining visual processing in reading are those by Legge and colleagues. Many of these have been concerned with developing ways of ameliorating the effects of poor vision on reading. However, the first of these studies (Legge, Pelli, Rubin, & Schleske, 1985) was concerned with normal vision.

Legge *et al.* (1985) used a moving window procedure in which a single line of text was swept across a display and the subject simply had to read aloud what was displayed. The speed with which the text swept across the display was increased until the subject could no longer read the text without making errors. A psychometric function was then obtained by plotting the reading rate (number of words correctly read) as a function of the scanning rate of the text. By varying a number of parameters which were thought to be important for visual perception, and measuring changes in the psychometric function, the effect of any parameter on the visual processing of the text could, it was argued, be determined. The size

of the text, spatial frequency bandwidth, and sampling density required for reading were examined.

The spatial frequency bandwidth (in other words, the spatial scale) required for reading was examined by blurring the text to different degrees by low-pass filtering using a ground glass diffuser, and measuring reading rate for different amounts of filtering. Two findings are of interest. The first is that the effect of filtering was independent of character size when spatial frequency was expressed as number of cycles per character. The second is that reading rate was unaffected by filtering which removed spatial frequencies above 2 cycles per character. Legge *et al.* concluded from this that only one spatial scale is necessary for reading.

Sampling density required for reading was examined by placing a grid over the display which had differing numbers of holes in it, thus varying density from 1.4 x 1.4 to 22 x 22 samples per character. From the Nyquist theorem, it can be shown that the information in a letter should be represented sufficiently well for reading if there are 2 samples per cycle, in this case (since 2 cycles per character was required for reading) 4x4 samples per character. However, at least twice this number (8 x 8 samples per character) was required for optimal reading for all sizes of text except the smallest. Legge *et al.* suggested that this finding could only be explained by proposing that (fine spatial scale) noise introduced by the sampling process interferes with information about characters contained at coarser spatial scales. This is interesting because this is what would be expected from visual processing according to one model of early vision: the MIRAGE model (Watt, 1988). This model is considered in some detail in Section 1.4.

Although the Legge *et al.* (1985) study is interesting, not least because it represents an attempt to conduct a psychophysical examination of the visual processing of text in reading, it has a number of methodological problems which unfortunately limit the usefulness of the findings and the conclusions reached. The first problem is the use of the moving window technique. This procedure is very different from that encountered in the normal reading situation in that the role of eye movements is both limited and atypical, and visually peripheral information is not available because of the narrow extent of the window itself. This must mean that any part played by information extracted in the periphery in the visual processing of text would not be found. The conclusion that Legge *et al.* reached, that only one spatial scale is necessary for reading, may therefore be an oversimplification in typical reading situations.

Secondly, the visual context in which words normally appear–in text–and its effects on visual processing were still not fully considered in this experiment; only one line of text was

displayed at a time. To reiterate, it has been argued above that the visual context in which words appear may have an important impact on the visual processing of text in reading and this experiment fails to address this issue properly.

Finally, Legge *et al.* used a measure of reading rate to determine what, and when, visual processing was being affected. Using reading rate as a measure of the visual processing being performed may introduce a number of problems in interpreting the data obtained by this method, as already mentioned (Section 1.3.5). Specifically, it is not possible, from a measure of reading rate, to determine which of the many visual and linguistic cues available in text is being affected or is used in perfroming a text processing task.

1.3.7 Summary

It was stated at the beginning of this chapter that in outlining some of the research into the role of visual processing in reading the aim was to discover and discuss the extent to which it has provided an understanding of the visual processing of text and to see how this might indicate the way in which a further study might contribute to the debate. It is hoped that this review has served to illustrate that the research has failed to provide anything like an adequate account of the visual processing of text, and indicate some of the factors contributing to this failure. Several reasons can be identified for the poor understanding of the visual processing of text which have important implications for any study and understanding of it.

The first reason concerns methodology. Although each methodology was found to have its specific problems, there is a fundamental problem common to all of these methodologies. It is the inability of methodologies employed in reading research (*i.e.* using measures of reading rate, lexical decision, probed letter detection, eye movements, and current psychophysical procedures) to be able to determine or control what, out of the various visual and linguistic cues available to perform the task required, is responsible for the behaviour observed.

The second reason is that most of the studies done to date have been concerned only with whether visual processing makes explicit information about whole word shape or only about letter features for the purposes of providing an orthographic representation of any word. In doing so, the assumption appears to be made (tacitly) that the *visual* context in which a word appears does not affect its visual processing. Isolated words are treated no differently from words in text. Certainly, such techniques are incapable of determining the effect of the visual context in which words appear on the visual processing they receive.

These shortcomings might be seen to be based at least partly on assumptions about what reading involves. Section 1.2 argued that reading involves a number of visual text

processing tasks, of which some must be performed before the orthography of each word can be determined. None of the research considers these issues explicitly. Eye movement studies bear upon some (for example word segmentation) only because eye movements suggest the operation of the task itself. McConkie & Zola (1987) discussed the importance of different text processing tasks, but did not attempt to examine them. From this discussion, the conclusion must be reached that for a proper study, the visual processing performed in individual text processing tasks needs to be studied, rather than the product of all the visual and linguistic text procesing tasks, the reading process. Failure to do so means that the inferences required to interpret the findings of studies examining the role of vision in reading are inevitably unsatisfactory. An experimental technique which can separate visual from other, non-visual (*e.g.*, linguistic) aspects of the reading process is needed for an effective study of the visual processing of text. However, devising such a technique is acknowledged to be a difficult obstacle to overcome.

The final, and most important point to make is that the preceding discussion indicates that both research and models of reading make assumptions about what vision is capable of delivering in the visual processing of text in reading. These assumptions are not based on any sound knowledge of what vision does, nor of the nature of the visual representation, particularly of text. It is suggested that only by giving full consideration to the operations which might be performed by early vision, and by attempting to discover how well a model of vision can account for human visual processing of text, can any proper understanding of the visual processing of text be gained, and any proper study of it be started.

1.4 Vision

A primary task for vision is to extract from the image information about which a decision needs to be made. The decision might be "move to the left because there is an obstacle" or it might be "read the next word." An important aspect of vision, stated by Marr (1976; 1982) is that vision is only possible by knowing something about how things in the world are organised. For example, vision expects that things which belong to the same object will be connected or close together, and conversely, that sudden discontinuities are likely to be due to boundaries between different objects (separate information).

It was the Gestalt psychologists (Wertheimer, 1923; Koffka, 1935; Köhler, 1947) who first drew attention to this constraint, which became embodied in a number of "laws" of perceptual organisation, based on grouping and segmentation. Two pertinent examples are the laws of proximity and similarity. The Gestalt law of proximity asserted that items which are close together will be grouped together and perceived as a whole. The law of similarity

asserted that items of similar attribute (*e.g.*, size, contrast) will be grouped together. Although the Gestalt principles, as stated, are vague the issues they raise about perceptual grouping and visual processing remain important and influential, and are pertinent to this discussion.

The constraints just outlined on the way in which vision must operate are important for understanding how typographical practice has, through its evolution, led to the way in which text is organised. A page is grouped into sentences and lines. Each line is perceived as separate from adjacent lines because of the amount of space between lines. Every line is actually a list of words which are close, *i.e.* grouped, together but the words themselves are perceived as separate from each other word by the amount of space between each word. The letters in each word are similar in size, and group together into a word, but each letter in a word is separated by a certain amount of space. It is important to note that the relationship between these (typographical) properties is governed by the way in which vision operates, and in particular the visual description and representation which provides the basis for this type of 'grouping'. It is therefore important to consider in some detail the visual processing concerned with making and using these representations and descriptions of images.

1.4.1 Image Algebra

In most natural images, changes in the light intensity which are attributable to those images do not correspond in a simple manner to the pattern of light intensity reaching the eye. Reflectance, illuminant sources, object texture, object position (causing, for instance, occlusion and shadows) and photon noise are some of the many factors which determine the complex and consequently noisy pattern of intensity projected onto the retina and entering the visual system. A very good way of specifying where in the image intensity changes occur that correspond to objects (information) is first to smooth (average) luminance values, which provides a method of determining the extent–or scale–of the intensity change, and then analyse the smoothed changes by the operation of differentiation. This replaces the intensity values in the image by a new set of values which records how much the intensity is changing at every point in the smoothed image. However, luminance changes can be caused by features other than objects, such as shadows, or the type of illuminant(s), and of course, it is necessary to be able to detect the source of these changes. Differentiating the first derivative (which yields the second derivative) records how rapidly this derivative is changing by causing a positive and/or negative peak, the statistics of which are dependent upon the type of luminance discontinuity, such as an edge or a shadow. This is illustrated in Figure 1.3.

There are many types of filters which could be used to smooth the image. In reality, the statistics of natural images serve to limit the choice of appropriate filters which are useful for a visual system. In attempting to determine how intensity values correspond to different objects, estimating the local mean luminance changes and the standard deviation of this change–a spatial scale term–in the second derivative are statistically reliable measures to be obtained. To get as reliable measure of these statistics as possible, it is desirable to obtain as many samples of luminance values as possible. Increasing the size–the spatial scaling term–of the filter increases the number of samples obtained from the image. However, increasing the size over which the samples are taken brings with it the danger of overlapping different, neighbouring, regions in the image, thus being unable to fulfil the requirement of being able to distinguish between different regions. A balance between these two requirements is needed.

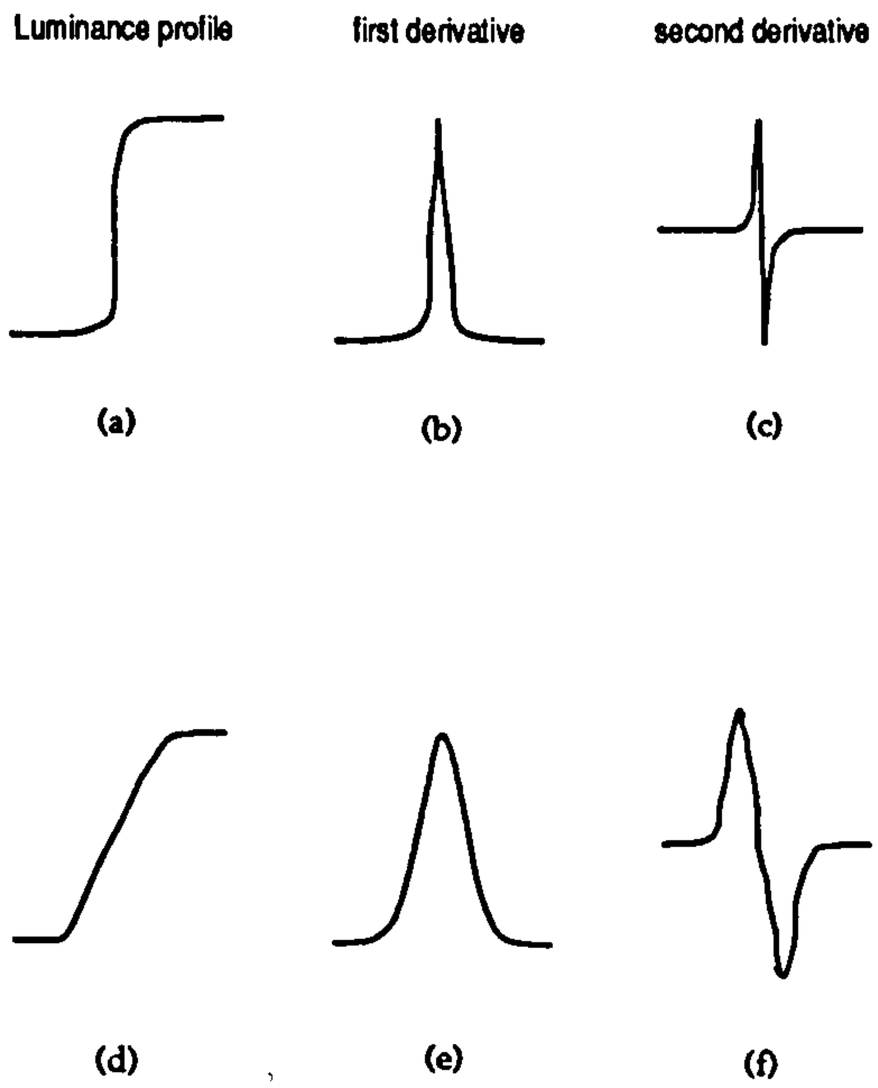


Figure 1.3. Two examples of different luminance profiles (a & d) and their first (b & e) and second (c & f) derivatives. Note how the different luminance profiles produce different derivatives.

In estimating the required statistics of mean and standard deviation of rate of intensity change in the image, it is useful to take into account the fact that the further away from some local point (neighbourhood) something is, the less likely it is to belong to the same thing. Using a weighting function which gives greater emphasis to local rather than distant points

in an image does take this constraint into account and importantly, serves to provide the required balance between the two requirements stated above. A weighting function which serves exactly this purpose, and one which also smoothes the image–the other necessary requirement–is a Gaussian filter function (see Watt, 1991; Marr & Hildreth, 1980 for a mathematical treatment of this conclusion).

In two dimensional images it is necessary to measure derivatives along a continuum of different directions. There is an operator which performs the required function of measuring the second derivative in every direction: the Laplacian operator. From the preceding discussion, it follows that a spatial filter is required which is a Laplacian operator, and has a Gaussian function. This is the Laplacian of Gaussian.

The algebraic properties of the required processes of smoothing and differentiation allows them both to be performed by a single operation–convolution–of the image with a filter. Therefore, convolving the image with a Laplacian of Gaussian fulfils all the necessary requirements for extracting information from images.

The amount of smoothing required to find information in the image will depend on the size–or scale–of the object, which cannot be known in advance. In many images, including text, objects have different types of information existing across the same spatial extent. For example, information about the letters in a word is contained at a small, or fine spatial scale. Across the same spatial extent, information about a whole word is contained at a larger spatial scale. This means that in order to be able to extract all the information required for all the visual tasks that might be required, there needs to be a range of spatial scales over which the convolution operation is performed on the image. The spatial scale term is then measured as the distance over which this filtering takes place.

The concept of spatial scale of image analysis is a very important one for this thesis. Several sources of evidence demonstrate that the visual system operates at a range of spatial scales (*e.g.*, Campbell & Robson, 1968; Blakemore & Campbell, 1969; Wilson & Bergen, 1979; Georgeson & Harris, 1984). Furthermore, there is evidence that early vision contains the mechanisms for implementing the type of convolution described (*e.g.*, Enroth-Cugell & Robson, 1966). These findings, and the arguments outlined above about what vision needs to do, provide the basis of a number of computational models of vision.

Two of the most influential of these are that proposed by Marr & Hildreth (1980) and that proposed by Watt (Watt & Morgan, 1985; Watt, 1988). In both, the effect of spatial scale is modelled by a convolution of the image with a range of Laplacian of Gaussian filters differing in space constant, in agreement with the evidence (above) for the existence of, and need for, filtering the image in the manner described, at a range of different spatial scales. In

the Marr & Hildreth algorithm, zero-crossings in the second derivative of a set of independent, different scale filters are used to detect edges of objects. The model has some modest success in accounting for some of the psychophysical data. However, it does have a number of serious shortcomings. Using zero-crossings in this way to locate information in the image turns out to have a very limited applicability (*e.g.*, Watt & Morgan, 1984). Watt & Morgan (1985) went on to propose a different account of the way in which information is recovered from the image by vision.

1.4.2 Watt & Morgan's (1985) MIRAGE algorithm

Unlike the Marr & Hildreth algorithm, the Watt & Morgan (1985) model, termed MIRAGE, proposes that although zero-crossings in the second derivative of the convolution are located, they are used only to separate the convolution of each filter into its positive and negative parts. On the grounds of psychophysical evidence (*e.g.*, Henning, Hertz, & Broadbent, 1975; Nachmias & Rogowitz, 1983; Watt & Morgan, 1983) Watt & Morgan (1985) proposed that the visual system does not have independent access to the responses of individual filters. Instead, the response from all the filtered images of different spatial scale are added together wherever the signs of the response from different scales are in agreement, to produce two response images, containing the sum of all the positive responses and the sum of all the negative responses. The location of the zero-crossings mark a zero-bounded response distribution for which the statistics of centroid (to provide spatial localisation), and mass (to measure contrast) are calculated. The orientation (direction) of each region is calculated by finding the direction of its principal axis. The length of a region is then estimated by calculating the standard deviation of the response distribution in the direction of its principal axis.

The resulting image is analysed by extracting all the positive and negative regions which are statistically significant different from the mean value. Each of these regions is then represented by a list of parameters which provide a symbolic image description of the location, length, orientation and mass. This sequence of operations is illustrated in Figure 1.4. (a more detailed account of the rationale behind the operation of MIRAGE and the empirical evidence for it can be found in Watt, 1988; 1991).

1.4.3 MIRAGE and grouping

A consequence of the way in which vision operates, according to MIRAGE, in terms of the local operations performed on an image (described above) is that the calculation of spatial position, which is a global operation, cannot be determined *absolutely*. As a consequence, what has to be estimated instead is *relative* (difference in) spatial position. Because of this

problem (compounded further by the problem of propagation of error in the visual system caused by noise and distortions), calculation of spatial position in the image needs (if local operations are performed) to be determined by an iterative process of comparison of all regions in the image with each other. The number of iterations required will increase as a function of the number of regions, so the necessary computations required to fulfil this obligation also increases very quickly, and therefore so does the danger of not being able to complete them in real-time. Given that this would be a disaster for a visual system, and the fact that the visual system solves these problems somehow (assuming it operates in this way) because it operates in real-time, it must employ a means to restrict the number of regions involved to a number for which the required iterations can be computed in real-time.

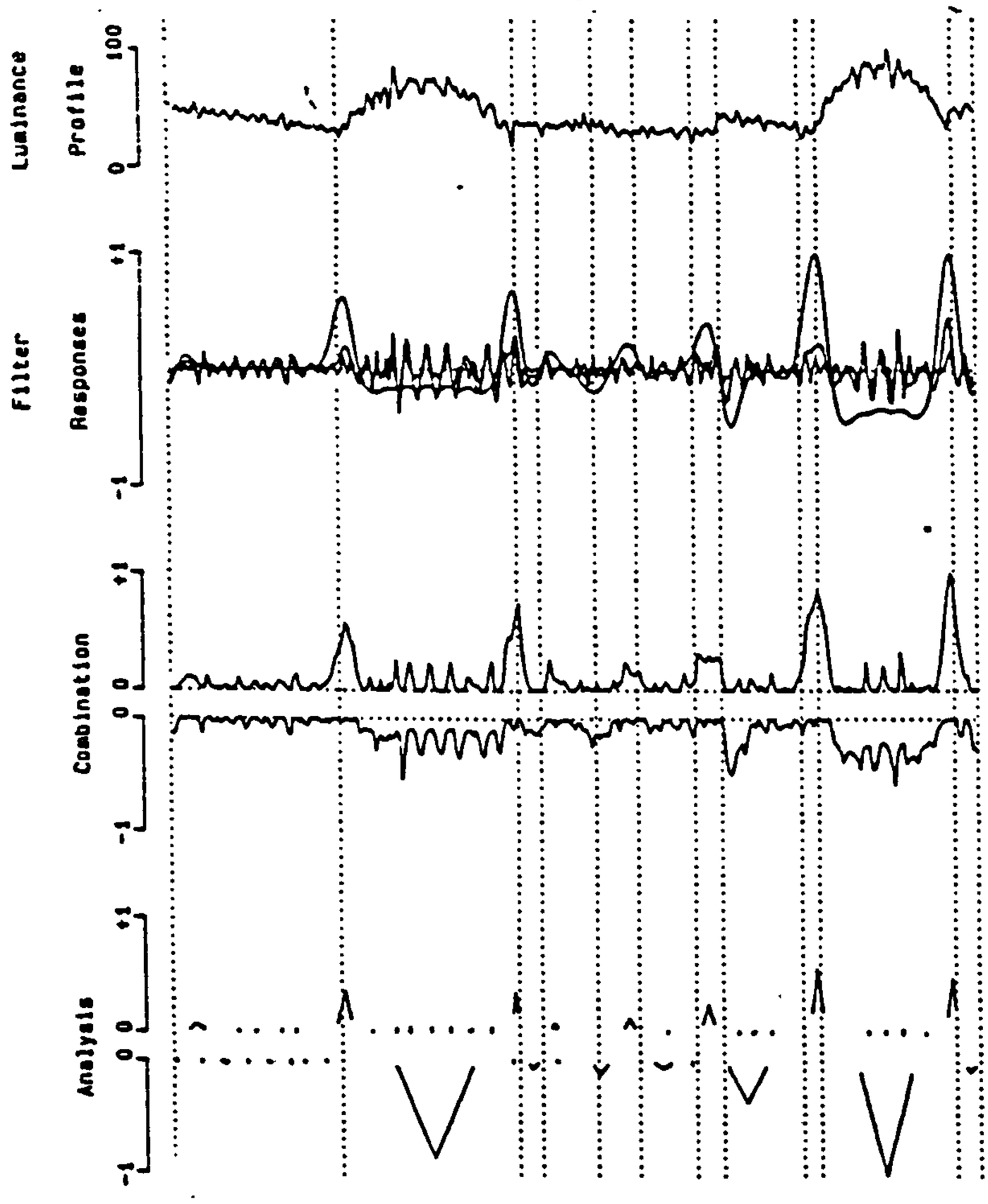


Figure 1.4 From luminance profile to representation, according to the MIRAGE algorithm. The top row shows a luminance profile of an image. The second row shows the responses of a range of Laplacian of Gaussian filters differing in spatial scale. The third row shows the combination of these spatial scales. The fourth row shows the MIRAGE analysis of the filter responses as a set of zero-bounded response distributions. Reproduced from Watt (1988) with permission.

Watt (1987, 1988) proposed that *grouping* regions of information would serve an important computational role by reducing the number of elements, and therefore the number of iterations and computations required to represent and describe the image. There are many types of grouping procedures. However, one which takes into account the constraints outlined earlier about how things in the world are organised, for example that adjacent regions of an image are likely to be a projected from adjacent parts of space–grouping on the basis of proximity–might be considered most appropriate. Note that grouping as a visual process means that the level of grouping suitable to represent and describe the image will depend on how the visual task determines what needs to be extracted from the image, and the degree of spatial position information that task requires.

Watt (1987) proposed an extension to the MIRAGE algorithm which solves these particular problems faced by the visual system. In this proposal, spatial position is represented initially at coarse spatial scale, at which grouping will produce few regions of information in the image because of the degree of smoothing, so the position of each group can be calculated quickly. Over a time-course of visual processing, the largest filter operating in the visual system is progressively switched out to provide successively finer representations of spatial position as the region of each group is gradually replaced by a representation of finer scale groups of regions. This is performed until all the information for a given visual task requiring spatial position to be calculated is analysed and resolved, or until only the smallest filter in the system remains active. Spatial position for elements within each group is determined with respect to the spatial position determined for that group represented by the previous largest filter: relative position is calculated. This operation allows a hierarchical grouping process in which the number of computations required to compute spatial position is generally kept to a low enough number to operate in real-time, but is also determined by the properties of the image. This process is illustrated in Figure 1.5.

This dynamic component to the model is based on the findings of Watt (1987). Watt proposed that the precision with which geometric positional information can be measured (in this case the orientation of a line) will depend on the degree of smoothing (the spatial scale of visual processing) and the length of the line. The size of the filter used to perform the task can be estimated by decreasing the length of the line until it becomes blurred by that filter to such a degree that its orientation cannot be resolved by the visual system. For this task, the orientation of large lines will be determinable at all but the coarsest spatial scales, whereas the orientation of short lines will require processing at fine spatial scales.

Watt (1987) obtained psychophysical measures of the precision of orientation estimation of lines of different lengths as a function of the exposure duration. It was found that the way that performance at detecting small departures from vertical varied with line length was indeed a function of stimulus duration, as predicted by the model on the basis of a coarse-to-fine spatial scale analysis. Visual resolution (which requires a fine spatial scale representation, but not about spatial position) did not vary with exposure duration.

It is important to note that in this way, MIRAGE has a property of grouping which is a consequence of the behaviour of the largest spatial scale operating in the system.

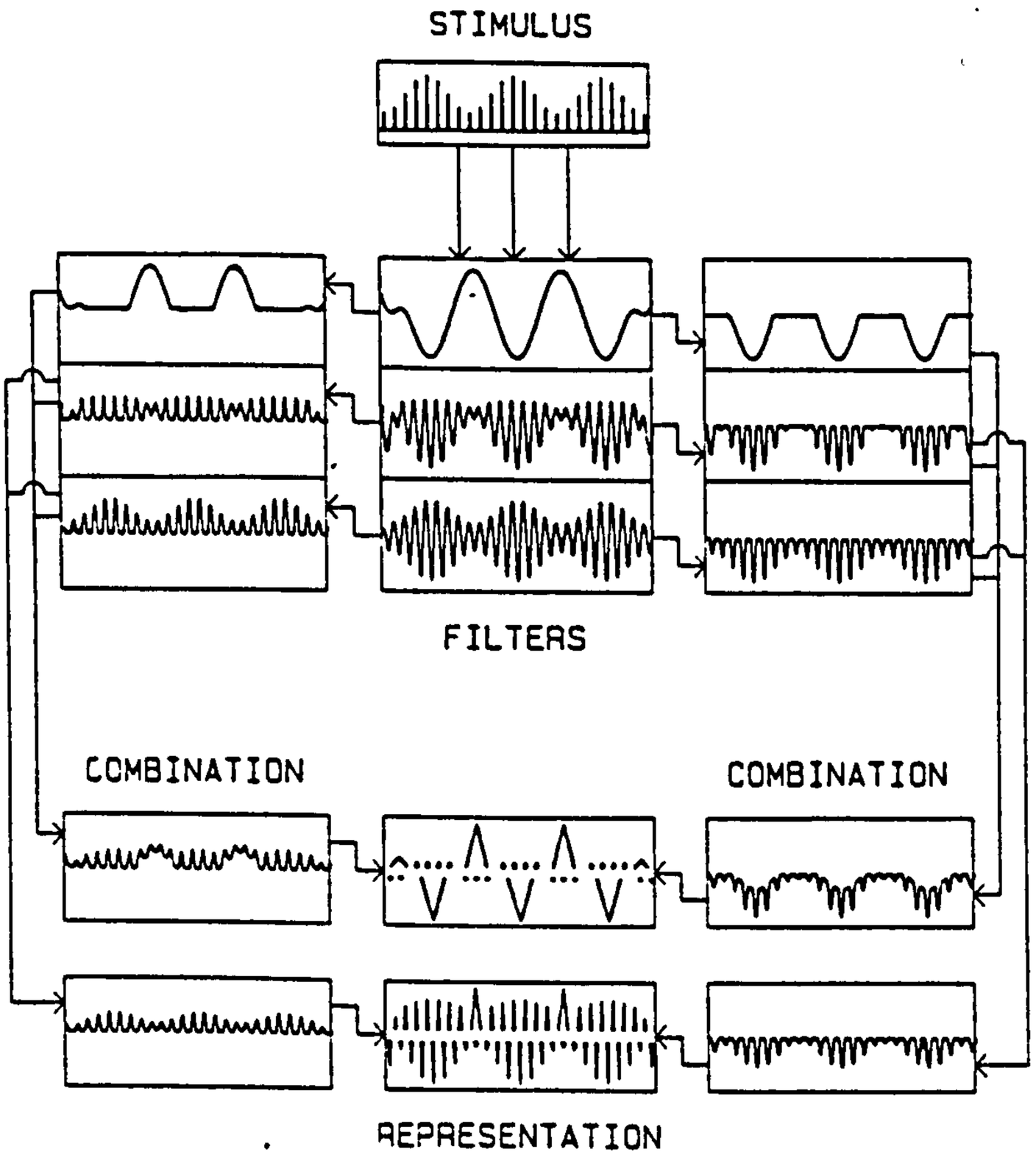


Figure 1.5. The effects of the largest active filter on the representation of an image, according to a dynamic MIRAGE. The top box shows a luminance profile. The next set of panels below this show the filter response from three different spatial scales: coarse (top), medium (middle) and fine (bottom). The bottom set of panels shows how the representation of the image differs according to the largest spatial filter active. In the top row, the representation has fewer response distributions: the largest active filter has 'grouped' elements of the image together. In the bottom row the representation is much finer as the largest filter active has now been switched out. Reproduced from Watt (1988) with permission.

Eye movements would be expected to play a central role in this spatial scale based hierarchical representation process. It is known that when the eyes make a saccade, they tend to move to the centroid of a group (Findlay, 1982). This suggests that the spatial position solution derived from the largest filter active is used to determine the alignment, within a large scale group, of the foveal point of fixation. This then enables the finer scale resolution at the fovea to continue producing a progressively finer scale representation. Eye movements will usually result in changes in the content of information in the image, and therefore the number of possible regions in the image. Because the most appropriate spatial scale of analysis will depend on the task, this means that the time required to compute a full representation of the image will depend on the time required for the visual system to reach the required level of visual analysis to perform the task based on this time-course of coarse-to-fine spatial scale of operation. Watt (1988) points out that the time-course of switching out the filters must also be determined by these factors. In order to match the time available for processing to the scale of processing required for a task the time-course of processing should be under control.

Eye movement control would serve as a particularly simple form of control of processing, and there is strong evidence, much of it from studies of eye movements in reading, to demonstrate that the decision of when and where to move the eyes is determined by the information obtained from the visual processing performed in both the periphery and central vision (*e.g.*, McConkie & Zola, 1990; Nyman, 1990). It seems not unreasonable to conclude that this decision is determined by whether an adequate representation of the image for a particular task has been completed.

A good example of this comes from reading. Reading rate varies with task or material difficulty (*e.g.*, technical material vs. a novel) and task requirements (*e.g.*, proofreading vs. skimming. See Tinker, 1965). Although there will be many factors which determine this rate, some of the reading time must be a consequence of the time the visual system would require, if it operates on a coarse-to-fine spatial scale of analysis, to reach a level of analysis (or spatial scale) at which the information required to visually process the text in each of the different reading situations is contained.

It is interesting to relate this description of vision to the findings of Legge *et al.* (1985), not least because the findings also turn out to serve as further evidence of the dynamic operation of MIRAGE just outlined. Legge *et al.* (1985) found that mean asymptotic reading rate was approximately 250 words per minute. From this it is possible to estimate that time to read a word was approximately 250 msec. Given that the optimal character size for reading was found to be 20 arc min (about 10– 12–point Times font read at 40–50 cm), and a required

critical spatial scale of 2 cycles per character was also found, a spatial scale of 10 arc min should have been optimal for the visual processing of text in this experiment. In this task, after 250 msec, the visual system would, on the basis of the data of Watt (1987), be expected to be operating at a scale of between 5–10 arc min, as was found.

Another finding of Legge *et al.* (1985) is also relevant in terms of the Watt (1988) model. It was noted earlier (Section 1.3.6) that Legge and colleagues found that at least twice the number of samples than that which would be sufficient on the basis of the Nyquist theorem was required for optimal reading for all sizes of text except the smallest. It is difficult to provide any explanation for this other than proposing that fine spatial scale noise introduced by the sampling process interfered with information about characters contained at coarser spatial scales, and indeed this was the conclusion reached by Legge *et al.* This supports an interaction between different spatial scales, as proposed in MIRAGE.

1.4.4 Summary

The Watt (1988) model of vision has been described in some detail here for a number of reasons. First, the operations of the model have specific consequences for the way in which it proposes that the visual system provides a representation and description of the information contained in text images, and therefore the visual processing of text. Second, the model is able to account for a wide range of psychophysical findings (a full account of the empirical evidence for the model has been described elsewhere *e.g.*, Watt, 1988). Third, the model is compatible with much of the available neurophysiological data. Indeed, in this vein it is interesting to note that the requirement to separate the positive and negative responses is justified on the basis of, and provides an explanation for, the finding that ON-centre and OFF-centre mechanisms are kept separate in the visual pathway (*e.g.*, Perry & Silveira, 1988; Wässle, 1988). For these reasons, the model is considered to be the most appropriate to use in this thesis, and the research itself will test the ability of the model to adequately describe human performance in a real visual processing task: text processing.

1.5 Modelling the early visual processing of text

1.5.1 Introduction

It was argued at the end of Section 1.3 that the failure of the research into visual processing in reading to provide an adequate account of the role of vision in reading was due in part to an inadequate description of early visual processing. It was concluded from this argument that a more appropriate approach to the study of the visual processing of text was therefore to start with a proper consideration of early vision. The aim of Section 1.4 was to provide

an account of the purpose of early visual processing, the computational requirements needed to fulfil this purpose. A computational model of early vision which proposes a means of extracting the information in images which provides an explanation for much of the available human visual (psychophysical) processing data was described.

Applying a computational model of vision (such as one based on the Watt, 1988 model, given its ability to account for much of the available data) to text might provide a useful contribution to an understanding of the visual processing of text. The potential usefulness of this approach has been argued for not only here but elsewhere (Monk, 1985; Watt *et al.*, 1990; Watt, 1993; and Bock, Monk & Hulme, 1993).

1.5.2 Computational studies of early visual processing of text

Three computational studies have so far been conducted, all of which can be considered to be preliminary analyses. The first of these was by Brady (1981) who used an implementation of the Marr & Hildreth (1980) algorithm to show the type of information which might be used for word segmentation. This analysis had some very limited success in making explicit word boundaries, but unfortunately, only when knowledge of what word boundaries 'looked like' was known in advance. It is very unlikely, however, that an edge finding algorithm based on locating only zero-crossings is implemented in human vision (*e.g.*, Watt & Morgan, 1984), and both these facts serve to limit the usefulness of this study.

The second and third studies: Watt *et al.* (1990) and Watt (1993) applied an implementation of the Watt (1988) MIRAGE model to different pages of text. Watt found that different levels of information, organised at a range of spatial scales, was made available by the model in a representation of the text image. Different scales contained information making explicit lines, words, word boundaries, and finally letters and letter features, in a coarse-to-fine spatial scale of analysis. Watt (1993) has also shown that the information made explicit in the representation of the model is dependent upon the typographical arrangement of the text. This is an issue which is taken up in the next section. The representation of text in the manner shown in these studies is far different from that assumed by models of reading, and on the basis of a coarse-to-fine visual analysis over time, would provide a representation of this information in a sequence not predicted by models such as the interactive-activation model of McClelland and colleagues.

The computational modelling of text done to date by Watt (Watt *et al.*, 1990; Watt, 1993) suggests that using a model of vision in this way may indeed be a promising way to proceed in a study of the visual processing of text in reading. However, the most important aim in modelling the visual processing of text is to establish whether the human visual processing of text can be described by the model. To this extent, none of the computational studies

performed to date have either attempted to do so (Watt *et al.*, 1990; Watt, 1993), or have been able do so (Brady, 1981). This is a specific aim of this thesis.

1.6 Typography

In the Watt (1993) analysis of text, outlined in Section 1.5, the information made explicit by the MIRAGE model of vision was found to be related to the typographical arrangement of the text. In Section 1.4 of this chapter, which discussed vision and vision and grouping, it was considered how the (typographical) arrangement of text might be seen to be constrained by the way in which vision operates. This has a number of implications for a study of the visual processing of text, and for this reason, the relationship between typography and vision is discussed here.

1.6.1 Typography and vision

The aim of typography is to ensure the efficient transmission of the writer's meaning to the reader through the visual appearance of the text (see Southall, 1988, for a discussion). Typography specifies the relationship between the size and proximity of the various marks on the page: the typeface, line spacing, word spacing and punctuation. This typographical arrangement determines the legibility, *i.e.* the speed and accuracy with which the information on a page can be extracted in reading.

There are several aspects of vision which serve to constrain the way in which text is laid out on a page which have consequences for the legibility of text. Because the resolution of the visual system decreases away from the fovea, and text is arranged spatially in rows of lines of words, it is necessary to align the fovea with different parts of the page by moving the eyes. In order to reduce the distance the eyes have to move, so as to enable rapid reading, arranging the text so that words and lines are as close to each other as possible is desirable[3]. On the other hand, there is a need to ensure that words and lines are perceived by the visual system as separate from adjacent words and lines. Thus, the typographical characteristics of the text must be adjusted to match the way in which the visual system operates because some characteristics of the visual system are fixed.

The basis of typographical effects is not understood. Typographical practice has developed through an iterative process of trial and error to ensure legibility. To discover something about the basis of typographical effects would be desirable, as it would go some way towards a proper understanding of typography. More importantly, because

---

3 Conventional printing practice was sometimes to put a large space in between words. However, this appears to have been motivated not by typographical considerations but seems more likely to be the cynical practice of compositors paid by the inch of typeset.

typography is constrained by vision, it would help to provide a better understanding of the nature of the visual processing of text.

1.6.2 Typographical research

There is a paucity of research specifically examining the effects of typography on legibility with the (claimed) aim of understanding the visual processing of text. What work there is therefore deserves discussion.

The most comprehensive set of studies examining the effect of typography on legibility was made between 1926 and 1965 by Tinker and colleagues. However, Tinker's studies, which run into hundreds, were concerned mainly with which physical aspects of text (primarily typographical arrangements, but other aspects such as contrast polarity, colour, and illumination were also studied) affected reading rate or eye movements. Details of the experiments can be found in Tinker (1965). It is sufficient here to describe a typical task in order to illustrate the type of methodology employed, and some of the problems of Tinker's approach.

The method most widely used by Tinker was the Tinker Speed of Reading Test (Tinker, 1955). This required the subject to read, as quickly as possible, two passages of text differing in some typographical parameter. Reading speed was measured as the number of paragraphs read in a given time (either 1.5 or 10 minutes). As a check for comprehension, subjects had to cross out particular words which spoilt the meaning. The text resulting in the fastest reading speed was simply judged to be the most legible.

Several problems with this approach are readily identified. First, the resulting measures of legibility were not independently comparable, because legibility was measured as a decrease or increase relative only to some other typographical setting. Second, closer inspection of a range of the texts compared in the tests reveals that although the different, compared, typefaces were typographically dissimilar, each was usually typographically 'correct' for that typeface. Third, reading distances were usually not specified, making the findings of studies difficult to interpret or compare. Fourth, as the discussion in Section 1.3 has shown, reading rate measures themselves may in any case have limited utility in most experimental situations. The final, but most important point to make in this respect is that Tinker did not offer any explanation as to why this pattern of results might occur. The conclusions of most of the studies are really no more than a re-statement of the findings: x can be read faster than y, so x is more legible: the work was completely atheoretical.

That Tinker's work remains widely cited as *the* classic psychological work on typography and legibility seems to owe more to the fact that it is really the *only* work, rather than to any insight into the relationship between vision and text it might have provided. It is clear that

this work does not go much way to establishing the visual basis of legibility, as Tinker claimed (Tinker, 1965).

1.6.3 Digital typography

The nature of the relationship between vision and text is changing rapidly with the development and increasing use of electronic text. Typography is no longer in the hands of typographers, but is instead under the control of the users of such systems. It is interesting then, to note that two other disciplines, human-computer interaction (HCI) and ergonomics, have devoted much research effort into two issues of relevance to this thesis. The first area of research is concerned with determining why reading is slower from visual display terminals (VDTs) than from paper; an issue still not resolved (*e.g.*, Dillon, 1992), despite the fact that numerous possibilities for the observed legibility differences, including illumination, display refresh rate, material difficulty differences, and contrast polarity have been explored.

It is suggested that inappropriate typographical arrangements resulting from either hardware (display) limitations, or more importantly the lack of appropriate typographical knowledge applied to the design of user-interfaces, may lie behind much of the observed reading differences between text presented on paper and on a VDT which remain even when the effects of all the factors mentioned above have been accounted for.

The second, and related area of research is the development of usable systems. "Usable" means, among other things, how well users can extract the information presented in terms of minimum effort, fewest errors and maximum speed. To this end, this area of research is also concerned with the development of specifications for the user-interface for a particular task. It is possible, and likely given the preceding discussion concerning the relationship between typography and vision, that many of the features determining usability are likely to be specified and determined by the visual processing requirements of the text displayed. It is possible then, that this is an area which might also benefit from a better understanding of the nature of the visual processing of text. This issue will be returned to in the last chapter.

1.7 Summary

This chapter began by outlining what is required in order to extract the information contained in a page of text (Section 1.2). It was argued that extracting the necessary information during reading is not a single stage event in which, from a whole page, the orthography of a single word was determined, but instead requires a number of visual processing tasks to be performed on the text.

A selective review of some of the key aspects of the research into the visual processing of text in reading then followed in Section 1.3. An important feature to emerge from this review was the failure of research to provide anything like an adequate account of the visual processing of text. A number of reasons were identified for this failure. Although outlined earlier, the importance of the issues raised by the reasons for this failure for any study of the visual processing of text in reading, and for an understanding of the issue itself make them worth restating.

The first reason is a methodological consideration. Specifically, the inability of the methodologies employed to date to be able to determine or control what, out of the various visual and linguistic cues available in texts, is responsible for the observed behaviour. In this respect, the importance (and difficulty) of developing a methodology which allows the separation and control of these factors to enable the visual processing of text to be investigated was stressed as necessary.

These methodological problems were seen to be rooted in further, conceptual, problems. A case was made for the first problem–assumptions made about the reading process–from the observation that research has concentrated upon the information made available from visual processing to provide only an orthographic representation of any single word. The extensive use of single words as stimuli supports this view. In using such techniques, there is an assumption that the *visual* context in which words appear does not affect their visual processing: isolated words are treated no differently from words in text. This ignores a number of important visual tasks normally involved in the processing of text, such as word segmentation.

However, it was suggested that there is a further, important, reason for the failure of the research to provide an adequate account of the visual processing of text in reading, and one which has a bearing on the methodological shortcomings. It is that both research and models of reading rest heavily on assumptions about what vision is capable of delivering. These are assumptions which have consequently ascribed a very limited role of vision in reading to that of providing a representation only of letter features (*e.g.*, McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), or at most, the shape of whole words, providing that the words are in lowercase (*e.g.*, Paap, Newsome & Noel, 1984).

The appeal to "attention" to attempt to provide an explanation of the processing of text (*e.g.*, McConkie & Zola, 1987) was suggested to appear to be based on the current poor understanding of what aspects of the processing of text in reading vision is capable of supporting, rather than any clear evidence that an attentional mechanism best accounts for any pattern of text processing performance in reading.

Because of this basic lack of consideration of what vision actually does, it was suggested that an understanding of the visual processing of text and any proper study of it, clearly needs to start by examining early visual processing itself, and the implications of this for the visual processing of text. This led to the need to consider the research into early visual processing (Section 1.4). The appeal of, and rationale for, computational approaches to vision was outlined. It was here that the concept of spatial scale of image analysis was introduced, and in particular with respect to the MIRAGE algorithm (Watt, 1988). This discussion led inevitably to how grouping, and representation of information within a group, was seen to be a consequence of the way in which vision operated. The possible consequences of the operation of this model for the visual processing of text were briefly outlined.

The small number of studies conducted which have applied computational models of vision to text were found to illustrate further the consequences of the way in which vision might operate may have for the processing of text. Specifically, the ability to provide a representation of the information contained in a page of text at successively finer spatial scales which 'un-groups' levels of detail in the image which might be used for reading. Noted also was the dependency of the various types of information made explicit by the MIRAGE model on specific typographical arrangements. This was seen as a feature which might possibly be exploited to study the visual processing of text.

It was concluded that applying a computational model of vision to text might be a fruitful way forward in understanding the visual processing of text if the difficult obstacle of devising a means of comparing modelled behaviour to human visual processing behaviour can be found.

1.8 Outline of the thesis

The discussion in this Chapter leads to the conclusion that the poor understanding of the visual processing of text is mainly due to the consequences of an inadequate account of early visual processing given, or assumed in reading research. Vision may be capable of supporting a far wider range of tasks in reading than commonly assumed, or at least ascribed, to vision by models of word recognition in reading (*e.g.*, McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; McClelland, 1986).

There is a clear need for a new, and different, theoretical and methodological approach to the issue of the visual processing of text in order to provide a better understanding of the role of vision in reading. This approach must be one which considers the nature of the visual tasks involved in the processing of text in the reading process, the nature of visual processing, and the need to very carefully separate visual from other non-visual (*e.g.*,

linguistic) factors in determining text processing behaviour. This would bring together both reading research and vision research, two areas of psychology which have been conducted in virtual isolation from each other.

This is the approach taken in this thesis.

There are three elements to the approach taken in this thesis which aim to meet the requirements for a proper study of the visual processing of text:

- To extend the computational modelling work in which a computational model of vision is applied to pages of text. It provides a new and useful framework within which a study of the visual processing of text can proceed.
- To carefully construct a set of suitable experimental text processing tasks which isolate the *visual* aspects of text processing.
- To compare the behaviour of the model to the behaviour of the human visual processing of text to establish to what extent the model can describe the human visual processing of text.

In Chapter 2, a computational model of visual processing based on the MIRAGE model (Watt, 1988) is applied to pages of text. The work of this chapter therefore adopts a similar approach to, and continues the work of, Watt *et al.* (1990) and Watt (1993). This computational image analysis will model the effects of using a range of different spatial scales similar to those found in human vision on the information made available in text images. This part of the thesis allows the generation of a set of definitive and testable hypotheses.

The rest of the thesis is devoted to testing these hypotheses in establishing how important the information made explicit by the model is for performing different text processing tasks of the sort performed in reading. That is, whether the same information might be represented by human vision and used in the visual processing of text.

Chapter 3 takes the first step in this aim by attempting to discover whether the human visual system does actually have access to the type of information made available by the model. A series of psychophysical experiments aim to determine the extent to which the human visual processing of text is spatial scale dependent. The next logical step is then to examine whether the extent of any spatial scale dependent text processing performance can be described by the model.

Chapters 4 and 5 aim to discover and establish how sensitive text processing performance is to changes in the pattern of information made available by the model of vision. If text processing performance changes in a way which can be described by changes to the information made available by the model, further support can be found for the

proposal that the information made available by the model is similar to that represented in human vision.

Chapters 6 and 7 adopt a similar rationale to examine the time-course of visual processing of text. In particular, the relationship between the visual representation of information across spatial scale and the time-course of the representation of this information based on a coarse-to-fine spatial scale of visual analysis is explored.

The final stage, dealt with in Chapter 8, examines the relationship between the model, the visual processing of text, and the product of this processing: reading. The aim is to establish the extent to which a measure of reading performance might be explained in terms of the visual processing component of that measure, and in particular how the model of the visual processing of text might account for this component of the reading process.

Finally, Chapter 9 discusses the interpretation and implications of the findings of thesis, and the extent to which the work can be considered to have provided any contribution to a better understanding of the visual processing of text.

2

A Visual Description of Text

A sensible first step in a study of the visual processing of text is to establish what information is made available by a model of early vision when applied to pages of text. In this chapter, details of the image processing operations necessary to model the basic operations of vision according to the Watt (1988) MIRAGE model of early vision are described. This is followed by the findings of applying this model to pages of text. There then follows a discussion of how the information made explicit by the model might describe the information expected to be required (on the basis of the discussion of the reading process outlined in Section 1.2) for performing different text processing tasks in reading.

This step in the thesis allows a detailed and precise specification of the information made available as a "visual description" of a text image, which is then used to generate a number of hypotheses about the behaviour of human visual processing of text. The chapter concludes with a proposal for a number of psychophysical experiments which mark the initial stage in testing these hypotheses.

2.1 Method

2.1.1 Text images

Nine samples of text were created from 3 pages of text taken from different academic publications. Each text sample was reproduced in 3 Apple Truetype fonts: Helvetica, a sans-serif font; Times Roman, and Palatino, both serifed fonts. All text was produced in 12–point (pts) size, left justified, with line spacing (defined as the vertical spacing between the baseline of one row of text and the topline of the row of text below that, ascending and descending characters excluded) set at 14pts. Hardcopy of each sample of text was then produced from an Apple LaserWriter IIg, at a resolution of 300 dpi (dots per inch). Nine text images were created from the nine hardcopy text samples by digitising 300x300 pixel sized

samples of the hardcopy text using a Hewlett-Packard ScanJet flatbed scanner onto a Hewlett-Packard Apollo workstation. This produced digital text images which had a resolution of 72 dpi and 256 greylevels (to provide anti-aliasing of the text). The text images were then subject to the following image processing transformations.

2.1.2 Image Processing

Each text image was subjected to a series of image processing operations representing the operations performed by early human vision, according to the Watt (1988) model, as outlined in Section 1.4. This series of operations is illustrated in Figure 2.1. The first stage was to subject each text image (Fig. 2.1a) to a spatial convolution with the isotropic operator Laplacian of Gaussian filter (Fig. 2.1b). The Laplacian of Gaussian has one free parameter, the space constant (*i.e.* size or scale) of the filter, defined as a standard deviation of the underlying Gaussian in pixels. The text images were filtered with Laplacian of Gaussians having standard deviations from 1 pixel to 32 pixels (in logarithmic increments). Computations were performed using floating point (64 bit) representations to avoid possible artefacts caused by rounding errors introduced by integer convolutions in images which have a high dynamic range, such as text images. The resulting filtered images were then subject to a second stage of processing.

Images which are filtered with a Laplacian of Gaussian have an expected mean value of zero, with a symmetric distribution of positive and negative values (representing the rate of change in image gradient; that is, the rate of change of the rate of change in image intensity) about this mean. These actual values were replaced by a value which represented the distance of that value from the mean value in numbers of standard deviations. This produced an image in which the value of each pixel was a measure of how different it was from the mean value of zero. The standard deviation of the values will depend on how much luminance change there is in the image: parts of an image in which there are little or no luminance changes (as measured by the second derivative of the Laplacian of Gaussian convolution), that is, where there is nothing interesting in the image to the visual system, will have low values. Parts of an image which have high luminance changes, such as edges or envelopes of shapes of objects will have high values. All image values which were less than a threshold of 1 standard deviation from the mean were set to zero, leaving an image having isolated positive and negative regions which differed significantly from the mean value of zero ("zero-bounded response distributions"). This is shown in Figure 2.1c.

The appeal of this approach is that provides for a convenient and reliable way of finding places in the image which are statistically likely to be either related or due to different sources, and is stable in the presence of random internal and external noise.

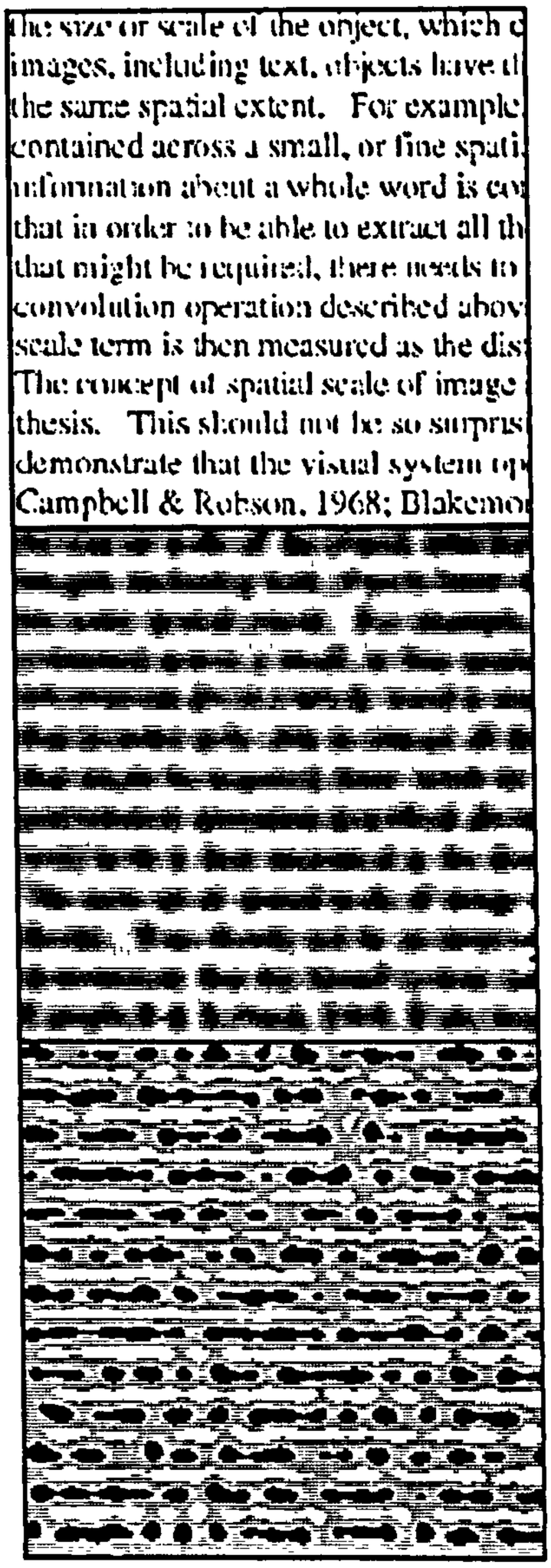
These two operations, filtering and thresholding, produced a resulting set of zero-bounded positive and negative response distributions, or *"regions"* found in the image at each spatial scale which was then analysed. Each region was described by measurement of the statistical central moments of mass, centroid, standard deviation and its principal axis, from which the orientation and length of each region can be determined (further detail can be found in Watt, 1988; 1991). The final step was then to provide a sentence-based symbolic representation in which each region in an image was described by its orientation, mass, length and position. A set of sentences describes the entire image at any spatial scale (Fig. 2.1d). This is an *"image description"*. Note that within any image, this set varies as the free parameter, the spatial scale of image analysis, varies.

2.1.3 Region analysis

Because all the regions found by this process can be described by their length and orientation, it is convenient to summarise the resulting image descriptions by constructing separate histograms of the distribution of mass of the resulting positive and negative regions as a function of the length and orientation of regions in the image at a particular spatial scale. The steps involved in producing this final description of a text image are shown in Figure 2.2. In each image (Fig. 2.1a), the mass of each region (or "blob") produced as a MIRAGE image description (Fig. 2.2a) was added to a histogram of region length at the point corresponding to the length of that region. Similarly, for each region, its mass was added to a histogram of region orientation at the point in the histogram corresponding to the orientation of the region. This produced a histogram of the distribution of region mass as a function of region length and orientation for any given spatial scale (an example of the histogram of region length at one spatial scale is given in Fig. 2.2b).

Because there are positive and negative regions, and these are kept separate, two histograms were created for each parameter of orientation and length, one describing the distribution of positive regions, the other describing the distribution of negative regions. By generating histograms in this manner at every spatial scale a complete set of histograms of the distribution of region length and orientation mass across spatial scale was constructed. An example of one such histogram is shown in Figure 2.2c. Notice that now, the distribution of region mass as a function of region length (abscissa) and spatial scale (ordinate) is shown as a density plot. The darker the area, the greater the region mass at that scale and region orientation and length.

a



b

c

d

```
FORMAT
 BLOBF-   Area          Mass   Centroid  Sd.(l,w)    Error   Axis -
METRIC
 ORIGIN<   0.00          0.00     0.00  0.00  0.00    0.00    0.00 >
  SCALE<   1.00          1.00     1.00  1.00  1.00    1.00    1.00 >
 STATS
 MEAN<    762.68      -1741.67     3.91  7.41  2.75    1.63    1.16 >
 SD<     6499.78     203735.42    90.10  8.92  7.13    3.11   21.73 >
 MAX<   71058.00    2196549.25   147.44 86.73 79.68   25.19   87.80 >
 MIN<      11.00     -66167.00  -147.54  1.30  0.73    0.03  -83.96 >
 TOTAL<89996.00    -205516.75   461.64 873.87 324.02 192.56137.21>
 DATA
  BLOBD(  24.00         11.95  -147.54   2.02 1.05   25.19   25.901)
BLOBD( 143.00   -18161.00 -141.45   4.93 2.47    0.63   -7.448)
BLOBD(1058.00  2196549.25    1.63  86.73 79.68   0.62  -12.096)
```

Figure 2.1. Operations performed on an image by the MIRAGE (Watt & Morgan, 1985) algorithm. An image (panel a) is first filtered with a Laplacian of Gaussian (LoG) differing in spatial scale (size of the filter), shown in panel b (filtered at a scale of 6 pixels). Spatial scale is defined as the standard deviation of the filter (in pixels). Following this, the filtered image (which now has positve and negative values and a mean value of zero) is subjected to a threshold procedure which extracts those parts, or positive and negative *regions* of the image which are significantly different from the expected mean value, shown in panel c. The image is then represented as a sentence-based symbolic description, shown in panel d - MIRAGE.

Figure 2.2. Stages leading to the final description of text images, as histograms. Top panel (a): Regions produced at one spatial scale (LoG filter s.d. = 6 pixels) by the image processing operations described by Fig 2.1. Middle panel (b): A histogram of the distribution of mass of negative regions (dark 'blobs'), shown on ordinate, of the image description shown in (a) as a function of region length, shown on abscissa. Bottom panel (c): Histogram of the distribution of mass of negative regions as a function of their length, as in (b) but one which now shows length (now on the abscissa) across the range of spatial scales (ordinate) shown as a density plot. The greater the region mass at any scale and length the darker the plot.

Because each region found has a particular length and orientation, construction of histograms of region parameters of orientation and length as a function of spatial scale in this way provides a very simple and direct means of a quantitative analysis of the information content of images as represented in the image description by the MIRAGE algorithm. As such, they form an important part of the computational work of this thesis.

2.2 Results

The results are summarised in two sections. The first (Section 2.2.1) is a summary of the image content from a visual inspection of the visual description (a reconstruction of the symbolic representation) of the image by the model. The second (Section 2.2.2) is a summary of the quantitative analysis of the same information content in the image, as provided by the histograms of regions parameters as a function of spatial scale of image analysis. Description of the findings in these ways shows how the representation of information in text images is made explicit by regions which emerge at particular spatial scales, and how this information changes with spatial scale.

2.2.1 Visual description

All three font types produced the same general pattern of results, and within each font type each exemplar produced almost identical patterns of response when, as was the case in this analysis, word and line spacing was set to the same parameters for all 3 fonts. Some small differences in response occurred between fonts at the finest scales of analysis but the present analysis will not be concerned by such small differences in feature representation at this level. The results are therefore given for only one font, Times Roman. The panels of Figure 2.3 shows a reconstruction of the resulting representation of one such text image at a range of spatial scales.

At a very coarse spatial scale (Fig. 2.3h, filter s.d. = 12 pixels, filter width = approximately 4 letters [letter is defined as the height of a lowercase 'x']) there is little information in the image which is represented by the model. Neither positive (light 'blob') or negative (dark 'blob') regions have any definitive structure to them. However, there is a correspondence between the presence and shape of positive regions and "rivers" in the text caused by alignment of word spacing from adjacent lines. More interestingly, vertical positive regions corresponding to each sentence boundary have appeared.

At the next spatial scale (Fig. 2.3g, filter s.d. = 8 pixels, filter width = approximately 2.5 letters) some negative regions correspond to whole words but others do not. However, note that there is now an interesting feature of the horizontal positive regions, which correspond to line spacing.

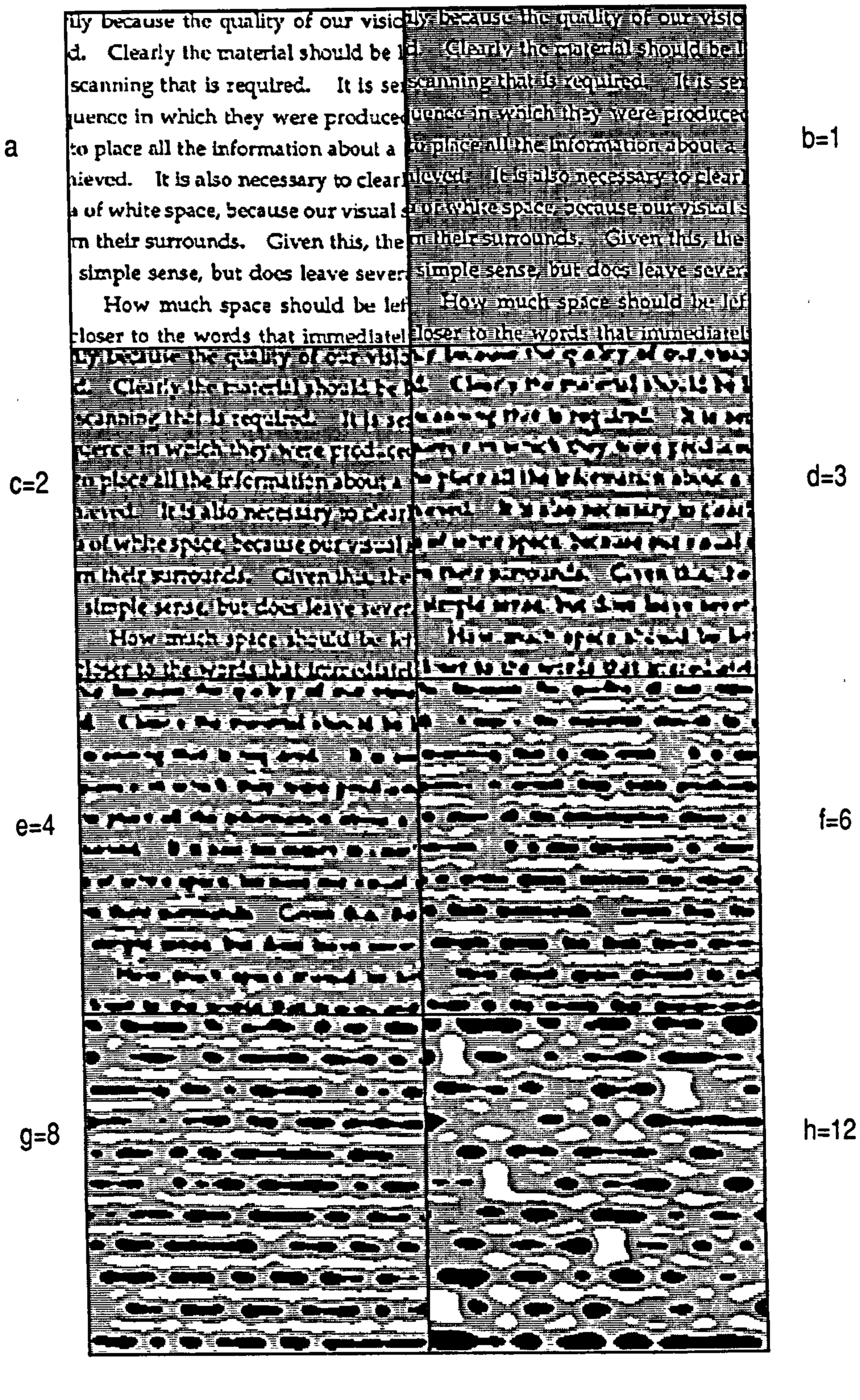


Figure 2.3. Visual description of the image as represented by the MIRAGE model of vision. Negative regions of image extracted by model are shown as dark 'blobs'. Positive regions are shown as light 'blobs'. Panel 'a' shows original image. Features extracted are shown as a function of spatial scale, from 1 pixel (filter s.d) in panel 'b' to spatial scale = 12 pixels, in panel 'h'. Number by each panel refers to filter s.d in number of pixels. See text for further detail.

At a slightly finer spatial scale (Fig. 2.3f, filter s.d. = 6 pixels, filter width = approximately 2 letters) there is a switch in the regions of interest. The interest now lies in the negative regions (dark 'blobs'), which have a definite horizontal shape, with the length of each region corresponding to the length of the group of letters making each word in that position in the text image. Positive regions (light 'blobs') are also horizontal and correspond to line spacing. These positive regions tend to break up at places in the text where ascenders or descenders, and especially both, from adjacent lines occur.

Moving again to a finer spatial scale (Fig. 2.3e, filter s.d. = 4 pixels, filter width = approximately 1.25 letters) the interest still lies in the negative regions, which still tend to horizontal. Not all word lengths appear at this scale; there has been some *un-grouping* of whole words in the image into regions having shapes corresponding to one or more letters. Interestingly, the shape of these regions is also determined by ascenders and descenders, especially when they occur in word boundary positions.

Positive regions are both horizontal and vertical, and of those that are vertical, a small number correspond to word breaks. The next spatial scale (Fig. 2.3d, filter s.d. = 3 pixels, filter width = approximately 1 letter) produces a switch in interest suddenly to the positive regions. At each word boundary, there appears a vertical positive region (light 'blob'). The negative regions (dark 'blobs') at this scale tend to correspond to individual letter envelopes, the length and orientation of which depend on the particular letter in that position in the image. A further process of un-grouping can be observed as the shape of many of the individual ascenders, descenders and small letters is made explicit.

At the next to finest scale (Fig. 2.3c, filter s.d. = 2 pixels, filter width = approximately 0.6 letter), the regions of interest are still the negative regions, which have specific letter shapes, but remain 'blob' like.

At the finest scale (Fig. 2.3b, filter s.d. = 1 pixel, filter width = approximately 0.3 letter) the interest once again lies in the negative regions which have the same shape as letters, but at this scale letter features are also represented. Positive regions are mainly vertical and occur at character spacing. At this spatial scale, the text visually appears little different from the original, unfiltered, version shown in Figure 2.3a.

2.2.2 Quantitative analysis

A quantitative analysis of the different information made explicit across spatial scales of analysis was performed by the construction of histograms of the distribution of the mass of region orientation and length, as a function of spatial scale and sign of response (positive and negative), as described in Section 2.1.3. A summary of this analysis is given below.

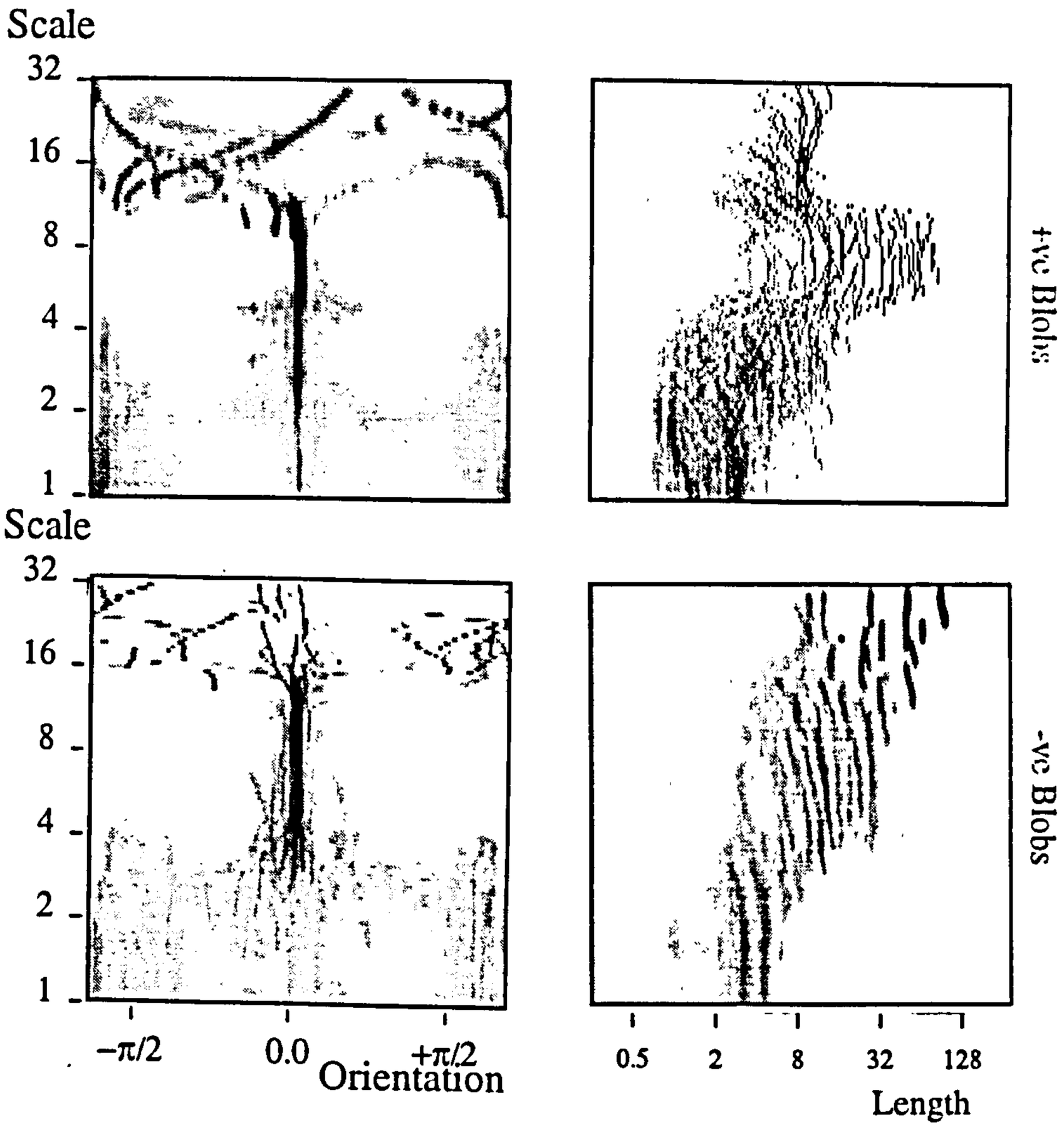


Figure 2.4. Histograms of the distribution of region (blob) mass as a function of spatial scale (filter s.d. in pixels), shown on ordinate, and region orientation (left panels) and region length (right panels). Distribution of region mass is shown as a density plot. The darker any point in the histogram, the greater the region mass at that scale and orientation or length. Two region orientation and length histograms are shown. The top two are positive regions (light blobs). The bottom two are negative regions (dark blobs). See text for further details.

Region orientation: Figure 2.4a shows the distribution of region orientation as a function of spatial scale. Examination of the orientation histograms reveals several findings which warrant comment.

The first thing to notice is that there is a definite pattern to the histograms which is determined by four bands of region mass across spatial scales, identified in the histograms as dark areas. Each of these bands is itself determined by how information about the text is made explicit at different spatial scales. Three of these bands are quite distinct, one is less easily identified until the difference between the positive and negative region distribution is examined. At the finest spatial scales (scale = 1–3 pixels) there is a distinct band in which both positive and negative regions are predominantly vertical or near vertical. From the observation of how regions corresponded to text features made from Figure 2.3 it is possible to conclude that this negative region distribution corresponds to the representation of information about letter strokes, which are predominantly vertical.

The next band (spatial scales 3-6 pixels) is best identified by examination of the *difference* in the distribution of region mass between positive and negative regions. This reveals a distribution of positive regions (light blobs), almost all vertical, with most concentrated within a narrow scale centred at spatial scale = 3 pixels. This corresponds to the emergence of the positive vertical regions at each word break. The negative regions (dark blobs) in this band are distributed across all orientations, but with a greatest frequency of horizontal regions. This corresponds to the representation of partial word length information.

A third band (spatial scales 6-12 pixels) is readily identified by a sharp switch in region distribution from that observed in the first band at scales 1-2 pixels. Both positive and negative regions are exclusively horizontal. This corresponds to whole word lengths in the case of negative regions, and line spacing in the case of positive regions.

Finally, a fourth band at the coarsest spatial scale: 12 pixels and above, can be identified by a switch again to both horizontal and vertical regions in both the positive and negative histograms. Note also that the frequency of horizontal negative regions extends over a wider range of spatial scales.

The positive regions correspond to 'rivers' being made explicit as a result of accidental alignment of word spaces, and to line spacing. The negative regions correspond to the grouping of several words into entire or almost entire line length regions due to the accidental alignment of ascenders and descenders.

Region length: The positive and negative response histograms for length, shown in Figure 2.4b also reveal a number of 'bands' of the distribution of region mass as a function of the

same range of spatial scales. However, they are generally not as clearly defined as those for orientation.

At the finest spatial scales (1–2 pixels) there is a bi-modal distribution of region length, one short, one even shorter, occurring in both the positive and negative histograms. This corresponds to the representation of tall (ascending and descending) and small ('x' height) letters respectively.

In the next, intermediate, band (spatial scales 3–6 pixels) the negative region distribution remains bi-modal, but regions decrease slightly in length. This corresponds to a change in the 'reliability' of the representation of letter features, especially the length of ascending and descending letters. Conversely, there is no longer a bi-modal distribution of positive regions. The regions which remain increase in length at a scale centred around 3 pixels. This corresponds to the representation of information about word breaks made available by the vertical positive regions.

At coarser spatial scales (6–12 pixels) a band can be identified by a switch in the distribution of region length from short to very much longer. This distribution of region mass is centred at a spatial scale of 6 pixels, with few regions of these lengths occurring at spatial scales of 4 or above 8 pixels. The negative regions are organised into 'stripes' of discrete lengths. These regions correspond to whole word lengths. The distribution of length of these regions being determined by the distribution of word lengths in the text. Positive region length distribution is found to be longer than that observed for negative regions. This is because they correspond to line spacing, which is longer than word length, often only being broken up by alignment of ascending and descending letters.

At the coarsest spatial scales (12 pixels and above) negative regions become less frequent and longer with increasing spatial scale. Positive regions switch to much shorter lengths. The longer negative regions correspond to the grouping of words into whole or near whole lines of text. The shorter regions correspond to 'rivers' emerging when ascenders and descenders are accidentally aligned. The positive regions correspond, similarly, to the white spacing between lines, and to the emergence of "rivers" in the text which may result when word spacings are accidentally aligned.

2.3 Discussion

There is a close match between the information made available in a page of text and the visual information plausibly required to process text during reading. At coarse spatial scales, sentence boundaries were extracted, which narratively, are common and sensible places to start reading. At the next scale down, the orientation and length of lines was

represented. Besides making explicit the location of each line, this could help determine the magnification of the text in order to scale the whole process and possibly select the initial largest useful spatial filter size.

The next step in the reading process would be expected to be to segment the page into words. At the next spatial scale (filter s.d. = 6 pixels) features were extracted making explicit whole words and their length in the negative regions. This occurred as the line-level regions were un-grouped as the largest filter size decreased. At a slightly finer scale, the positive regions occurring at word boundaries emerged. These too might be used as cues to where words begin and end. Also at this scale, a different set of features were made explicit about whole word shape, particularly about the relative location of ascenders and descenders. This is reminiscent of the type of "supraletter" features which Monk & Hulme (1983) proposed might be used for word recognition.

If a word cannot be identified on the basis of information contained at this scale, the identification and position of each letter in the word needs to be known. This information was represented at the finest spatial scales by the model, which extracted letter shaped regions, and finally, letter-feature regions.

An important feature of the way in which the model makes particular aspects of the text 'visible' for the processing of text is that of the grouping, or more precisely the un-grouping, of elements. Line-level regions were un-grouped into word length regions, word regions were then un-grouped into letter regions, and letter-level regions were finally un-grouped into letter feature regions. This un-grouping was defined by the operations of MIRAGE on the basis of a coarse-to-fine spatial scale of visual analysis, as the largest active filter is progressively switched out.

Two other points are worth noting. The first is that the distinctive image structures forming each of the 'bands' generally covers 1 octave of spatial scale. This agrees with other, psychophysical, findings concerning the size of the range of filters in the human visual system and their bandwidth (Philips & Wilson, 1984; Georgeson & Harris, 1984). It may also provide one explanation for Bock, Monk & Hulme's (1993) suggestion that the grouping of letters into words in reading may depend on image features differing by less than one octave.

The second point is that the analysis shows that different levels of structure are made available at different scales. If Watt's coarse-to-fine analysis occurs in the visual system, then it is interesting that the order in which the different levels of structure revealed is the same as that plausibly required for reading.

Finally, note that typographical arrangements are very likely to determine what structures are made available at particular spatial scales. The reason for this is as follows. As outlined in Chapter 1, the proximity of one element to another will be a determining factor in what features of text are represented and made explicit as regions at particular spatial scales. These features of text are determined by the proximity–or spacing–of lines, words and characters, which are determined by typographical practice. For instance, if word spacing is too narrow, information about word length contained in the negative regions (dark 'blobs') at the coarse spatial scale (filter s.d. = 6 pixels) and word breaks contained in the positive regions at the intermediate scale (filter s.d. = 3 pixels) might no longer be represented. This might be expected to affect the visual processing of text and consequently slow reading as the reader has to use a finer level of visual detail at which appropriate information, for example the individual letters, would still be contained. Thus, a possible explanation for typographical effects in terms of the nature of early visual processing begins to emerge.

The purpose of the rest of thesis is to discover whether the information made available by early vision and used in the processing of text during reading might be the same as that by the model when applied to pages of text, as identified here. In this respect, the findings of this computational analysis of text allow the generation of a number of hypotheses based on the predictions of the model, concerning how human visual processing should produce specific behaviour under specific conditions. These are stated below, and are tested out in subsequent chapters.

Hypothesis 1: Text processing performance should be spatial scale dependent.

Information suggested here to be important for specific text processing tasks, such as word segmentation, letter position identification or sentence boundary location was made available at different spatial scales by the model. Human visual processing of text should be similarly dependent upon different spatial scales, and moreover, be dependent on the same, or similar, spatial scales as those of the model.

Hypothesis 2: The sensitivity of visual text processing behaviour to changes in physical parameters of text should be explicable in terms of changes in features represented by the model as a function of the same changes to text.

Performance of a text processing task should vary in a manner which is predicted by the model in terms of changes to the information hypothesised to be important for that task, and at the same spatial scale.

Hypothesis 3: The visual processing of text should be consistent with a time-course of visual processing which is predicted by a coarse-to-fine spatial scale of visual analysis.

If the visual information used to process text is distributed across spatial scale, then different text processing tasks requiring spatial position to be calculated which have different time-courses of processing will require different durations for a given level of performance which are predicted on the basis of a coarse-to-fine spatial scale of visual analysis.

3

Text Processing and Spatial Scales of Visual Analysis

The computational analysis reported in Chapter 2 demonstrated how a model of visual processing extracted a set of features, or 'regions,' in text images at a range of different spatial scales. If this modelled description of information contained in a page of text is similar to, or indeed the same as, the human visual representation of information in a text image it should be possible to devise a number of psychophysical tasks to show that the predictions of the model are consistent with human visual processing of text performance.

This chapter reports an initial series of experiments which sought to determine the first prediction of the model which needs to be tested: that the visual processing of text should be spatial scale dependent. Three experiments are described which aimed to show two things. First, whether the human visual system has access to information contained in different spatial scales which is used to process text. Second, whether any spatial scale dependency of visual processing behaviour is predicted by the model of vision.

3.1 General methods

A primary and important problem to overcome in conducting any experiment examining the visual processing of text is that of devising a suitable experimental task to capture only the *visual* aspects of text processing. Chapter 1 argued for two important requirements of a study if the problems associated with existing methodologies are to be avoided. The first requirement was that the methodology must capture the visual aspects of text processing only, and must avoid the possibility of confounding visual and linguistic components of text processing which could determine behaviour. The second requirement was the need to examine individual text processing tasks.

Psychophysics is a good way of isolating a physical attribute of an image and measuring how the information in that image is used by the visual system, so designing suitable

psychophysical tasks would be expected to fulfil the first requirement. However, psychophysical procedures have usually been confined to measuring the basic properties of the visual system such as acuity, contrast detection or curvature. It has been necessary here to extend this approach to examine the visual processing of text.

Three psychophysical visual processing tasks were carefully designed to represent the type of text processing tasks of the sort required during the processing of text in reading, and isolate the *visual* processing aspects of each of these tasks. The tasks seem somewhat odd at first sight because they bear little resemblance to reading itself. However, to reiterate, what is required is a measure of the visual processing of individual text processing tasks of the type performed during reading, and not a general measure of the product of all the components of the reading process.

The three text processing tasks were 'word segmentation', 'letter position identification' and 'sentence boundary location'. In the word segmentation experiment (Experiment 1), the subject's task was to discriminate between two text images which differed in the mean length (defined by the number of letters) of words contained in the page image. This task only requires the subject to be able to segment the words on the page in order to perform the task correctly.

In the letter position identification experiment (Experiment 2), the subject's task was to discriminate between two text images, one of which contained ascenders or descenders at particular positions in each word, the other contained ascenders or descenders in any position in each word: that is, identify letter position.

In the sentence boundary location experiment (Experiment 3), the subject's task was to discriminate between two text images which differed only in the number of sentences each one contained. This task required only sentence boundaries and their number to be located.

Within the word segmentation task (Experiment 1), an additional issue was also explored. The probability of ascenders and descenders occurring at word boundary positions is greater, in English, than that for small letters (Walker, 1987). It is worthwhile to speculate about whether this finding has any basis in the visual processing of text. In particular, whether the physical characteristics of ascenders and descenders (*i.e.* having elongated vertical strokes) may provide some kind of early visual processing aid for word segmentation. Any evidence that word segmentation performance is better when words have ascenders or descenders at boundary positions would provide some support for this suggestion.

3.1.1 Text

(i) Text characteristics. It was important to constrain several characteristics of the text. There were 3 constraints: letter type (ascenders, descenders and small), letter position (where about in the word any of the letters could occur) and word length distribution (used as a cue to perform the psychophysical task in Experiment 1). The details of each of these constraints (the statistics of the text for each experiment) are provided in the appropriate Section for each experiment.

(ii) Text image digitisation. Images were created by scanning pages of text at a resolution of 300 dpi (dots per inch) and 256 grey levels using a Hewlett-Packard ScanJet digital scanner and selecting 300 pixel square samples of the digitised pages. Pixel values were scaled such that the mean greylevel was zero, and the standard deviation of the range of greylevels was 16.

(iii) Text image filtering. The resulting images were band-passed filtered using filters whose functions were Laplacian of Gaussians (LoG). Images were filtered using a range of 8 filter scales with space constants of 1, 2, 3, 4, 6, 8, 12, and 16 pixels. This scale term refers to the standard deviation of the filter. Examples of the appearance of the text images at the end of this image processing are shown in Figure 3.1.

(iv) Visual masking noise. Visual noise, which was to be arithmetically added to the images, was generated from a set of the original, unfiltered images by subjecting them to a Fast Fourier Transform (FFT), phase randomisation and inverse FFT. This process produced an image with a spatial frequency and power spectrum which was identical to the text image but one which contained no phase information. These images were also filtered using the same range of LoGs, and normalised to the appropriate display range.

3.1.2 Technique

The technique of adding bandpassed noise to a stimulus has been used previously as a method of estimating the 'visibility' or the reliability of the information contained in a narrow spatial band or scale (*e.g.* Stromeyer & Julez, 1972; Julez, 1980; Nothdurft, 1991). The same basic rationale is adopted here, but the present technique makes an important departure from that employed in previous studies using this technique. That is that both the noise *and* the text images were limited in spatial scale. This was because there is evidence to suggest that there is an interaction (a non-independence) of spatial scales at supra-contrast threshold levels, (*e.g.*, Henning, Hertz & Broadbent, 1975; Jamar & Koenderinck, 1985).

Thus, in the case of a spatially extended (spatial frequency, that is) image and spatially limited noise, the effect of the noise cannot be estimated reliably (Pavel *et al.*, 1987) because there may be information contained at one or more scales which is outside the masking effects of the noise and is therefore left to influence the task threshold.

This technique produces an estimate of a signal-to-noise ratio (SNR) required for a threshold level of performance on a text processing task at a number of different spatial scales. The amount of signal level (SL) required to reach task threshold provided an estimate of the 'visibility' of the information used to perform the task as a function of spatial scale. That is, the more sensitive the subject's vision is to the information in the image at a particular spatial scale, the lower the SNR (and thus the lower the SL) is required to extract that information and perform the task.

3.1.3 Procedure

In each experiment a two alternative forced choice (2AFC) procedure was used. The phase randomised visual noise was added to the stimulus with SNRs varying from trial to trial. An adaptive method of constant stimuli (APE; Watt and Andrews, 1981) was used to select representative SNRs on the psychometric function. APE generated a range of stimulus levels between 0–8. These values were used to combine image and noise signals. Thus for an APE level of 8, the full image with no noise added was presented and the subject could perform at 100% correct level. At a value of 0 the full noise signal was presented with no image at which point the task was impossible. Probit analysis applied to the data determined the standard deviation of the best fitting cumulative Gaussian[4]. This value, which corresponds to the probability of 83% correct point, was defined as the threshold.

Two adjacent text images having a mean luminance of 170 cd/m^2 were presented simultaneously for 3000msec either side of the centre of a CRT display against a grey background having a mean luminance of 65 cd/m^2. Each text image subtended, at the viewing distance of approximately 50cm, a visual angle of approximately 6°. Ambient illumination varied from day-to-day, but was typically 200 Lux.

Subjects were given extensive practice before data collection began. In each run at each spatial scale, 64 presentations were made to the subject. The subject responded by pressing a mouse button indicating which one of the two text images (*i.e.* left or right) differed in the size of the particular cue (see each individual experiment for what each cue was).

---

[4] Data from each trial were fitted to this standard form of psychometric function and the goodness-of-fit estimated using a chi-square. This function provided a good fit to the data on over 90% of trials, but where it did not, data from that trial was rejected.

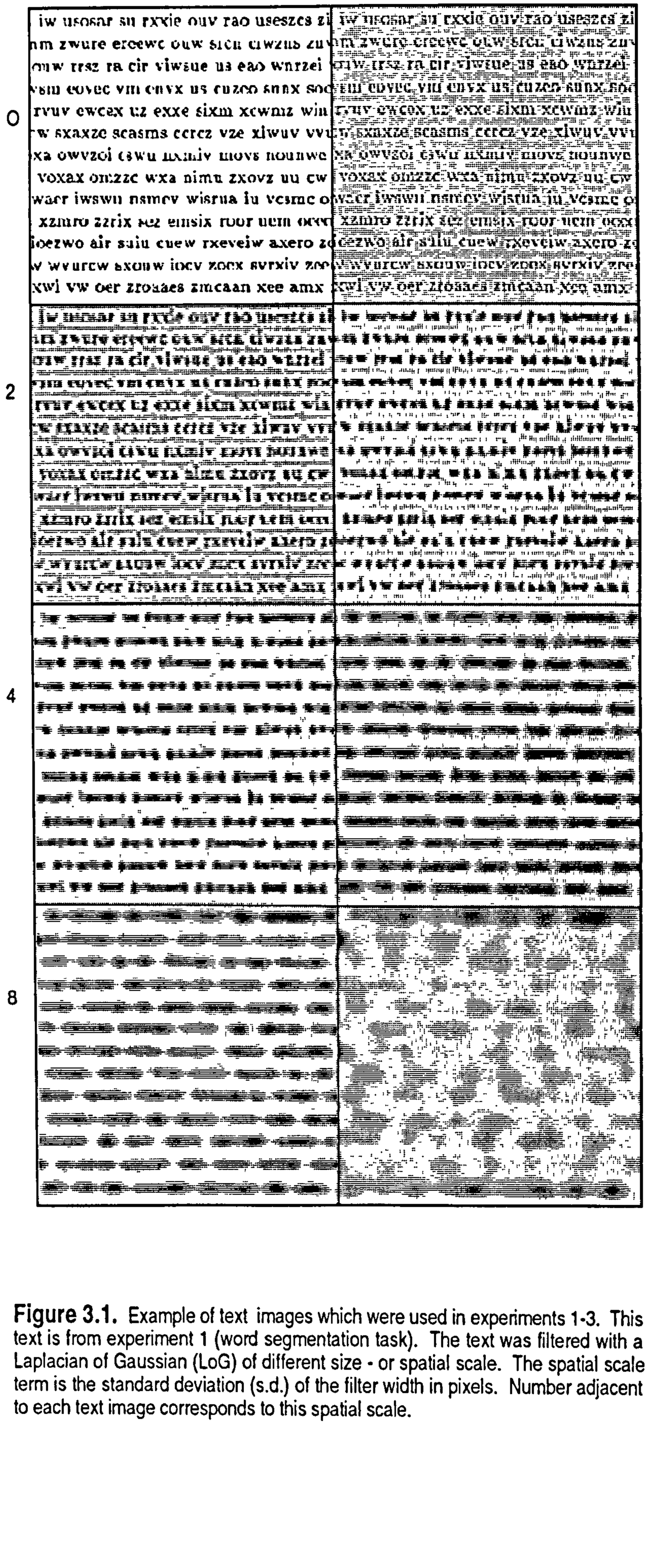


Figure 3.1. Example of text images which were used in experiments 1-3. This text is from experiment 1 (word segmentation task). The text was filtered with a Laplacian of Gaussian (LoG) of different size - or spatial scale. The spatial scale term is the standard deviation (s.d.) of the filter width in pixels. Number adjacent to each text image corresponds to this spatial scale.

The threshold signal level (SL) required to perform the task at any single spatial scale was taken as the mean of the SL determined in each of three such runs.

Images were presented on a 21 inch Trinitron display with a refresh rate of 66Hz, controlled by a Sun 3/80 workstation. The luminance profile of the monitor was linearised using a look-up table.

3.1.4 Subjects

3 subjects participated in all 3 experiments. RAO and PMC were paid participants. The other, SJE, was the author. RAO and PMC were naive as to the purpose of the experiments but had previously participated extensively as psychophysical subjects. Both were emmetropic. SJE was a corrected hypermetrope.

3.2 Experiment 1: Word segmentation

3.2.1 Method

(i) Images. 40 text images (20 with a mean word length of 4 letters; 20 with a mean word length of 7 letters) for each of 3 conditions were generated. The term 'word' in the context of Experiment 1 refers to a letter string of particular characteristics conforming to one of the conditions outlined below. Text was generated from a 'C' program developed (by the author) to generate a page of text with a specified mean word length and standard deviation having an approximation to a Gaussian distribution. Each page of text was printed in 12pt Times Roman from a LaTeX file. Text was page-centred to ensure that word spacing was uniform. The standard deviation of mean word length distribution on each page was 1.5 letters. The range of lengths was between 1 and 11 letters, which, given the difference in mean word length, meant that strictly speaking, the distribution of word length was slightly skewed from a Gaussian.

There were 3 conditions in which word structure differed: 'Boundary', 'Small' and 'Mixed'. In the boundary condition, word boundary letters (first and last letters of each word) were any randomly selected ascender or descender (q, t, y, p, d, f, g, h, j, k, l, b); each non-boundary letter was any randomly selected small letter of the English alphabet (w, e, u, i, o, a, s, z, x, c, v, n, m).

In the small condition, all words were constructed from any, randomly selected, small letters, and in the 'mixed' condition all words were constructed from randomly selected ascenders, descenders and smalls so that any letter had an equal probability of occurring in any position in the letter string. Examples of text from each condition are shown in Figure 3.2.

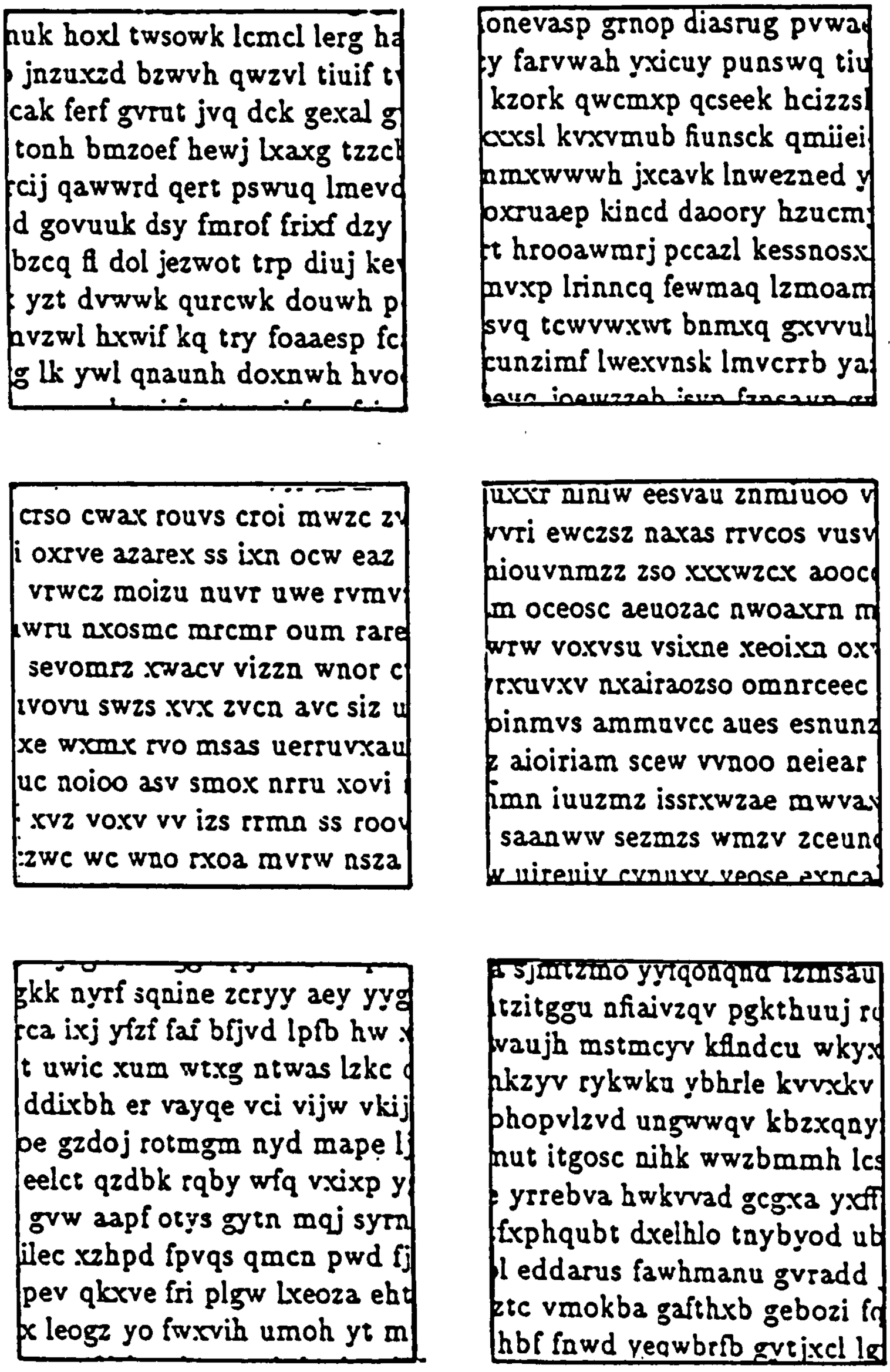


Figure 3.2. Examples of text images used in experiment 1. Panels show the 2 different mean word lengths (left panel = mean 4 letters: *reference*; right panel = mean 7 letters: *cue*) and the 3 different word structure conditions (top = 'boundary'; middle = 'small'; bottom = 'mixed'). Refer to text for more details.

(ii) Procedure. The general technique and procedure used in Experiment 1 followed that described in the General methods Section (3.1). Therefore, only experiment specific details of the procedure are given.

Subjects were presented with two 300x300 pixel text images appearing simultaneously in adjacent positions on the display for 3 seconds. Both text images were displayed at the same spatial scale, but the spatial scale condition was interleaved and so varied randomly on each presentation. Text images had varying levels of noise added to them which was determined by APE (further detail provided in Section 3.1).

The subject's task was to decide which text image had the longest mean word length. Instructions given to subjects were only that they needed to decide which text image had the longest mean word length.

3.2.2 Results

The data of Experiment 1, shown in Figure 3.3a-c, reveals that in all 3 word structure conditions, performance varied systematically with spatial scale of image content. Optimum performance for all 3 subjects occurred at a single spatial scale (filter s.d. = 6 pixels). That is, spatial scale = 6 pixels contained the most useful, or reliable, information for word segmentation. In addition, there was some evidence of a second scale (filter s.d. = 3 pixels) at which some information was contained which facilitated word segmentation performance.

Figure 3.4 shows the comparison of the effect of word structure on word segmentation performance. It can be seen that when words contained ascenders/descenders at word boundary positions ('boundary' condition) the signal level (SL) required to reach task threshold tended to reach a slightly lower minimum than for small and mixed conditions for all 3 subjects within the range of scales = 3–6 pixels.

This result appears to lend some support to the suggestion made earlier that information in the image which was used for word segmentation might be made more reliable by the effects of ascenders or descenders at word boundary positions, facilitating word segmentation. However, an analysis of variance (ANOVA) found that the difference in performance as a function of word structure condition failed to reach statistical significance [$F_{(2,20)} = 1.43$, n.s.]. Care needs to be taken though, in concluding from this analysis that word structure had *no* effect on performance. Performing any inferential statistical analysis, such as an ANOVA, as is typically required in these circumstances, using such a small sample of data imposes what might be considered to be unreasonably strict criteria for determining the reliability of an observed, if albeit small, effect.

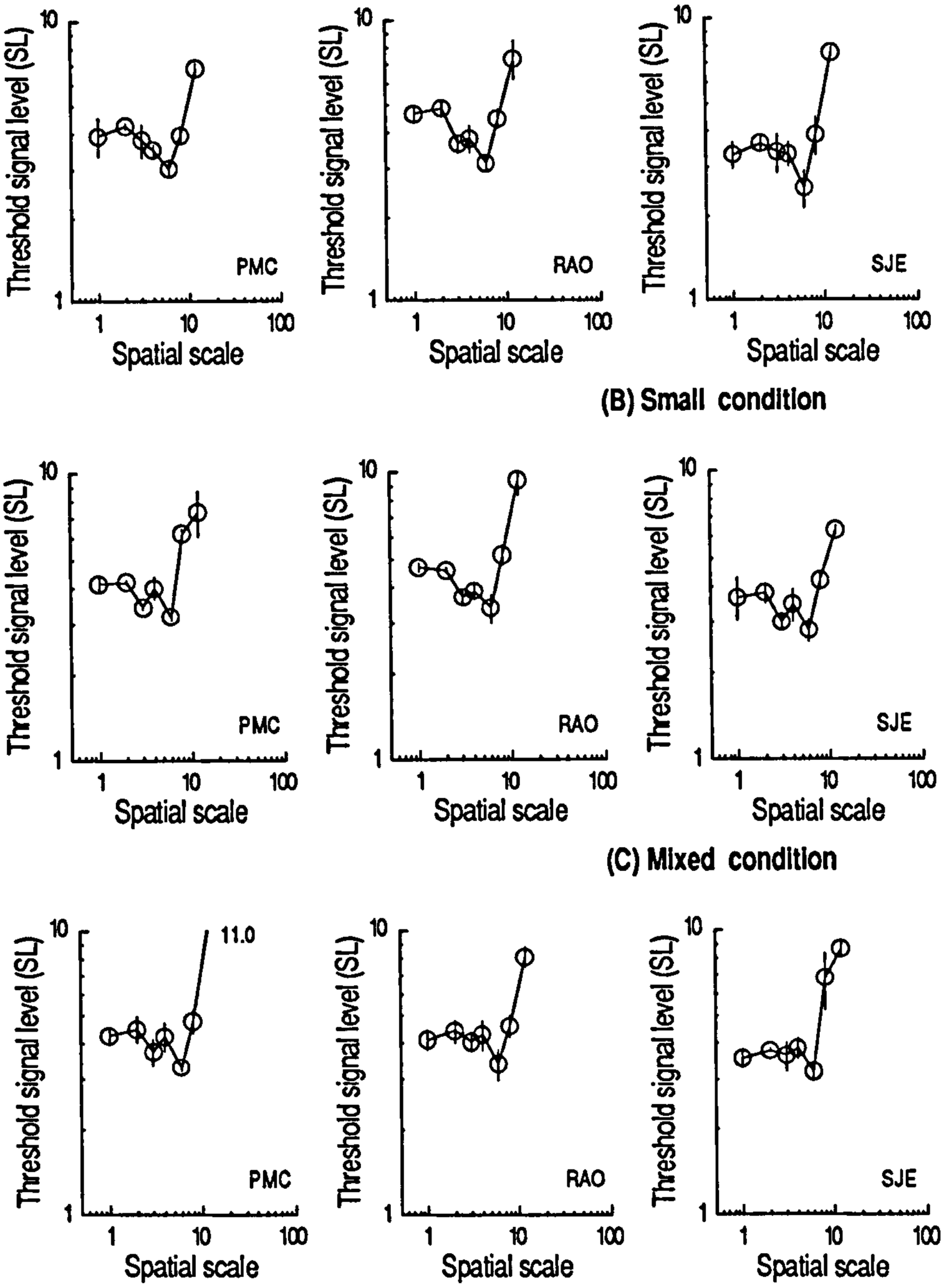


Figure 3.3. Psychophysically determined word segmentation performance as a function of spatial scale of image content and word structure (boundary [A], small [B] and mixed [C]) measured as mean word length estimation performance. Ordinate is the signal level required to reach threshold performance determined by APE (arbitrary units). Abscissa is the spatial scale of the image, where the scale term is the standard deviation of the Laplacian of Gaussian filter width in number of pixels. Each data point is the mean of 3 runs. Each run contained 64 trials. Bars show standard error.

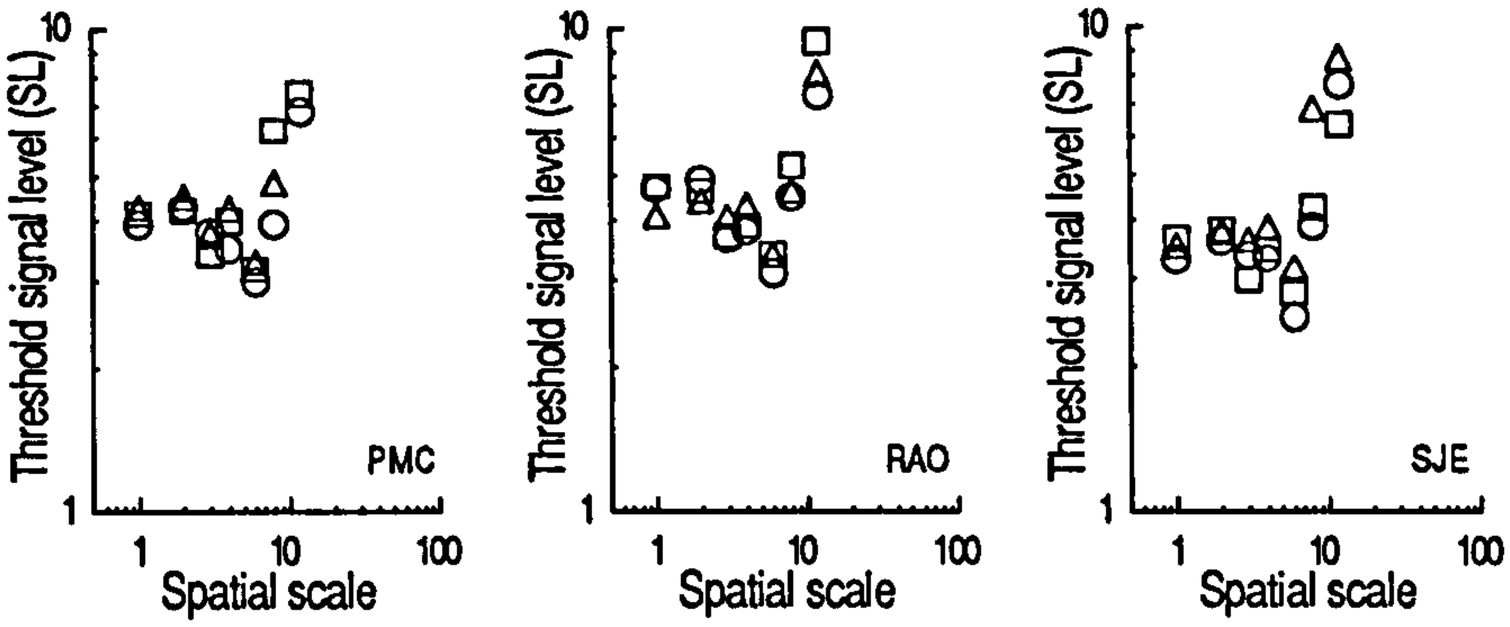


Figure 3.4. A comparison of word structure on word segmentation performance for 3 subjects. Circles = 'boundary' condition; Squares = 'small' condition; Triangles = 'mixed' condition.

Finally, note that the results also show that optimum psychophysical performance appears to occur at the same spatial scales in which information suggested to be of importance for word segmentation was made explicit by the model, as found in the analysis in the previous chapter. Chapter 2 found that at a spatial scale of 6 pixels, negative regions (dark 'blobs') corresponding to whole word shape, particularly length were extracted by the model. It was suggested then that, on the basis of the visual information likely to be required to segment words (as discussed in Chapter 1), information about word length contained at this scale might be useful for this text processing task. This issue is discussed further in Section 3.5.

3.3 Experiment 2: Letter position identification

3.3.1 Method

(i) Text images. 20 text images were generated in which text had ascenders or descenders at every word boundary position, with small letters at every other letter position. A further 20 images were generated which had either ascenders, descenders or small letters at randomly determined word boundary positions, and also at any other non-boundary position. All text images had a mean word length of 4 letters with a standard deviation of 1.5 letters. The text was subject to the image processing operations described in Section 3.1 which provided a set of LoG filtered text images and a set of noise images at a range of spatial scales. The resulting text images had identical characteristics to those of the 'boundary' and 'mixed' conditions of Experiment 1. An example of the text images used in Experiment 2 is shown in Figure 3.5.

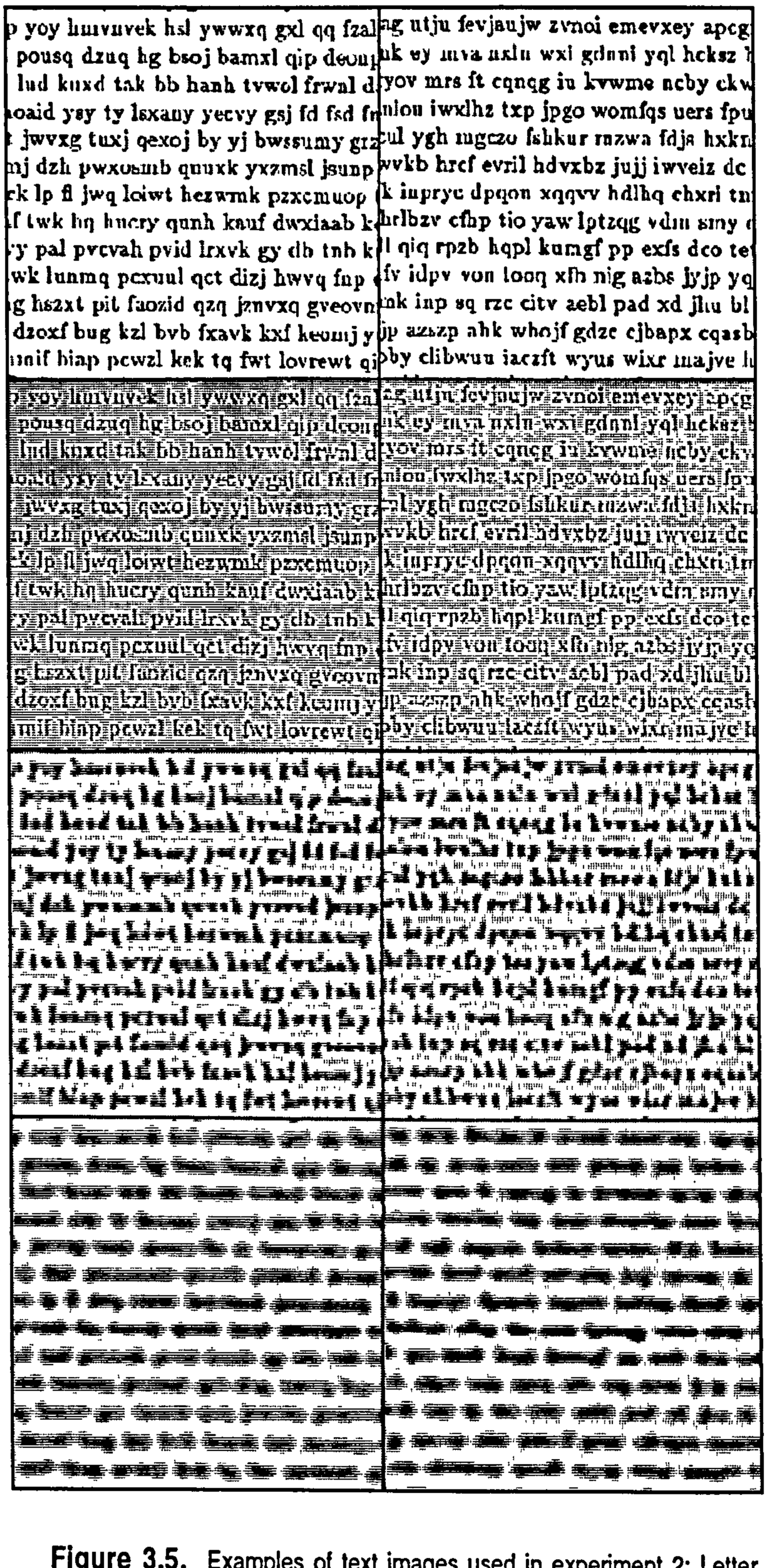


Figure 3.5. Examples of text images used in experiment 2: Letter position identification. Text was filtered with a Laplacian of Gaussian (LoG) differing in spatial scale. The spatial scale term is the standard deviation (s.d.) of the filter width in number of pixels. Number adjacent to panel is spatial scale in pixels, so illustrated is original (C & R) and examples filtered at 1, 3 and 6 pixels. R = Reference stimulus; C = Cue stimulus.

(ii) Procedure. The procedure was the same as that for Experiment 1 (described in Section 3.1 (General methods). The subject's task was to decide which text image had ascenders or descenders in boundary positions only, and which had ascenders, descenders or smalls distributed randomly.

(iii) Subjects. The same 3 subjects that participated in Experiment 1 participated in Experiment 2.

3.3.2 Results

Figure 3.6 shows how letter position identification performance varied as a function of spatial scale. The scale at which optimum performance occurred in this experiment was at a different spatial scale to that observed in Experiment 1. Optimum performance occurred at the finest spatial scale (filter s.d. = 1 pixel), observed by the threshold SL minimum. Performance decreased with spatial scale to the point at which above spatial scale = 6 pixels, it was no longer possible to correctly identify letter position.

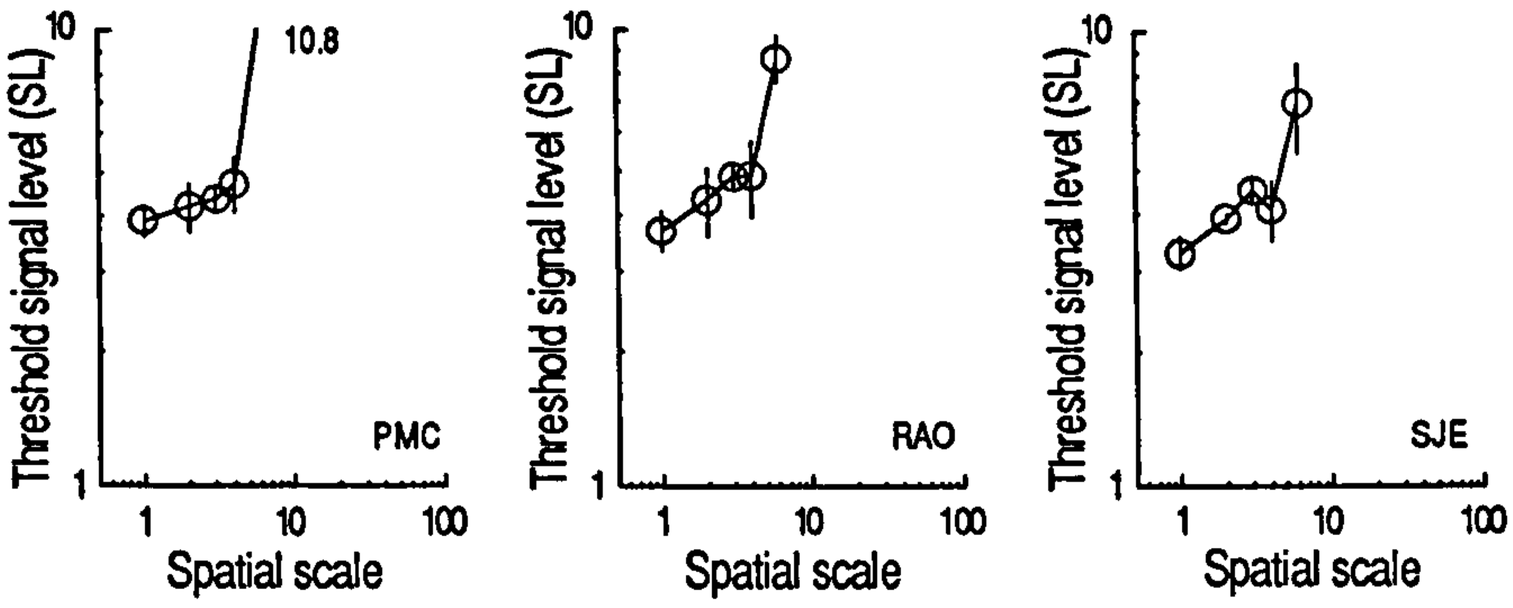


Figure 3.6. Letter position identification performance as a function of spatial scale of image content. Threshold signal level (SL) = signal level required to reach task threshold (arbitrary units). Spatial scale is the bandwidth of the image LoG filtered where the scale term is the standard deviation of the filter in number of pixels. Each data point is the mean of 6 runs. Each run contained 64 trials. Bars show standard error.

Within the range of spatial scales = 1–4 pixels performance did not decline rapidly, suggesting that although the most reliable information in the image was contained at very fine spatial scales, there was some information contained at coarser scales, up to a scale of 4 pixels which still allowed the task to be performed. It is possible that letter position identification is possible at such relatively coarse scales from some type of word envelope information which preserves supraletter features such as word shape.

In this respect, it is worth noting again that the spatial scales at which best performance occurred were the same as those at which features suggested to be useful for identifying letter position were made available by the model in Computational Analysis 1 (Chapter 2). It may be remembered that regions corresponding to letters were extracted at the finest spatial scale by the model (1–2 pixels). In addition, some partial information about letter shape, in particular whether the letter was either a small, ascender, or descender was also made available at slightly coarser spatial scales (3–4 pixels).

3.4 Experiment 3: Sentence boundary location

3.4.1 Method

(i) Text images. Two sets of 20 text images were constructed from text passages taken from several sources (newspapers, journal articles, encyclopaedia) and printed in 12pt Times Roman using the same method as that of Experiments 1 and 2. One set of text images contained 3 sentences, the other set contained 6 sentences. Sentence spacing was set to twice normal word spacing. An example of the text used in Experiment 3 is shown in Figure 3.7.

(ii) Procedure . Experimental procedure was the same as that described in Section 3.1. The subject's task on each trial was to decide which text image contained the most sentences.

3.4.2 Results

Figure 3.8 shows how sentence boundary location varied with spatial scale of image content. It can be seen that optimum performance for all 3 subjects occurred at spatial scale = 8 pixels, indicated by the SL minimum required to reach the threshold level of performance. Again, note that the scale at which optimum performance occurred is different to that at which optimum performance occurred in Experiments 1 and 2.

Unlike Experiments 1 and 2, the spatial scale at which optimum psychophysical performance occurred here was not at the same spatial scale in which information suggested to be important for this text processing task was made available by the model, as observed in Computational Analysis 1, in Chapter 2. In Computational Analysis 1, regions corresponding to sentence breaks which were suggested to be of importance for sentence boundary location appeared at a coarser spatial scale of 12 pixels (filter s.d.).

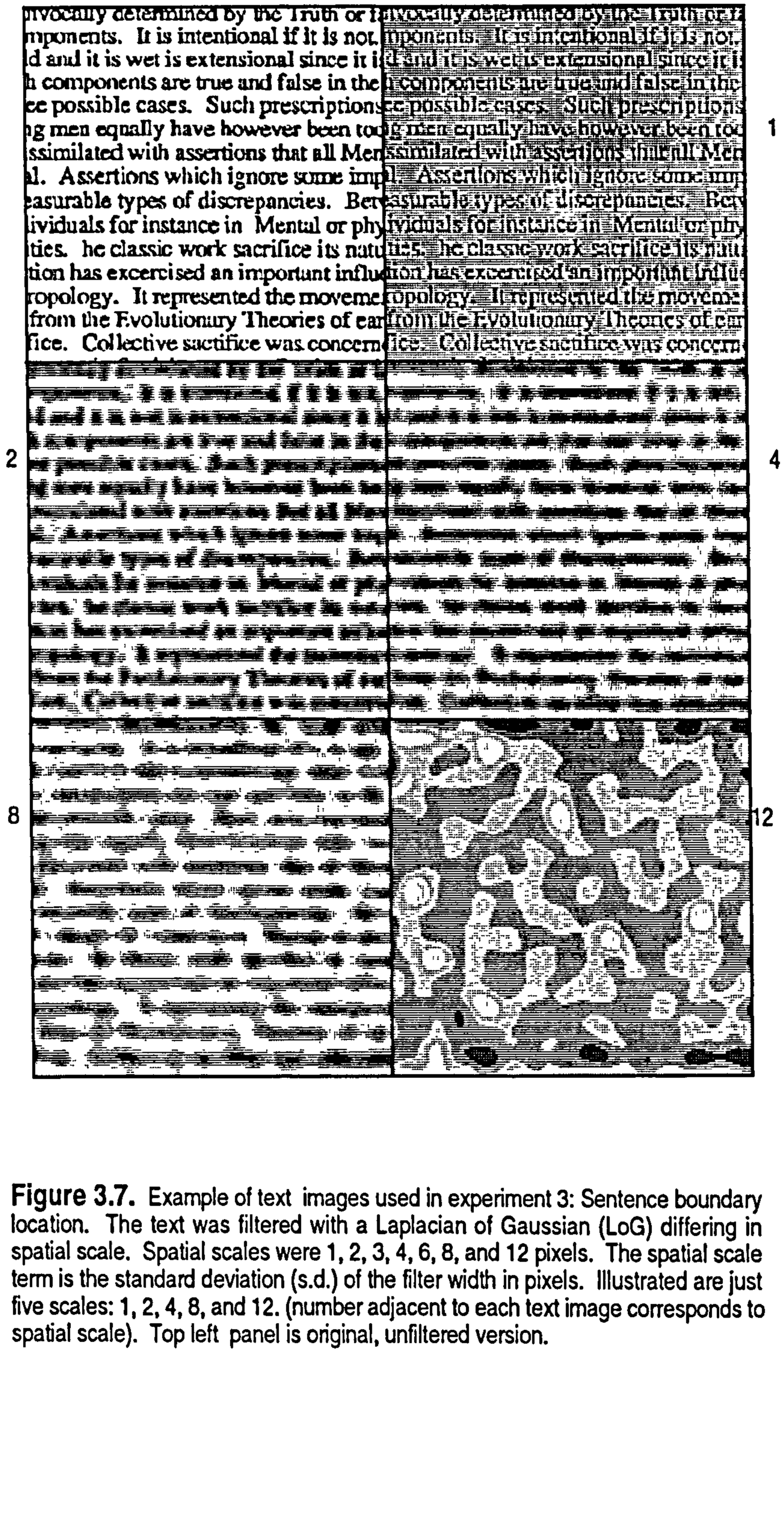


Figure 3.7. Example of text images used in experiment 3: Sentence boundary location. The text was filtered with a Laplacian of Gaussian (LoG) differing in spatial scale. Spatial scales were 1, 2, 3, 4, 6, 8, and 12 pixels. The spatial scale term is the standard deviation (s.d.) of the filter width in pixels. Illustrated are just five scales: 1, 2, 4, 8, and 12. (number adjacent to each text image corresponds to spatial scale). Top left panel is original, unfiltered version.

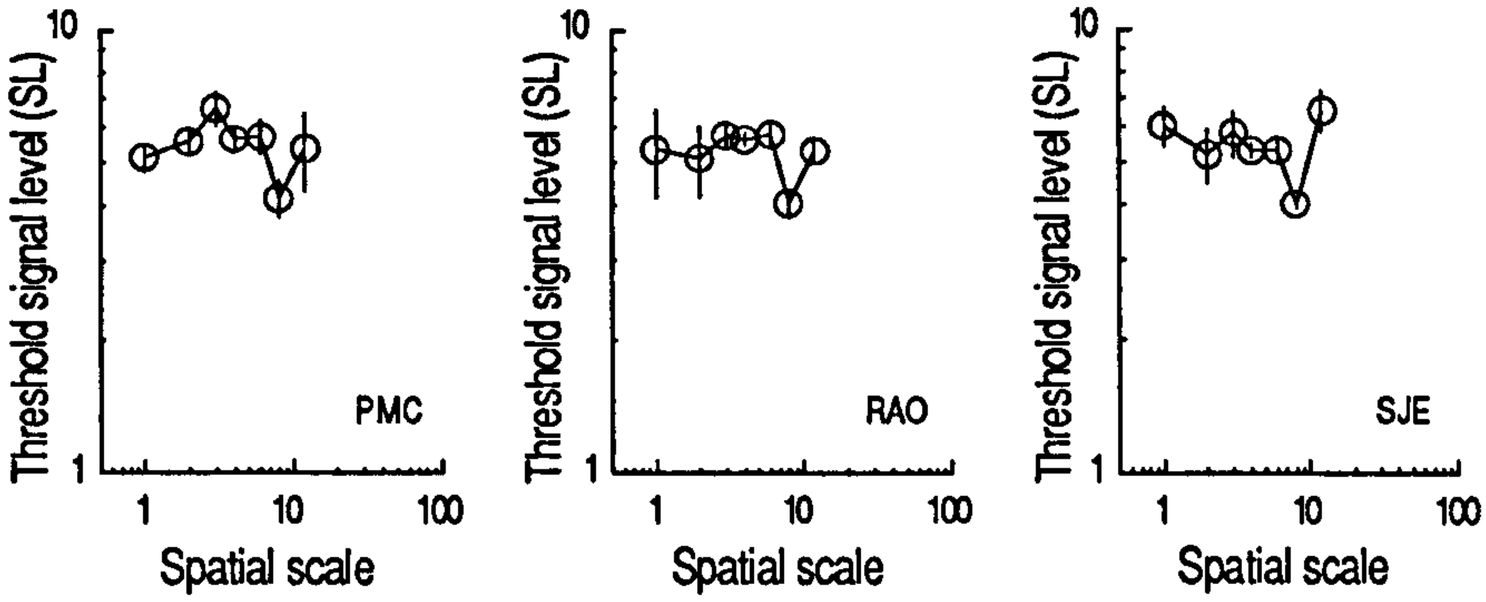


Figure 3.8. Sentence boundary location performance as a function of spatial scale of image content. Threshold signal level (SL) = signal level required to reach task threshold (arbitrary units) is shown on ordinate. Spatial scale is the bandwidth of the image LoG filtered where the scale term is the standard deviation of the filter in number of pixels. Data points are the mean of 6 runs of 64 trials per run.

3.5 Discussion

Experiments 1–3 showed that optimum performance in each of the text processing experiments occurred at a different spatial scale. In Experiment 1 (word segmentation) this was at spatial scale = 6 pixels (filter s.d.), with some evidence that information contained at spatial scale = 3–4 pixels was also more useful than that contained at other scales in performing the task. In Experiment 2 (letter position identification) performance was best at finer spatial scales, optimum performance occurring at scale = 1 pixel. In Experiment 3 (sentence boundary location) optimum performance occurred at the coarse spatial scale of 8 pixels.

Furthermore, initial comparison with Computational Analysis 1 (Chapter 2) showed that the spatial scale at which optimum performance occurred for each task was at the same, or similar, spatial scale at which features capable of supporting each task were extracted by the model in Computational Analysis 1. The findings of this initial series of experiments, that optimum text processing performance occurred at different spatial scales for different tasks, therefore provides some support for *Hypothesis 1:* that the visual processing of text should spatial scale dependent.

The results of Experiment 3 showed the only real deviation from the correspondence between the model and the psychophysical data. Comparison of the text used in Experiment 3 and Computational Analysis 1 reveals that each had slightly different sentence spacing: Experiment 3 text was slightly smaller (3pts vs. 4pts). This is may provide an explanation for the difference between psychophysical performance and the findings of

Computational Analysis 1, given that Watt (1993) showed that the information represented by MIRAGE varied according to the particular typographical arrangement.

It is interesting to note that the psychophysical data showed that performance on each of the tasks was best at a scale which, on the basis of a coarse-to-fine spatial scale of visual analysis, suggests that the information was made available in the order it might plausibly be expected to be required for reading (see Section 1.2). That is, the reader might need to find sentences first, then segment words, and finally, identify letter position in each of the segmented words. This concurs with the general findings of Computational Analysis 1.

Furthermore, the findings of Experiment 1, while not supported by statistical analysis, did lend some support to the suggestion that the information contained in the image and extracted by the visual system to perform word segmentation was affected by word structure. Further support for this possibility, in conditions in which an effect might be more appropriately expected, is sought in Experiment 4 (next Chapter).

While comparison of the results of Experiments 1–3 and Computational Analysis 1 is interesting, it does, however, have limited utility. It is necessary to provide a more appropriate and direct comparison of the psychophysical data obtained here with the behaviour of the model under comparable circumstances. At this stage, there is no clear way the model can be used to account for these particular experimental findings for two reasons. The first is that the text used had minor differences to that used in the computational analysis. Second, and more importantly, a method is required which provides a quantitative analysis to establish whether the information subjects used in processing the text (*i.e.* in making the actual psychophysical judgements) can be described by the model. This was investigated in Computational Analysis 2.

3.6 Computational Analysis 2

The aim of Computational Analysis 2 was to provide an initial measure of the ability of the model of vision to describe the pattern of *psychophysically determined* spatial scale dependency of one of the text processing tasks: word segmentation. Details of this method used to achieve this, and the findings of the analysis, are provided below.

3.6.1 General method

The set of image processing operations performed in this analysis was basically the same as those of the previous computational analysis, details of which are given in the method section of Chapter 2. However, this analysis contains a single, but important, departure from the previous one. In this analysis it was necessary to produce a final description of the

visual representation of features in the text images that the model would make available, from which the information available to make *psychophysical judgements* could be modelled. Furthermore, it was necessary to do this in a manner which provided a quantitative analysis of this information. This was achieved by obtaining a metric of the model's *discriminability* of text images by a series of operations (details given next) which produced a set of histograms containing only those regions in which the model found a difference between the 2 image sets used to perform the psychophysical discriminations in Experiment 1. The process essentially makes explicit the features the model is able to extract to make the same discrimination as that required by the subjects in performing the word segmentation task.

(i) Histogram construction. Because of the computationally demanding nature of this type of analysis it has been confined to an examination of Experiment 1: word segmentation.

Half of the total sample of text images of each word structure type (boundary, small and mixed) used in Experiment 1 were randomly selected. Half of the images selected were from the reference set (mean word length=4 letters) and the other half were from the cue set (mean word length =7 letters). The same image processing operations performed in Computational Analysis 1 were the performed on each text image, viz. (i) a spatial convolution with a range of Laplacian of Gaussian filters; (ii) a series of operations performed on the resulting filtered image at each spatial scale in which the regions or 'blobs' of the resulting images were found for both positive and negative signs of response (see Section 2.2); (iii) the construction of a set of histograms of the distribution of the mass of these regions according to the parameters of orientation and length (see Chapter 2 for further detail).

It was the following subsequent steps which departed from the method of the previous analysis. The aim of these computations was to model the psychophysical discriminability–in other words, the perceptual variability or sensitivity–of the text image sets. Thus, features of the image sets which were very variable contributed less to the final analysis than features with little or no variance in the sets. This was justified on the basis that the more variable some possible cue or information is the less reliable it is as a perceptual cue. The computational procedures employed to achieve this are familiar as the formula for a *t*-test statistic. Indeed, the same basic rationale lies behind this analysis.

The first stage was to compute a 'mean response' histogram for each of the 2 text images sets. This was done simply by computing the sum of the response of each text image in each set (n = 10 in each set) and dividing the resulting response values by the number of the set. This produced two histograms: x_4 (mean response of mean word length = 4 letters) and x_7 (mean response of mean word length = 7 letters).

The second step was to compute a 'difference between means' (d) histogram which was obtained by subtracting the mean response of one set from the mean response of the other set ($x_7 - x_4$).

The third step was to compute the 'standard error of the difference between means' response, (obtained once the intermediate process of computing the variance of the differences had been performed).

The final step was to compute the equivalent of a t-statistic (t). This can be expressed as:

$$t = \frac{x_i - x_j}{\sqrt{\frac{\sum d^2 - \left((\sum d)^2 / n\right)}{n(n-1)}}}$$

Where x = the mean text image set, i = mean word length =4, j = mean word length = 7 response, d = the difference between mean responses and n is the number of histograms. From this, a final histogram, termed a 't' histogram could then be constructed. The t histograms thus provide a simple but reliable quantitative metric which describes the information available to perform the psychophysical task, according to the model.

3.6.2 Results: visual inspection

Figure 3.9 shows a visual description of the regions produced by MIRAGE across the range spatial scales of one example of an Experiment 1 ('boundary' word structure) text image.

The first thing to notice is that the features which the model extracts at each scale are, perhaps unsurprisingly, very similar to those found in Computational Analysis 1 (a full description of how regions correspond to features: *e.g.*, word breaks, word shape and letter strokes etc. was provided in Chapter 2, and is therefore not given here).

Coarsest spatial scale: filter s.d. = 12 pixels. Inspection of Figure 3.9 shows first, that at the coarsest spatial scale of 12 pixels (filter s.d.), shown in the bottom right hand panel, there is little in the image that has been extracted which corresponds to any features of the text. Some of the positive regions (light blobs) emerging tend to correspond to accidental alignment of word spaces from adjacent lines of text. However, for the task required of the subject–word segmentation–there seems to be nothing of use at this scale.

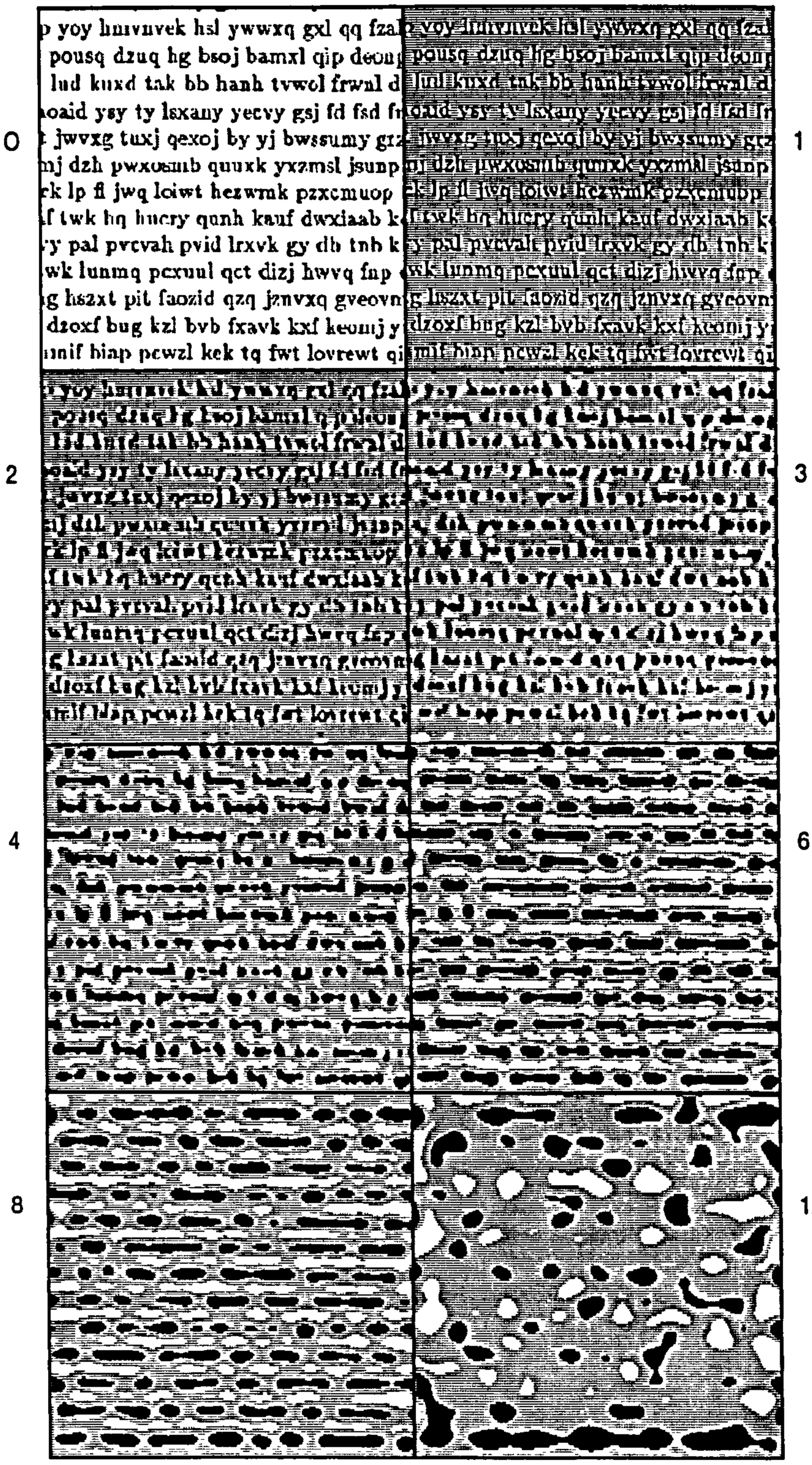


Figure 3.9. Information made explicit across spatial scale by the MIRAGE model represented as positive regions (bright 'blobs') and negative regions ('dark blobs') in an example of the text used in experiment 1. This page of text is the same as that illustrated in Fig. 3.1. Number adjacent to panels corresponds to the spatial scale(filter s.d. in pixels). See text for further details.

Spatial scale: filter s.d. = 8 pixels. There is some correspondence between negative regions and whole words. However, the length of regions does not correspond well with the length of words. Regions tend to break-up, which would suggest that using information at this scale to segment words would be unreliable. Positive regions tend to correspond to line spacing, and points where it is broken up occur at points where ascenders and descenders from adjacent lines align.

Spatial scale: filter s.d. = 6 pixels. The length and orientation of negative regions correspond almost exclusively to the length of each word at that place in the original text image. This is interesting because it was at this spatial scale at which optimum word segmentation performance occurred. Positive regions emerging correspond to line spacing.

Spatial scale: filter s.d. = 4 pixels. Negative regions (dark 'blobs') no longer correspond to whole word length, but have broken up in many parts. However, it can be seen that ascending and descending letter shapes are represented by the vertical orientation and length of some of the negative regions.

Spatial scale: filter s.d. = 3 pixels. A small vertical positive region now appears at almost every word spacing (word break). The results of Experiment 1 show that information used to segment words was contained at this spatial scale. It is possible that these features may be used in word segmentation when more salient or reliable features, particularly word length information, is unavailable, as would have been the case in Experiment 1 when the text was filtered at 3 pixels, whereupon whole word length would not be available.

The negative regions are completely broken up at this scale, but correspond very loosely with the position of individual letters.

Finest spatial scales: filter s.d. = 1–2 pixels. At the 2 finest spatial scales (filter s.d. = 1–2 pixels) the negative regions correspond to individual letters and letter strokes.

3.6.3 Results: summary of quantitative analysis

Figure 3.10 shows two *t* histograms of the difference in distribution of regions between the two image sets of Experiment 1 ('mixed' word structure condition). The left panel shows the positive region orientation *t* histogram, the right panel shows the negative region length *t* histogram. The pattern of response in the *t* histograms was very similar for all 3 word structure conditions, so only that for the 'mixed' condition is illustrated. Note that it is the positive region orientation histogram and the negative region length histogram which are

shown because it was only these parameters which produced any significant features in the histograms of the image analysis.

Taking the histograms of region orientation first, it can be seen that in each word structure condition, the significant area of region mass (to reiterate, the histograms are shown as density plots, so that the darker the area, the greater the region mass at that point) is contained in the vertical regions centred around spatial scale = 3 pixels. This distribution must correspond to the difference in the distribution of word breaks between the Experiment 1 image sets.

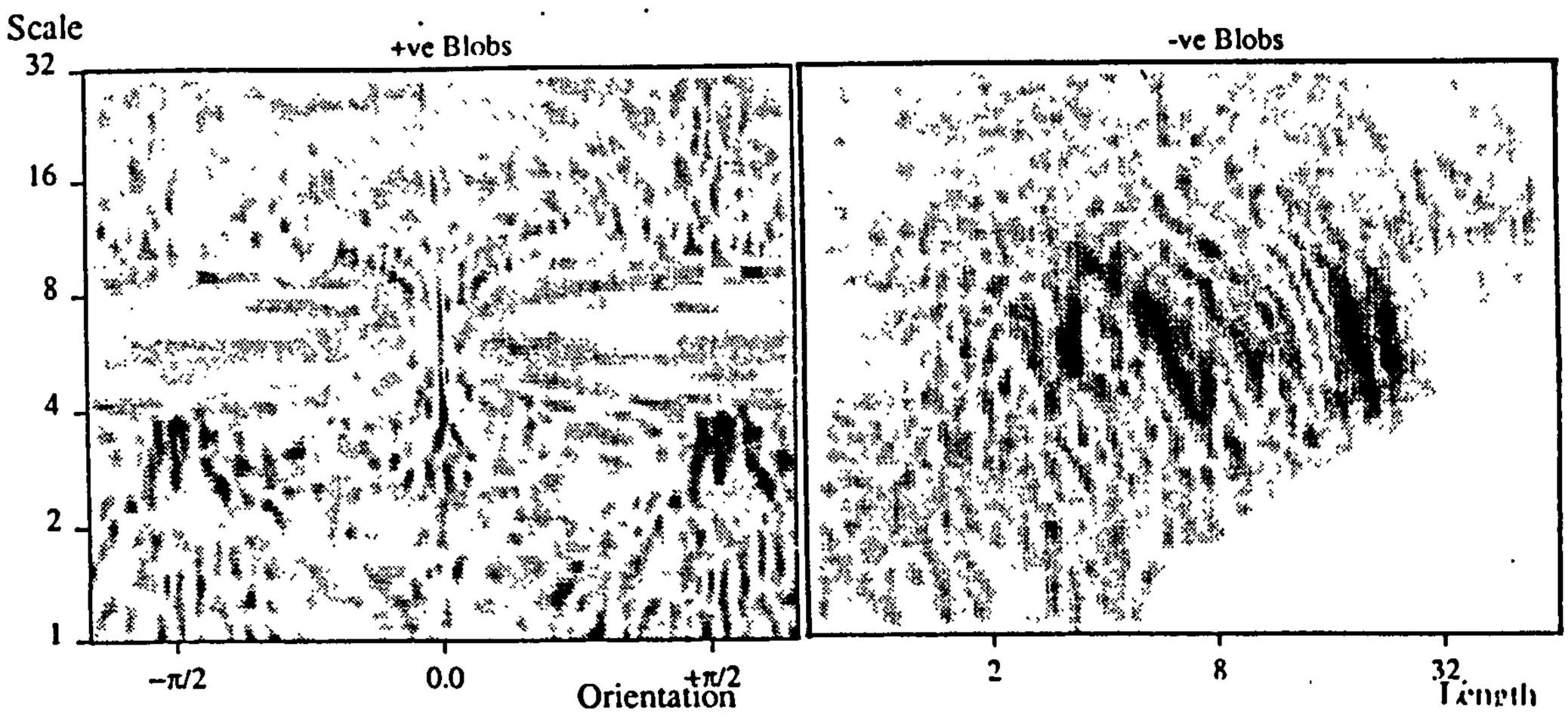


Figure 3.10. Histograms of the difference in distribution of region mass (*t* histograms) between the 2 image sets used in Experiment 1 ('mixed' word structure condition). Left panel shows positive region orientation and right panel shows negative region length. Histograms are shown as density plots, so the darker any point is the greater the region mass at that point. See text for explanation.

Examining the histogram of region length, a band corresponding to the difference in the distribution of region mass can clearly be identified around a spatial scale = 6 pixels. The distribution extends from 8 pixels to 40 pixels. This distribution of region mass corresponds to the difference in word length between the Experiment 1 image sets.

The final step is to make some direct comparison between the structures found in the *t* histograms (the model's discrimination between the image sets of Experiment 1) and the pattern of psychophysical performance (the subject's discrimination between the same two

image sets of Experiment 1). By doing so it should be possible to show how the model is able to describe visual text processing performance. This was achieved by taking the numerical value of region mass, in number of pixels (obtained from a numerical, as opposed to a density plot of the histogram of region mass) at the range of spatial scales for each parameter, length and orientation. The result of this is shown in Figure 3.11. This comparison shows that the model is able to provide a reasonable fit to the data in each of the word structure conditions of Experiment 1.

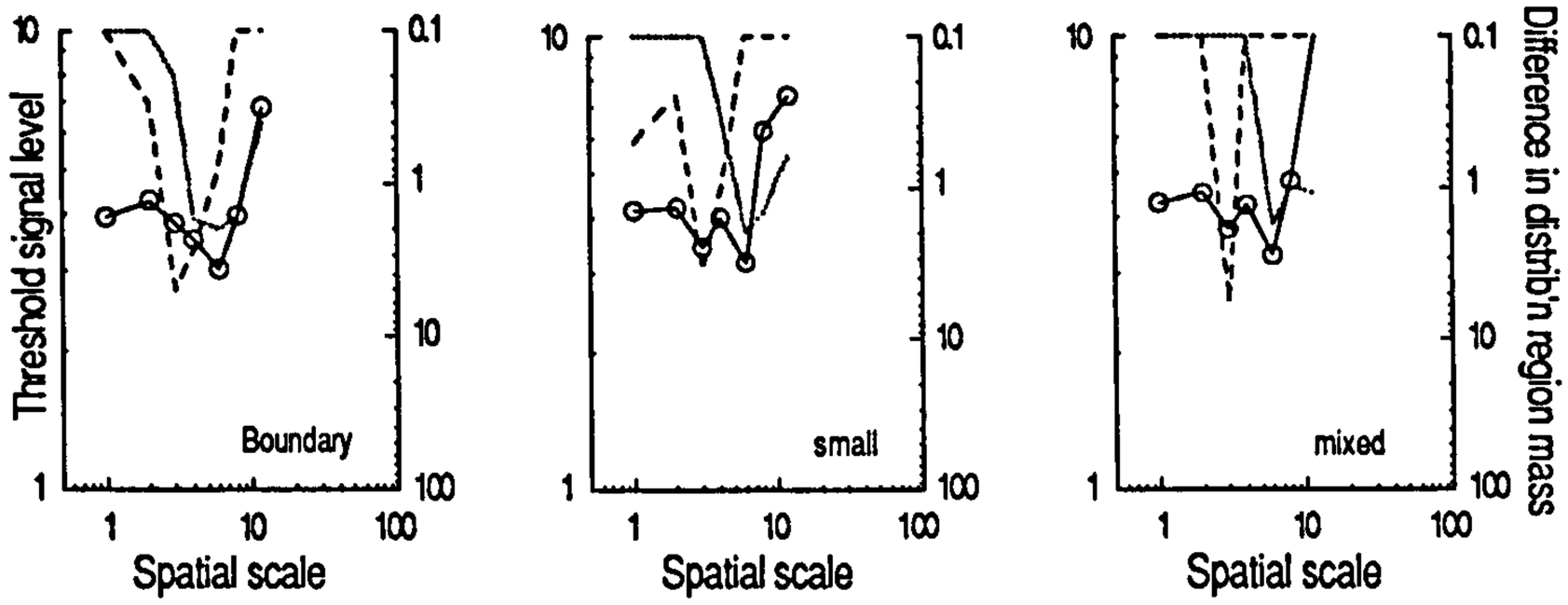


Figure 3.11. Comparison of psychophysical data of Experiment 1: word segmentation performance (subject PMC) plotted, on left ordinate, as open circles. Model performance is plotted, on right ordinate, as the difference in the distribution of region mass of the 2 Experiment 1 image sets (mean word length = 4 letters; mean word length = 7 letters). Dashed line is the difference in the distribution of positive regions. Dotted line is the distribution of the length of negative regions. No fitting has been applied. Model data is plotted directly against psychophysical data.

3.7 General Discussion

This chapter began by asking the first question which needed to be asked in this study of the visual processing of text: whether the visual processing of text is spatial scale dependent. In order to try to answer this question, a number of methodological problems first needed to be addressed. These were based on the need to isolate individual text processing tasks of the type expected to be performed in reading, and to isolate the visual aspects of text processing performance.

Experiments 1–3 demonstrated that the visual processing of text is spatial scale dependent. Psychophysical examination revealed that the most useful information for performing 3 text processing tasks was contained at three different spatial scales.

It was then necessary to examine the basis of this observed pattern of text processing performance. In particular, to try to determine something about the nature of the visual information used to perform these text processing tasks. Computational Analysis 2 (this

chapter) sought to examine this issue. This required devising a procedure in which the information available to perform the psychophysical judgements required in Experiment 1 could be modelled. Having devised a suitable procedure (the construction of a *t* histogram) for describing the psychophysical discriminability of images by the model, the results of this analysis showed that the model was able to provide a reasonable description of the information used in the visual processing of this task. The findings of Experiments 1–3 and Computational Analysis 2 therefore lend support for the hypothesis (1) that the visual processing of text should be spatial scale dependent, and in a way which is predicted by the model.

The implications of the findings are thus. First, these initial findings show that the model of the visual processing of text is at least consistent, so far, with human visual processing behaviour, suggesting that the model may be able to provide an adequate description of the visual processing of text. Second, if this first implication finds further support, then the possibility exists for extending our understanding about the possible nature of the early visual representation–and the visual processing–of text.

Further, and clearer, support for these possibilities is required. In this respect it is hypothesised that if readers use the type of information represented by the model in the visual processing of text, then changes in text processing behaviour resulting from changes in some physical parameter of the text should be accompanied by changes in model behaviour. This is *Hypothesis* 2 (see Section 2.3), and it is tested in Chapter 4.

4

Visual Processing of Text: Word–Level Effects

The findings of Experiments 1–3 and the comparison of the word segmentation data of Experiment 1 with the results of Computational Analysis 2 provided encouraging, but only initial, support for the model of the visual processing of text, suggested at the end of Chapter 2. A good way of testing further this model is to examine how sensitive human visual processing behaviour is to changes in the output of the model as a function of systematic changes to one or more of the physical characteristics of text. It is the next logical step in the process of trying to discover the nature of the visual processing of text, and it is the aim of this Chapter.

4.1 Experiment 4: Word segmentation as a function of word spacing

Experiment 4 establishes how sensitive human text processing (word segmentation) performance is to changes in the typographical parameter of word spacing. This is done using a psychophysical procedure similar to that used in Experiment 1. The main methodological difference is that in this experiment the text was unfiltered (had normal bandwidth).

The findings of Experiment 4 are then compared to a third computational analysis in which the sensitivity of model of vision to the same physical changes in the text under modelled psychophysical conditions (*i.e.* the same procedure as that performed in Computational Analysis 2) is determined.

4.1.1 Method

(i) Text characteristics. 120 text images (60 with a mean word length of 4 letters; 60 with a mean word length of 7 letters) for 6 levels of word spacing in each of 3 word structure conditions identical to those of Experiment 1 were created. Therefore the term 'word' refers

to a letter string whose characteristics conformed to one of 3 word structure conditions: 'boundary', 'small' or 'mixed', as outlined in Chapter 3. All other aspects of the text were identical to those of Experiment 1, with the exception that word spacing had 6 levels: 0.0, 0.6, 1.2, 1.8, 2.4 and 3.0pts. Examples are shown in Figure 4.1.

(ii) Text image digitisation. The digitisation process was identical to that described in Section 3.1. The other image processing operations described in Section 3.1 are irrelevant to this experiment, since the text was unfiltered, and presented at full bandwidth.

(iii) Procedure. Procedure was similar to that of Experiment 1 (see Section 3.1 for details of general procedure). So, subjects were presented with 2 300x300 pixel text images which appeared simultaneously at adjacent positions either side of the centre of the display for 3000msec. In a two alternative forced choice (2AFC) procedure, the subject's task was to decide which text image had the greatest mean word length. An adaptive method of constant stimuli (APE) was used to select a range of word spacings on the psychometric function over the 64 presentations. APE generated a range of 6 stimulus levels. Thus for a stimulus value of 6, text with typographically conventional word spacing (3.0 pts) was presented and the subject was able to perform at 100% correct. At a stimulus value of 0 text with no word spacing (0pts) was presented and the task was impossible.

(iv) Subjects. The same 3 subjects who took part in Experiments 1–3 (PMC RAO and the author, SJE) also took part in this experiment.

4.1.2 Results and discussion

Table 4.1 summarises mean word spacing required for word segmentation for 3 subjects in each of the 3 word structure conditions. Figure 4.2 shows the individual psychometric functions for Experiment 4 for the 3 subjects in each of the 3 word structure conditions from which these thresholds were estimated. The data was fitted to the best fitting cumulative Gaussian curve which described the mean and standard deviation (sensitivity) of the subject's response error distribution. Inspection of Figure 4.2 reveals two aspects of the data. The most obvious, but least interesting, feature is that word segmentation performance was dependent on word spacing. However, the nature of this relationship with respect to word structure is rather more interesting. Word spacing required for word segmentation varied according word structure. Having ascending/descending letters in word boundary positions reduced the word spacing required to segment the words compared to the small and mixed word structure by a factor of 1.4 and 2 respectively for

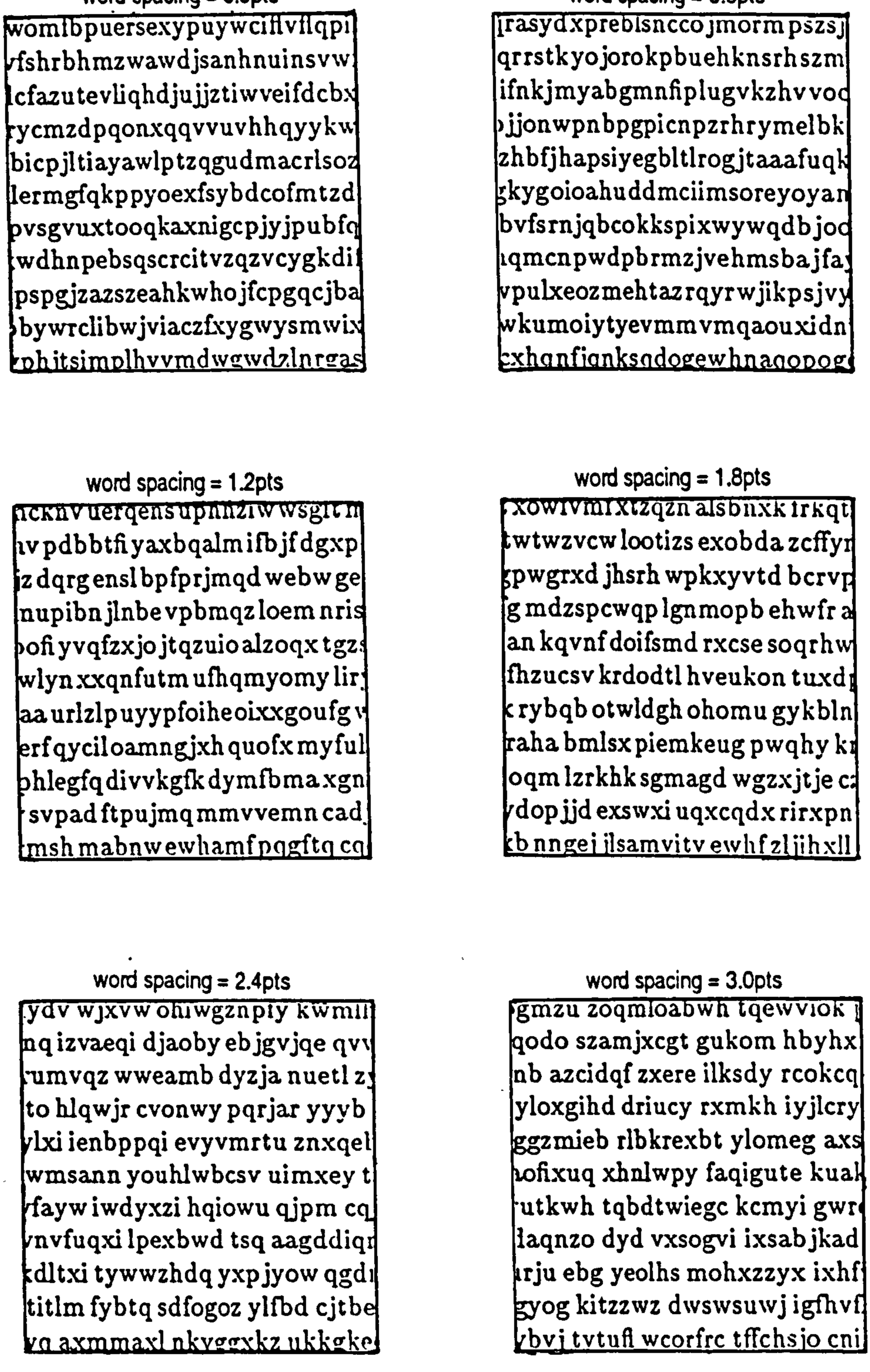


Figure 4.1. Examples of text images used in experiment 4. Each panel shows text having a different word spacing. Word spacing varied between 0-3pts in 5 0.6pt increments. Number adjacent to panel corresponds to word spacing in pts. Shown here is mean word length =7 letters for 'mixed' word structure condition only. See text for further details.

PMC and by a factor of approximately 3.5 and 5 respectively for SJE. It is not clear why RAO did not exhibit the same type of performance dependency on word structure.

The finding that word structure influenced word segmentation performance provides some support for the suggestion made by Walker (1987) that the higher probability of ascenders or descenders occurring at word boundary positions might be beneficial in reading (it is assumed here that what Walker meant was to provide some additional cue to word segmentation) in texts where word spacing is very narrow. This may be the reason why the effect of word structure on word segmentation was found to be weak in Experiment 1, which used text which had a fixed, relatively wide word spacing.

Threshold (P[c]=83%) word spacing

	Boundary	Small	Mixed
PMC	0.66	1.30	0.95
RAO	1.40	1.40	1.37
SJE	0.35	1.63	1.29

Table 4.1. Experiment 4 summary data. Word spacing (in points) required to reach word segmentation performance threshold for each of 3 word structure conditions: boundary, small and mixed.

Given the findings of this thesis so far, it is possible that the reason for this latter feature of the results is that ascenders and descenders, when at word boundary positions, make available (otherwise unavailable) information corresponding to either whole words or to word spacing, or both. This possibility requires further examination to determine which, if any, of these features is represented at narrower word spacings. However, this issue is not addressed any further in this thesis. The primary reason for this is that this issue has essentially been an aside, not a core issue, and to examine it any further required very large amounts of computing time and space which was simply not available in our laboratory. Instead, text from the 'mixed' word structure was chosen for further analysis because of its greater similarity to the statistics of English than any of the other word structure conditions. Details of this analysis are given in Section 4.2.

The main interest in the pattern of performance is the actual word spacing required to reach threshold word segmentation performance. To discover the basis for the word spacing needed to reach threshold word segmentation performance it is necessary to compare the pattern of word segmentation performance found in Experiment 4 to the pattern of modelled visual processing as a function of the same changes in word spacing.

The aim of this is to provide a reliable and informative measure of the ability of the MIRAGE model to describe the pattern of text processing performance.

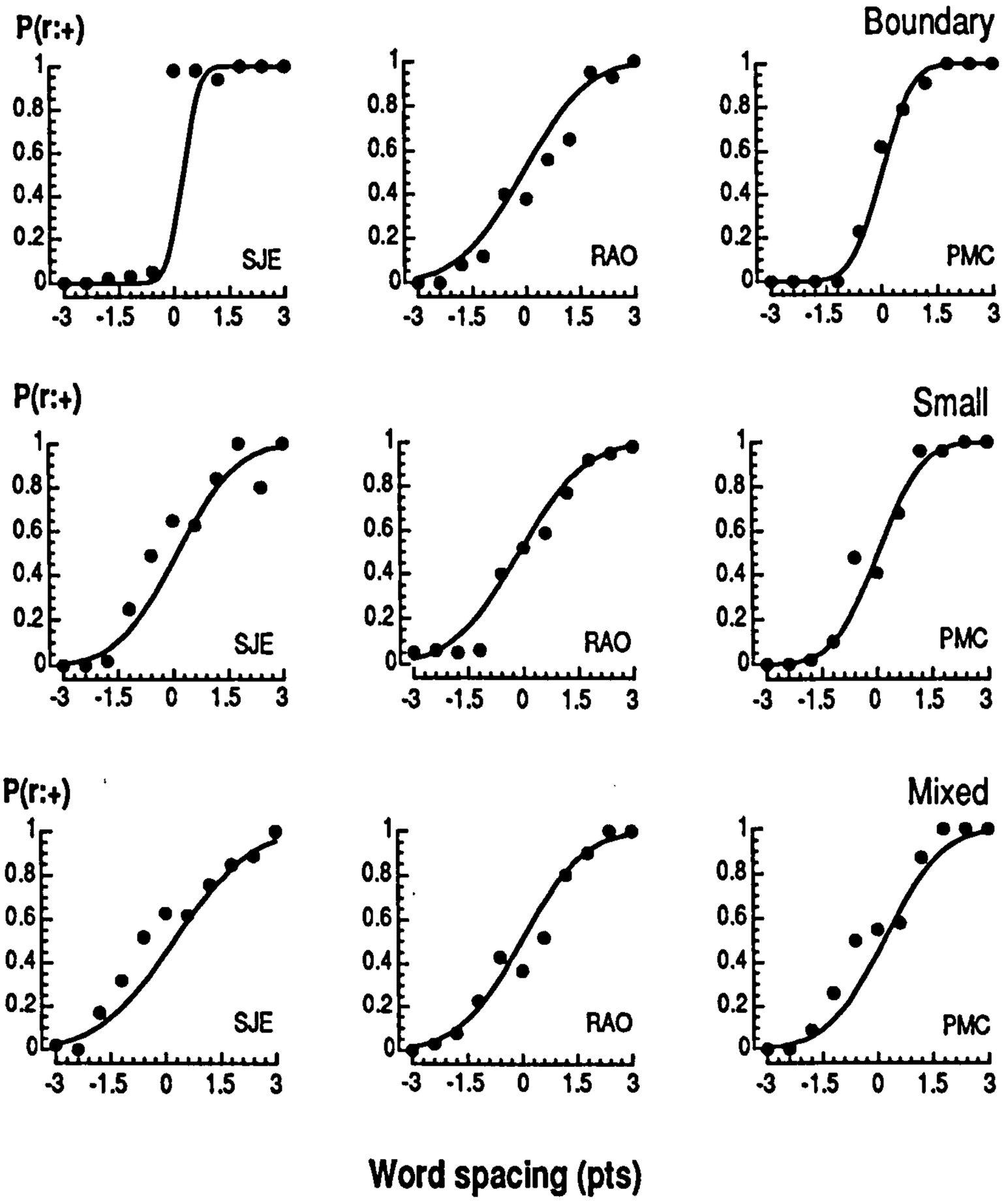


Figure 4.2. Psychometric functions of word segmentation performance as a mean word length estimation discrimination as a function of word spacing for 3 subjects in each of 3 word structure conditions. The ordinate, P(r:+), is the probability (P) of responding correctly (+) when the text image containing the cue (mean word length=7 letters) appeared on the right side (r) of the display. Abscissa is the word spacing of the text in points. Negative numbers refer to text presented on the left side of the display. Data is plotted this way to determine bias in response (in this case there was none, *i.e.* chance levels of performance always occur at 0pts word spacing). The top panel is the boundary condition, middle panel is small condition and bottom panel is the mixed condition. Each data point is the mean of 3 runs. Each run contained 64 measurements of response at the positions on the psychometric function given by each data point.

On the basis of the findings of the comparison of Experiment 1 with Computational Analysis 2, it is predicted that the features corresponding to whole words (word length)

contained in the negative regions emerging at spatial scale = 6 pixels should describe the sensitivity of word segmentation performance as a function of word spacing. This was examined in Computational Analysis 3.

4.2 Computational Analysis 3

4.2.1 Method

(i) Text images. Half of the original text images from each of the six different word spacing sets used in Experiment 4 'mixed' condition were randomly selected for the analysis.

(ii) Procedure. The procedure was identical to that employed in Computational Analysis 2, and is therefore not described again here. The reader will find details of the procedure for the image processing operations and *'t'* histogram construction in Chapter 3, Section 3.6.

4.2.2 Results

Figures 4.3 and 4.4 illustrate how the pattern of response of the model of vision varies as a function of three word spacings (ws). The text image in the top left panel of Figure 4.3 (Fig. 4.3a) has a word spacing just narrower than that required to reach threshold word segmentation performance (ws = 0.6pts). The one in the middle left panel (Fig. 4.3b) has a word spacing nearest to that required to reach a threshold level of performance (ws = 1.2pts) and the text image in the bottom left panel (Fig. 4.3c) has a word spacing just wider than that required for threshold performance (ws = 1.8pts).

The right hand panels show the 'visual description' of the regions the model extracts in each of these text images at one spatial scale (6 pixels) for each of these word spacings. In the top right panel (Fig. 4.3d), the negative regions do not correspond to word lengths. In the middle and bottom right panels (Fig. 4.3e-f), the length of the negative regions does now tend to correspond to the length of the words at the word spacings at which threshold psychophysical word segmentation performance occurred.

Figure 4.4 shows the histograms of the difference in the distribution of the length of region mass (the *'t'* histograms) as a function of spatial scale each at one of the 3 word spacings shown in Figure 4.3. Negative region length histograms only are shown as there was no structure at all to the orientation histograms at these word spacings.

The first thing to notice about the histograms is that the distribution of region mass (shown as the dark areas in the histogram density plot) changes with word spacing. In the top panel (ws = 0.6pts) the model is unable to extract any features of the text images to discriminate between the 2 text image sets at this word spacing.

Figure 4.3. The visual description of the representation of 3 pages of text used in Experiment 4 by the MIRAGE model. Shown are 3 examples of text (left panels), and the MIRAGE description of the same text images (right panels) at one spatial scale (filter s.d. 6 pixels). Each example differs in word spacing. In the top panels, the word spacing is 0.6pts: just below that required to reach threshold word segmentation performance. Notice how the negative regions (dark 'blobs') tend to merge into very elongated blobs. Text in the middle panels have a word spacing of 1.2pts: approximately that required to segment words. Notice that now, the negative regions suddenly tend to correspond with whole words, i.e. the words tend to be segmented. Text in the bottom panels has a word spacing of 1.8pts: just above that required for threshold word segmentation performance.

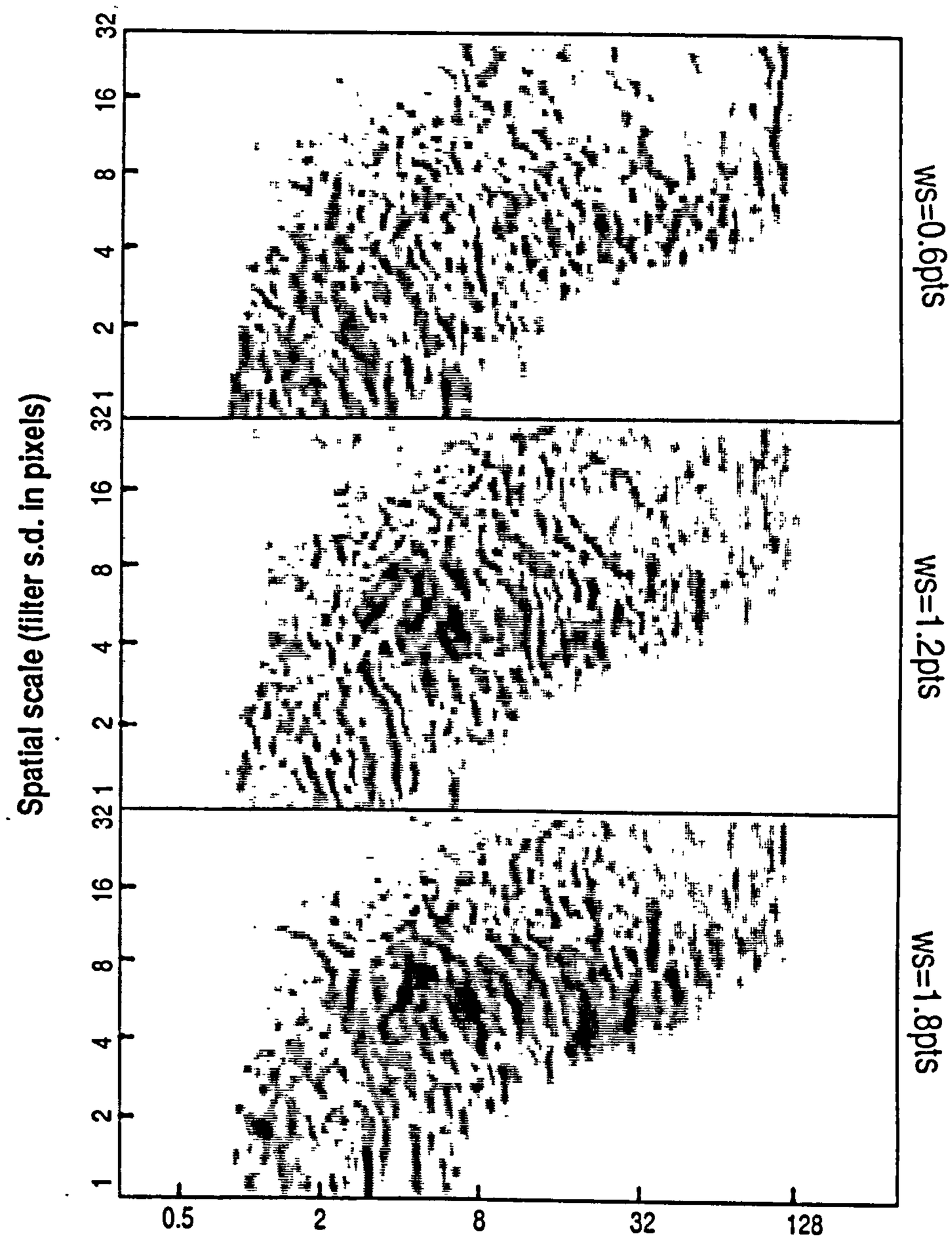


Figure 4.4. 'r' Histogram of the difference in the distribution of negative region length as a fiunction of spatial scale (filter s.d. in pixels) shown on ordinate and region length (in pixels) shown on abcsissa. Analysis is the modelled discrimination of the 2 text image sets (mean word length=4 letters vs. mean word length=7 letters) used in Experiment 4 (mixed word structure condition) at 3 word spacings (ws). The top panel shows the representation of regions made available in modelling the discrimination between text image sets with a ws=0.6pts (just below that required to reach word segmentation threshold). Middle panel shows same for text with ws=1.2pts, nearest to that required to reach threshold word segmentation performance. Bottom panel shows same for text with a ws=1.8pts, just wider than that required for word segmentation. Emergence of regions discriminating two image sets is at word spacing needed for threshold word segmentation performance.

In the middle panel of Figure 4.4 (ws = 1.2pts) an emergence of region mass (a dark area) can now be seen, centred at a spatial scale = 6 pixels. The model has been able to extract some feature which discriminates the 2 image sets at this word spacing. This is the same word spacing at which threshold word segmentation performance occurred. Inspection of the text image at this word spacing in Figure 4.3 (middle panel) shows that this is the point at which the regions tend to correspond to whole words.

In the bottom panel of Figure 4.4 (ws = 1.8pts) this area is now more pronounced. Inspection of the bottom panel of Figure 4.3 shows that now there is an even more reliable correspondence between the negative regions (dark 'blobs') and the length of each word in the image.

The way in which the difference in the distribution of region mass varied across the entire range of six word spacings can be seen very clearly in Figure 4.5, which shows basically the same information as that of Figure 4.4 (a *t* histogram), but which is displayed in a slightly different form. The axis showing the distribution of region length in Figure 4.4 has now been collapsed, leaving just the distribution of region mass (ordinate) and spatial scale (abscissa).

Each of the curves represents the difference in the distribution of region mass *i.e*, the modelled discriminability of the 2 text images as a function of one of the six word spacings. It is very clear to see that as word spacing changes, the distribution of region mass also changes. Importantly, it can also be seen that the distribution of region mass is centred around the spatial scale = 6 pixels, the same spatial scale which was found to contain the information used for optimum word segmentation performance in Experiment 1.

Whether this pattern of response of the model can adequately describe the sensitivity of word segmentation performance can be finally determined by fitting the output of the model to the psychophysical data.

This was done by first measuring the peak value of region mass as determined from the peak value in Figure 4.5 for each word spacing. Having obtained this measure of 'peak value of region mass' at each word spacing, the value at ws=0.0pts was set at P = 0.5 probability correct (chance) and 'peak value of region mass' at ws=0.6pts was then set to the same as that of the psychophysical estimate at ws=0.6pts (P =0.56). This allowed the model to be fit to the data with only one free parameter. The model data ('peak value of region mass') at every word spacing was then fitted to the psychophysical data on the basis of a proportional difference between the difference in word spacing between ws=0.0–0.6pts and the difference in the 'peak value of region mass' between these two word spacings. It must be noted that this procedure does not prevent the model doing better than P >1.0. Indeed,

this turned out to be the case for the widest word spacing (3.0pts), where the modelled probability was in the order of $P = 1.23$. However, it was clamped to P=1.0, justification for which is on the basis that the subjects could not do better than P=1.0.

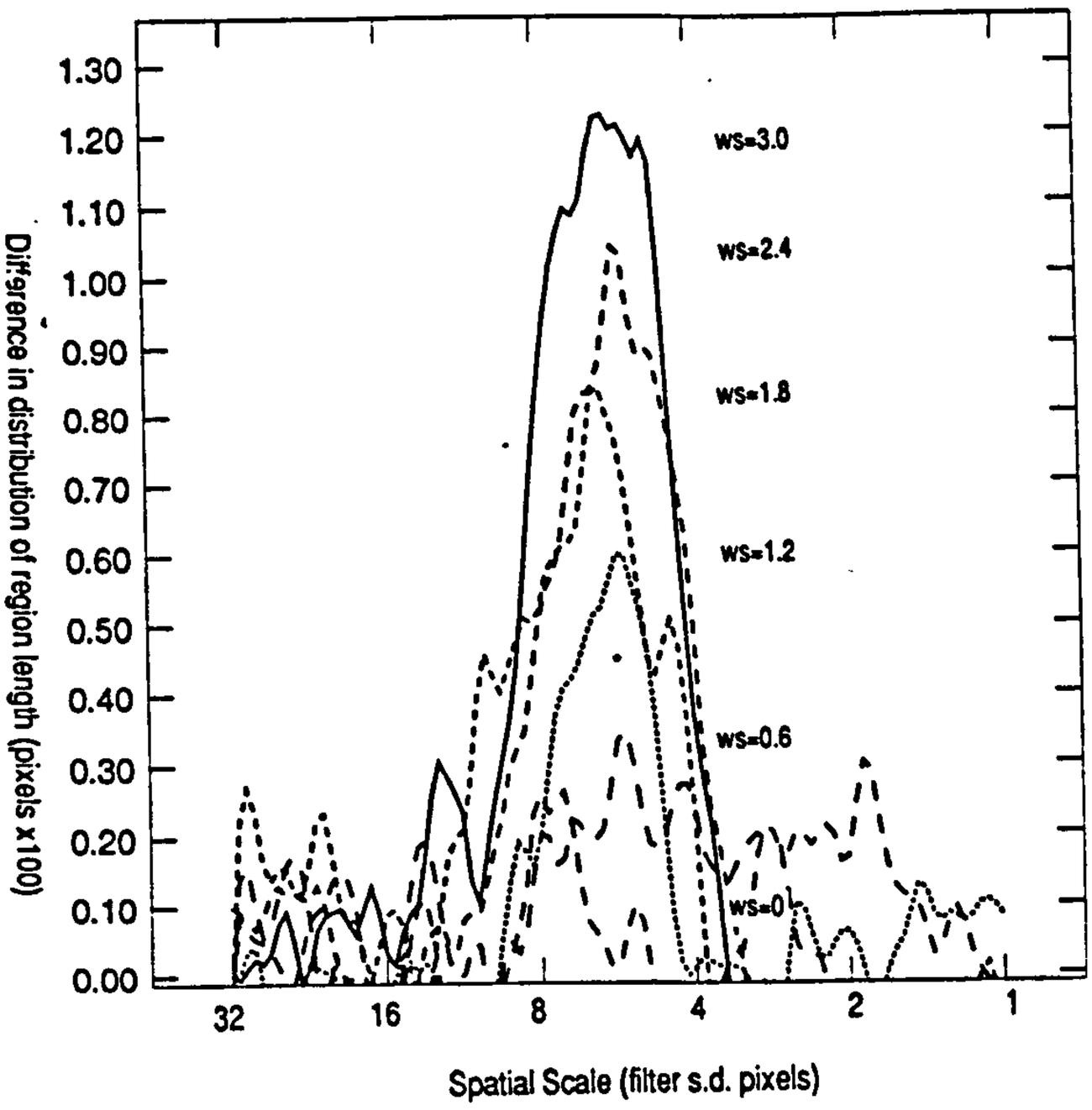


Figure 4.5. A series of *t* histograms showing the difference in the distribution of negative region mass (ordinate) between the cue(mean word length = 7 letters) and reference (mean word length = 4 letters) image sets of Experiment 4 as a function of spatial scale (abscissa) at each of the 6 word spacings (ws) used in Experiment 4 (one of the 6 curves). See text for further details.

Figure 4.6 shows how the model fits to the psychophysical data in this manner. It shows that the model provides a good fit to the pattern of visual processing of text (word segmentation performance) as a function of word spacing. This is supported by a test of goodness of fit using a chi square test ($\chi^2 = 0.58; p > 0.98$).

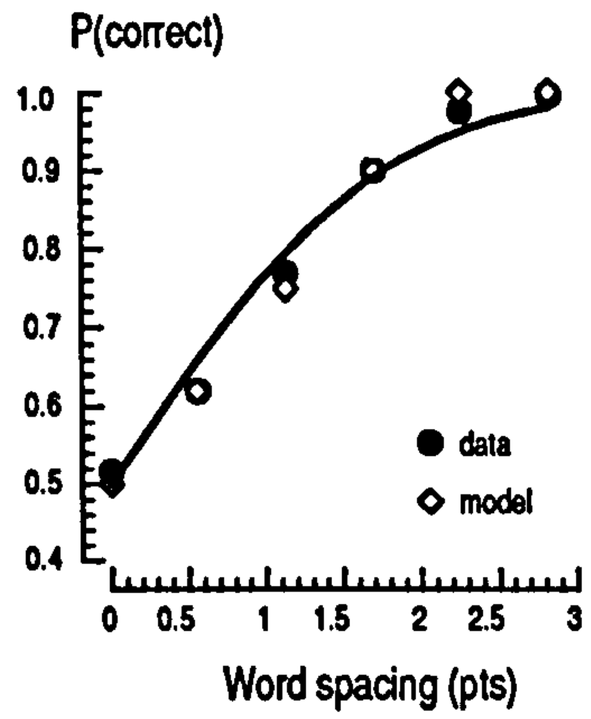


Figure 4.6. Mean of 3 subjects data from mixed word structure condition of Experiment 4 (filled circles) and data from model (open diamonds) both fitted to a psychometric function described in Section 4.2. See text for further details.

4.3 Discussion

The findings of Experiment 4 and Computational Analysis 3 have shown how the pattern of visual processing of one text processing task in reading, word segmentation, can be described by the pattern of the response of the MIRAGE model of early vision as a function of varying word spacing. Thus, these findings support *Hypothesis* 2 which stated that this is exactly what should occur if the model of vision delivers a representation of text which is similar to the human visual processing of text.

Furthermore, the findings also lend support to the MIRAGE model itself as a working model of early human vision. This is interesting because it demonstrates that the model is able to predict and describe the performance of real visual tasks, and not just 'basic' visual tasks usually examined in psychophysical experiments.

5

A Visual Description of Text: Typographical Effects

It is possible to extend the approach taken in Chapter 4 to examine further the nature of the visual processing of text. The work of Chapter 4 examined the effects of word spacing on word segmentation. This can be considered as a 'word-level' analysis. However, the existence of typographically conventional (or at least acceptable) word spacing and line spacing relationship (*e.g.*, Spencer, 1968) suggests that the effects of word spacing on the visual processing of text, and indeed on the pattern of response of the model, are also likely to be influenced by the relationship between typographical parameters of the text which occur at a *'page-level'* of description. The aim of the work described in this Chapter (Computational Analysis 4) is to explore the effects the typographical layout of a page of text might have on the visual processing of text at this 'page-level'. In Computational Analysis 4, the features the MIRAGE model of early vision extracts from text as a function of both word spacing and line spacing are analysed. The pattern of response of the model as a function of these page-level characteristics is then compared to the pattern of human performance in the visual processing of text in subsequent chapters.

5.1 Computational Analysis 4

5.1.1 Method

(i) Histogram construction. The same basic set of image processing operations employed in the previous computational analyses was applied in this computational analysis, except for one difference. This analysis had to describe the performance of several, different types of experiments, one of which was not a psychophysical experiment. Therefore, constructing a *t* histogram from the analysis was considered to be inappropriate, and it is argued that it was unnecessary. Having identified in previous analyses the important visual information which subjects may use to perform text processing tasks, it was possible to examine the fate

of that information successfully without having to first identify what that information was by the construction of a *t* histogram. Therefore, this analysis, whose construction otherwise follows an identical method to that of Computational Analysis 1, still provided a simple and reliable quantitative analysis of the image content across spatial scale by the operations performed by the MIRAGE algorithm.

(ii) Text images. Text used in Experiment 8 served as the input to the analysis. 40 pages of text were created. Each page was taken from *The Independent* newspaper magazine satirical commentary 'Up and Down the City Road', which has as its subject current affairs. The mean passage length was 152 words, s.d. = 15.3 words. Mean word length was 4.7 letters.

Text was generated in Apple Truetype 12pt Times Roman font, left justified, using an Apple Macintosh IIcx computer running Microsoft Word. A hardcopy of each passage was produced by an Apple LaserWriter IIg onto white A4 paper.

A 300x300 pixel section of each page of text was then digitised in the same manner to that described in Section 2.1 using a Hewlett-Packard Scanjet digital scanner. A set of image processing operations and steps leading to the construction of a set of histograms of the distribution of mass of region orientation and length were performed on the resulting text images. All these steps were identical to those performed in Computational Analysis 1, and are described in full in Chapter 2, Section 2.1.

There were 2 typographical parameters of the text: line spacing and word spacing. Line spacing (ls) had 4 levels: 0, 2, 8, and 16pts (defined as the spacing between the bottom of one line to the top of the line below). Word spacing (ws) had 5 levels: 0, 1, 2, 4 and 8pts. There were 2 passages for each typographical combination, totalling 40 passages. Examples of the text can be seen in Figure 5.1.

5.1.2 Results

Figure 5.2a–c illustrates some examples of the MIRAGE visual description of image content at several spatial scales, and at a number of different word spacing and line spacing combinations. Figures 5.3 and 5.4 show how the distribution of region mass varied with word and line spacing combinations. In Figure 5.3 is a set of histograms, each one showing positive region orientation as a function of word and line spacing combinations. Each one of the histograms shown in Figure 5.4 describes the distribution of negative region length as a function of the same word spacing and line spacing combinations.

General. Before beginning a detailed description of the results of the analysis, there are two general features to note about the findings. The first is that there is a definite pattern to the

histograms of region distributions, and moreover, a systematic change in the pattern as a function of word and line spacing. The second is that the effect of word spacing and line spacing on the features the model extracts (shown as the distribution of regions in the histograms) is not separable. That is, the effects that word spacing has on the features extracted from a page of text (the distribution of regions in the histograms) depends on the line spacing in that page of text.

From the findings of previous Chapters, the expected features of interest are already known. The first is the negative horizontally elongated regions (the dark 'blobs') corresponding to the words themselves, whose distribution of length is determined by the distribution of word length. A second possible feature, identified in Experiment 1 and Computational Analysis 2, is the positive, usually vertical, regions (the light 'blobs') which correspond to word breaks. The effects of word and line spacing are considered in each of the 'bands' of word spacings (ws) below.

ws=0-2pt: Inspection of the examples shown in Figure 5.2, and the histograms in Figure 5.3 and 5.4 shows first that, unsurprisingly, when there was zero word spacing, neither the negative regions corresponding to words nor the positive regions corresponding to word breaks were made available by the model. However, when word spacing had increased to 2pts, positive regions corresponding to word breaks emerged at scale = 3 pixels for line spacings between 0–8pts. At the widest line spacing (16pts) these positive regions disappeared.

At even the narrowest word spacing (ws=1pt), providing line spacing was 8pts (typographically 'conventional') there was an emergence of negative regions corresponding to word length at spatial scale =6 pixels. When line spacing was very wide (ls=16pts) and word spacing was very narrow (ws=1pt) the negative length regions had a tendency to correspond to whole lines, and whole word length was unable to be extracted. This is even though at word spacings of a similar width at narrower line spacings, word length regions were made available. There was also an emergence of positive length regions corresponding to whole line spacings at this arrangement making the text look very 'stripy'. However, even at such wide line spacings (16pts) when word spacing had reached 2pts, negative regions corresponding to whole word length were made explicit, although the response in the histogram looks weaker than at more favourable arrangements (*e.g.*, ws = 4pts and ls = 8pts) because there were fewer words on each page at this point.

ws=2-4pts: The spatial scale at which positive regions corresponding to word breaks appeared changed as both word and line spacing increased, so that for line spacings of 0–2 pts, word breaks were made explicit first at scale =3 pixels for narrow word spacings of 2pt, then spatial scale 3 and 4 pixels when word spacing was 4pts. There was a clear distribution of negative regions corresponding to word length at spatial scale = 6 pixels at this word spacing when line spacing had reached 8pts.

ws=4-8pts: When word spacing was twice conventional width at 8pts, a distribution of positive regions corresponding to word breaks appeared at spatial scale = 4 pixels, at narrow to conventional line spacings (0–8pts), at which point they were no longer vertical but horizontal regions. When line spacing and word spacing were twice conventional size (ls=16pts; ws=8pts) word breaks were not represented at all: these regions disappeared altogether. When word spacing was twice conventional width (ws=8pts) negative regions corresponding to whole word length were made explicit across a larger range of spatial scales (4–8 pixels), although it can be seen from Figure 5.4 that the response was weaker than that for the conventional word and line spacing arrangement (ws=4pts and ls=8pts).

Figure 5.1. (shown overleaf). One example of each of the text images (taken from Experiment 8 text) differing in word spacing and line spacing which was used in Computational Analysis 4. Figure 5.1a shows text at the full range (0–8pts) of word spacings (ws) for 2 of the 4 line spacings (ls) (0 and 2pts). Figure 5.1b shows the same range of word spacings (ws) for the other 2 line spacings (8 and 16pts).

Figure 5.2 (shown overleaf). Examples of the 'visual description' (a visualisation of the representation) of the features, or regions, of the text images extracted by the model of vision at 3 spatial scales. In each Figure the top row of panels shows the original text image. The next rows of panels down shows the regions extracted at spatial scale = 1, 3 and 6 pixels (denoted by the number to the left of the row). ws and ls shown at top of each Figure refers to word spacing and line spacing, in points (pts). See text for explanation of emergence of regions as a function of word spacing and line spacing.

Figures 5.3 and 5.4 (shown overleaf). Histograms of the distribution of the mass of region orientation (Fig. 5.3) and length (Fig. 5.4) as a function of word spacing and line spacings. See text for further details and explanation.

Fig.5.1a

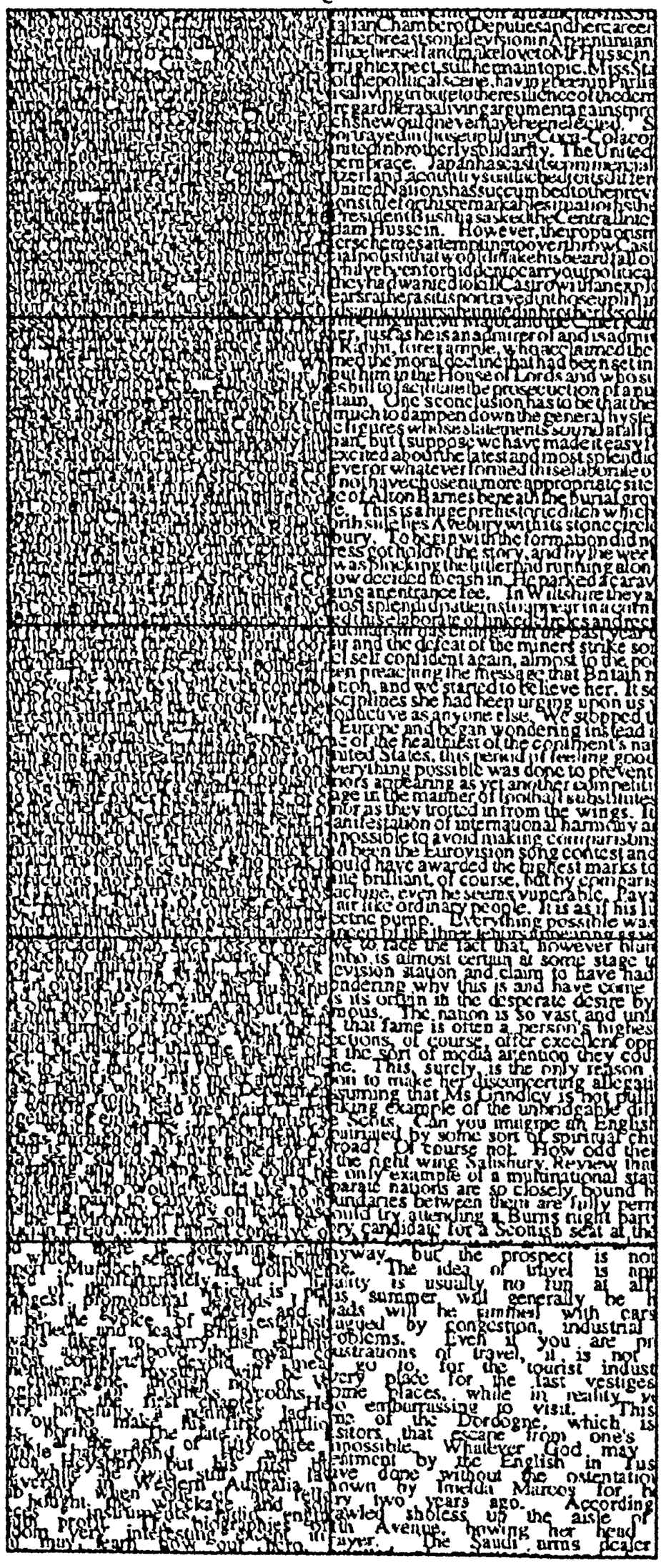
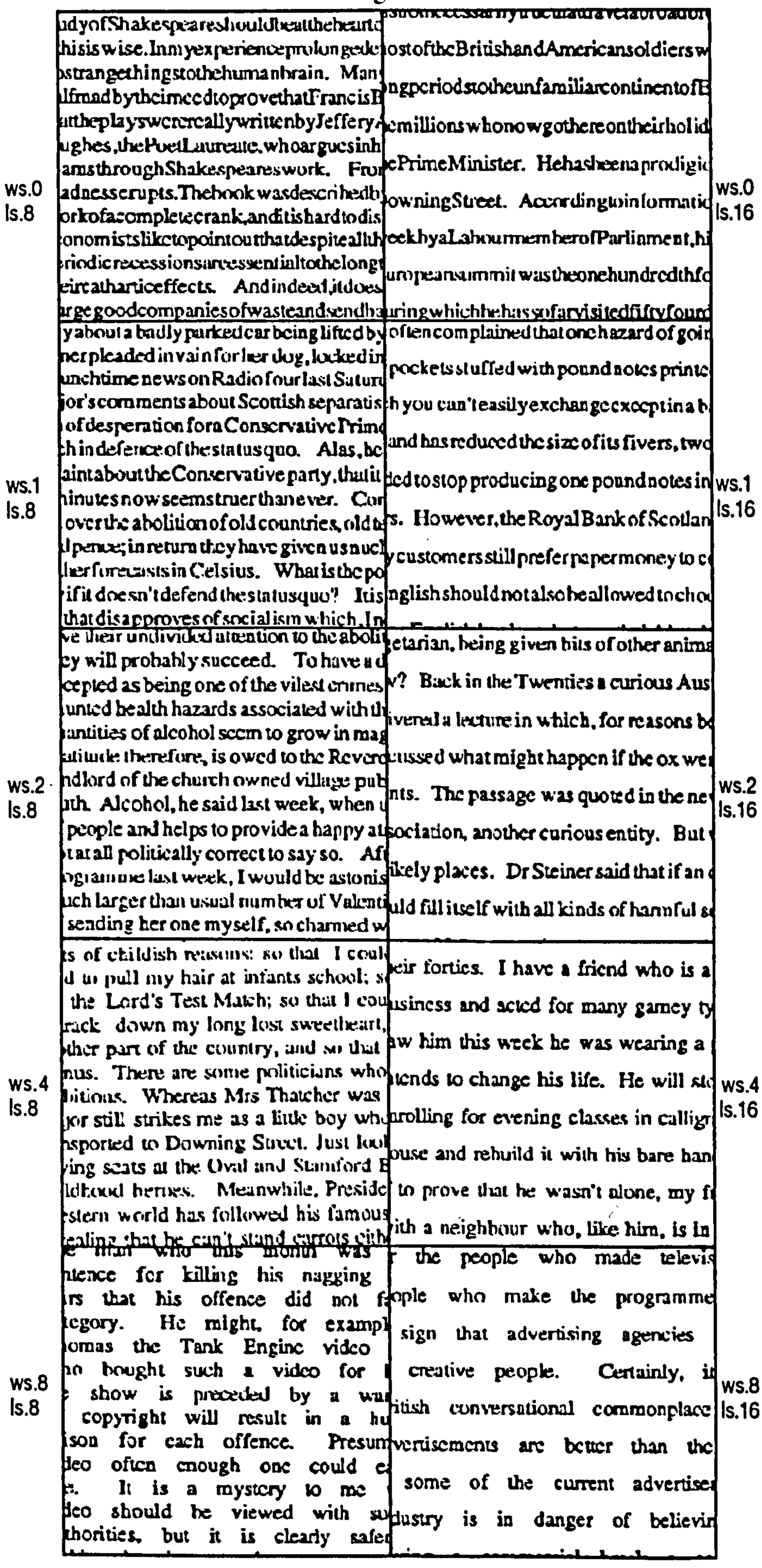

ws.0
ls.0
ws.1
ls.0
ws.2
ls.0
ws.4
ls.0
ws.8
ls.0
ws.0
ls.2
ws.1
ls.2
ws.2
ls.2
ws.4
ls.2
ws.8
ls.2

Fig.5.1b


ws.0 ls.8
ws.1 ls.8
ws.2 ls.8
ws.4 ls.8
ws.8 ls.8
ws.0 ls.16
ws.1 ls.16
ws.2 ls.16
ws.4 ls.16
ws.8 ls.16

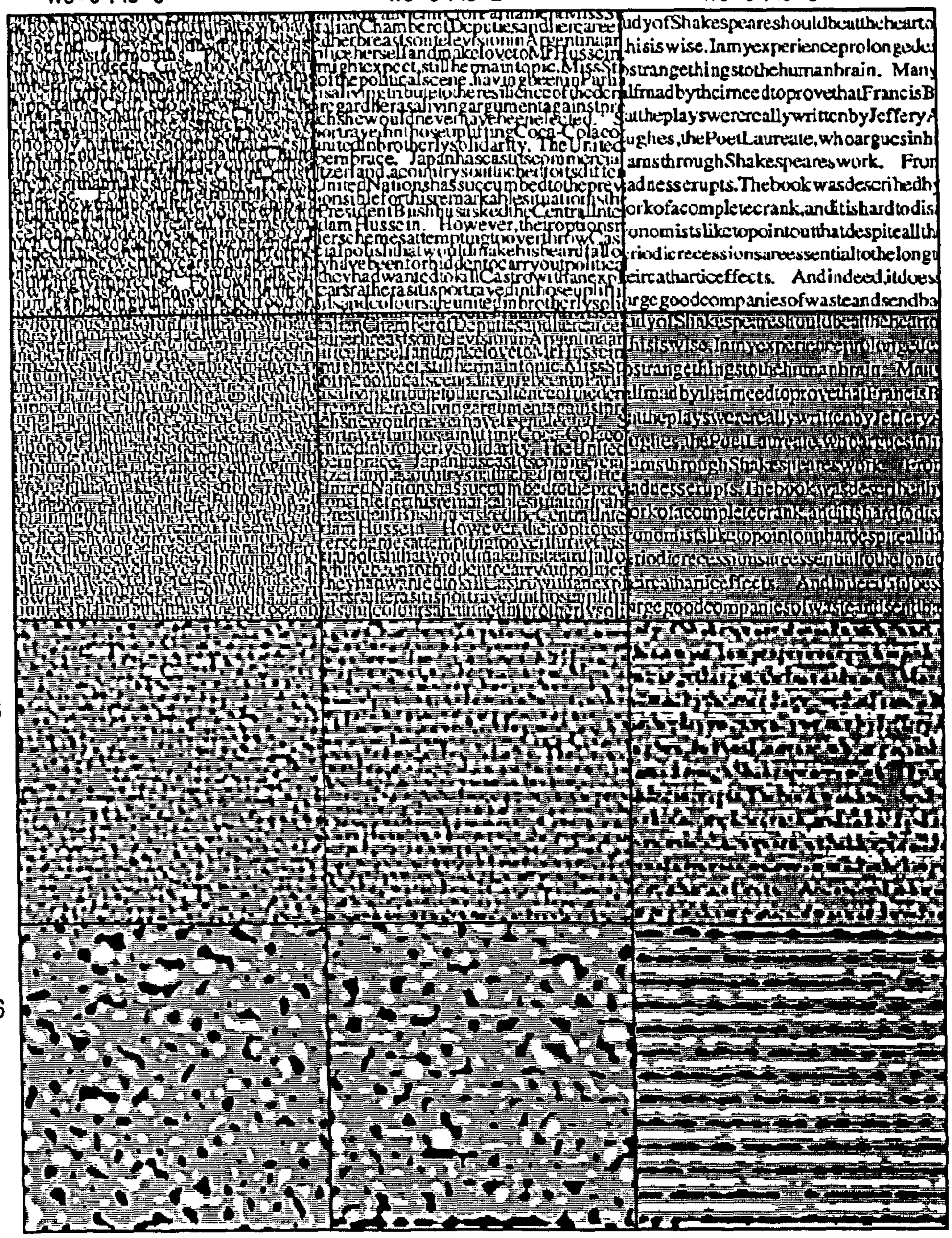


Figure 5.2a. A MIRAGE description of an example of a page of text used in computational analysis 4 represented as positive (light 'blobs') and negative (dark 'blobs') regions. The typographical parameter of line spacing varies in each column. Line spacing in left column is 0pts, middle column =2pts and right column = 8pts. Word spacing in this illustration is always 0pts. (Word spacing in Fig. 5.2b = 2pts and in Fig. 5.2c = 4pts). Top row is original text example. Second row is the representation at spatial scale = 1 pixel (filter s.d.); Third row is representation at spatial scale = 3 pixels; Bottom row is representation at spatial scale = 6 pixels.

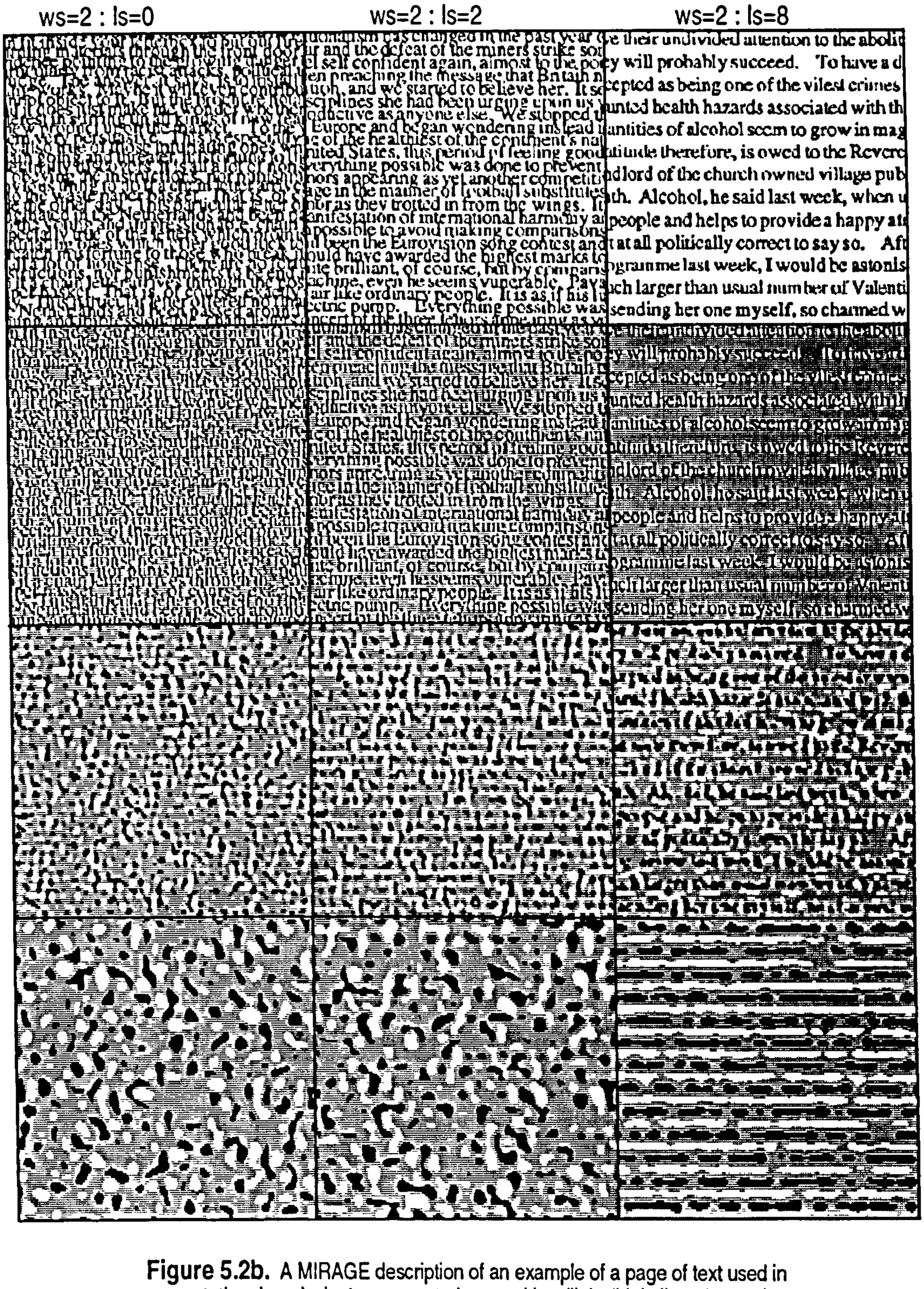


Figure 5.2b. A MIRAGE description of an example of a page of text used in computational analysis 4 represented as positive (light 'blobs') and negative (dark 'blobs') regions. The typographical parameter of line spacing varies in each column. Line spacing in left column is 0pts, middle column =2pts and right column = 8pts. Word spacing in this illustration is always 2pts. (Word spacing in Fig. 5.2a = 0pts and in Fig. 5.2c = 4pts). Top row is original text example. Second row is the representation at spatial scale = 1 pixel (filter s.d.); Third row is representation at spatial scale = 3 pixels; Bottom row is representation at spatial scale = 6 pixels.

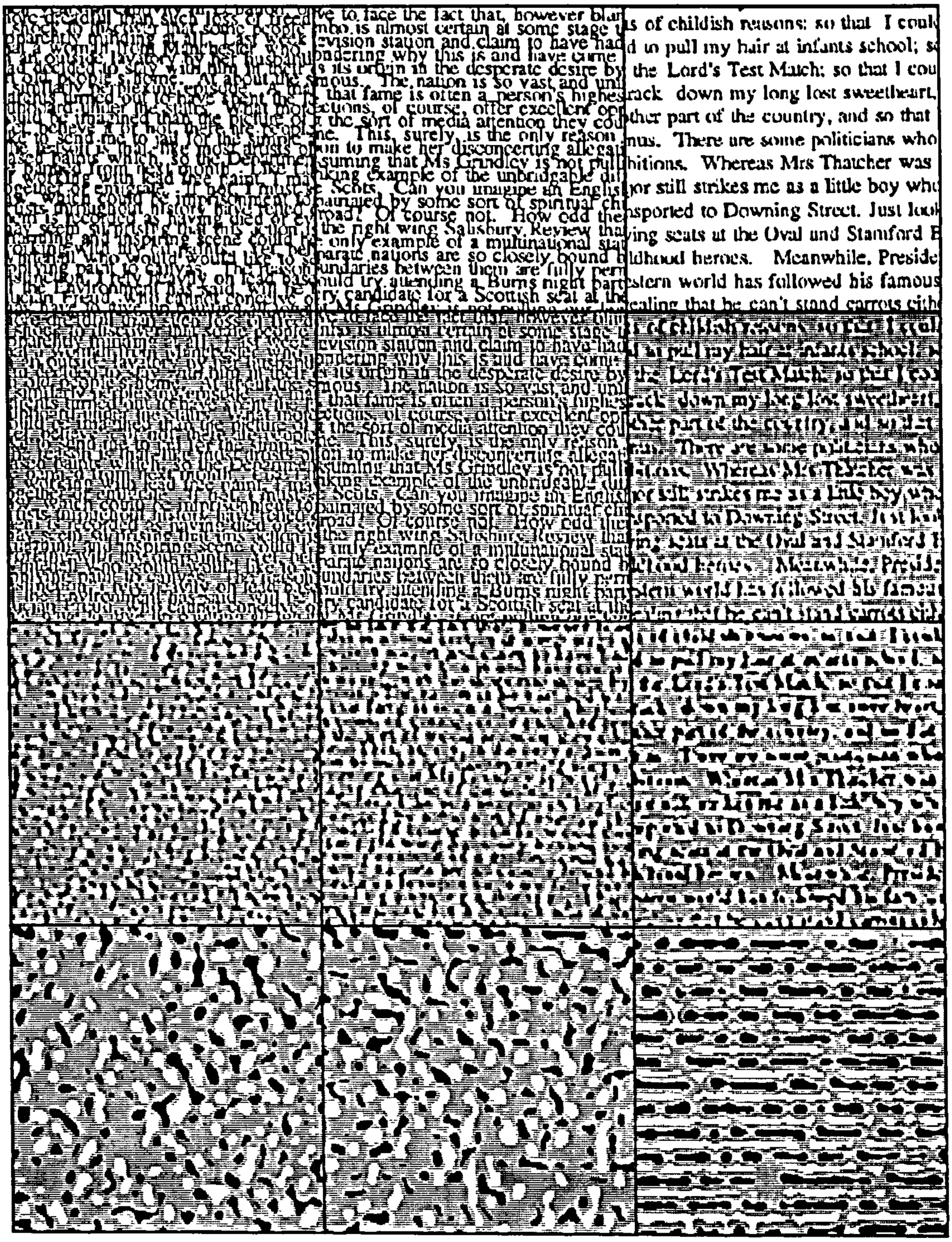


Figure 5.2c. A MIRAGE description of an example of a page of text used in computational analysis 4 represented as positive (light 'blobs') and negative (dark 'blobs') regions. The typographical parameter of line spacing varies in each column. Line spacing in left column is 0pts, middle column =2pts and right column = 8pts. Word spacing in this illustration is always 4pts. (Word spacing in Fig. 5.2a = 0pts and in Fig. 5.2b = 2pts). Top row is original text example. Second row is the representation at spatial scale = 1 pixel (filter s.d.); Third row is representation at spatial scale = 3 pixels; Bottom row is representation at spatial scale = 6 pixels.

Figure 5.3

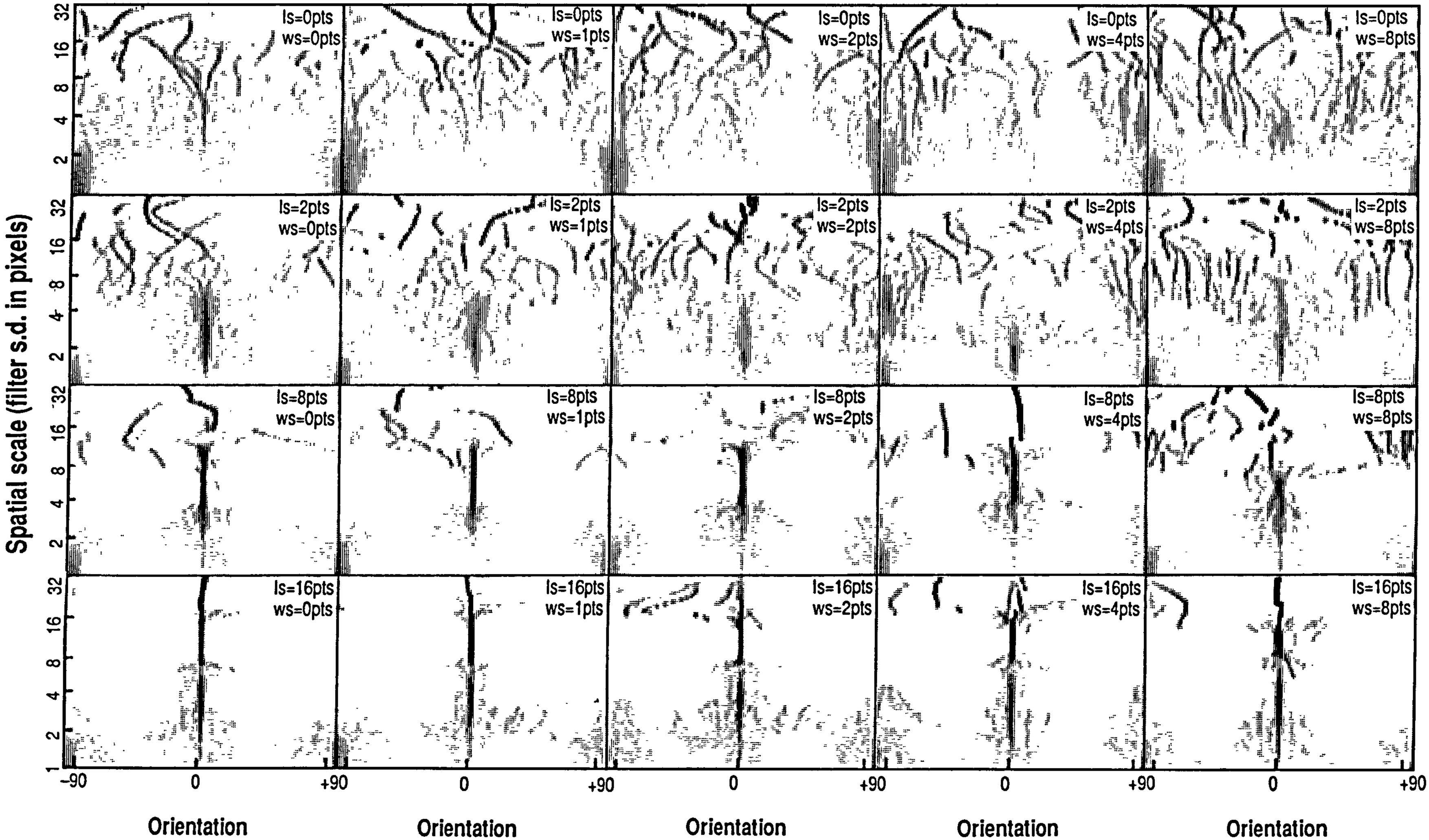
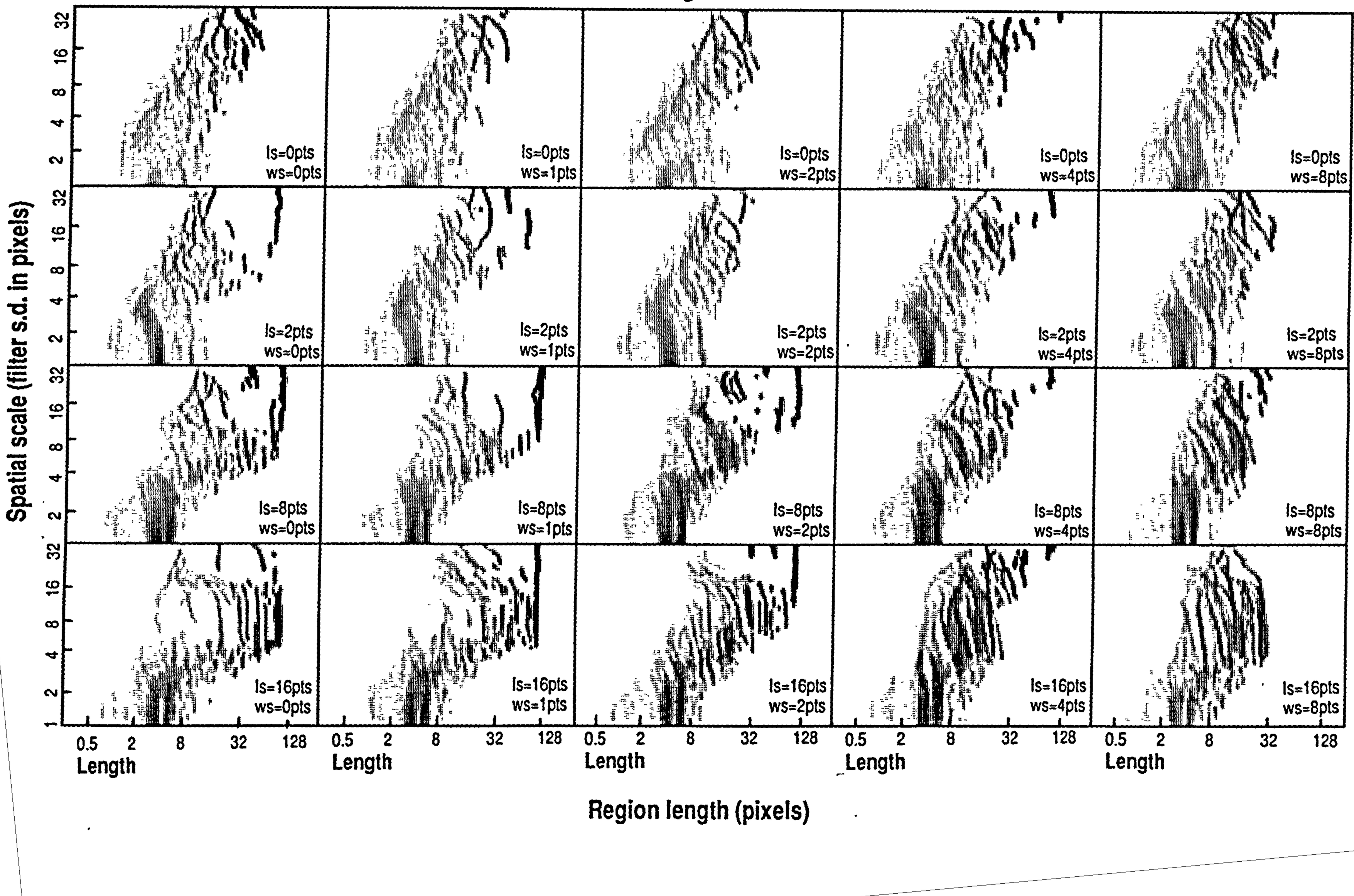

ls=0pts ws=0pts
ls=0pts ws=1pts
ls=0pts ws=2pts
ls=0pts ws=4pts
ls=0pts ws=8pts
ls=2pts ws=0pts
ls=2pts ws=1pts
ls=2pts ws=2pts
ls=2pts ws=4pts
ls=2pts ws=8pts
ls=8pts ws=0pts
ls=8pts ws=1pts
ls=8pts ws=2pts
ls=8pts ws=4pts
ls=8pts ws=8pts
ls=16pts ws=0pts
ls=16pts ws=1pts
ls=16pts ws=2pts
ls=16pts ws=4pts
ls=16pts ws=8pts
Spatial scale (filter s.d. in pixels)
1
2
4
8
16
32
−90
0
+90
Orientation

Figure 5.4


Spatial scale (filter s.d. in pixels)
1
2
4
8
16
32
ls=0pts ws=0pts
ls=0pts ws=1pts
ls=0pts ws=2pts
ls=0pts ws=4pts
ls=0pts ws=8pts
ls=2pts ws=0pts
ls=2pts ws=1pts
ls=2pts ws=2pts
ls=2pts ws=4pts
ls=2pts ws=8pts
ls=8pts ws=0pts
ls=8pts ws=1pts
ls=8pts ws=2pts
ls=8pts ws=4pts
ls=8pts ws=8pts
ls=16pts ws=0pts
ls=16pts ws=1pts
ls=16pts ws=2pts
ls=16pts ws=4pts
ls=16pts ws=8pts
0.5
2
8
32
128
Length
Region length (pixels)

5.1.3 Discussion

This analysis examined the relationship between page-level typographical arrangements (word spacing and line spacing combinations) and how these influenced the ability of the model of vision to extract and represent (make available) particular features of the text. The analysis revealed that the effects of word and line spacing were not separable. At conventional typographical arrangements, the model extracted a set of features which provided a representation of the scale (size) of the text and its orientation (orientation of the lines), and the information required to segment the words on each line in the orientation and length of both positive regions (representing line spacing) and negative regions (representing word length on each line) at coarse spatial scales (8–6 pixels). As spatial scale decreased to 3 pixels, word spacing was made explicit in the positive regions occurring in between each word. As spatial scale decreased further to the finest scales (1–2 pixels) regions corresponding to the individual letters themselves were extracted. Note that as the spatial scale decreased, so the un-grouping of regions at one scale led to the emergence of a new set of regions over the same spatial extent (position) extended by the previous, larger, region. So, line regions became un-grouped so as to be represented as a set of word regions. These word regions in turn became un-grouped into a set of letter regions.

Importantly, it was found that, as word and line spacing departed from their conventional typographical arrangements, the relative proximity of neighbouring lines and words resulted in inappropriate grouping and subsequent un-grouping of these text elements, so that important features which may be used in the visual processing of text (on the basis of the findings of Experiments 1–4), such as negative regions representing whole word shape were no longer made available at the scale at which they were previously.

The analysis also shows that, at this page-level of description, the interaction between word spacing and line spacing suggests that there may be typographical arrangements at which the features corresponding to word breaks represented by positive regions occurring at word spaces may be the only features, or information, available to allow word segmentation to be performed.

The findings of this analysis yield definite predictions for human text processing performance. The simplest, and most obvious, is that for a given line spacing, the word spacing required to segment the words in text should be predicted by the model's ability to represent word length in the negative regions at spatial scale=6 pixels. Furthermore, the model also now predicts that as word spacing and line spacing are varied, the positive vertically oriented regions corresponding to word breaks might, under the previously unexamined conditions of variable line spacing, also be used for word segmentation. If this

were the case, then at line spacing of 0–2pts, word spacing required should be between 1–2pt, as positive regions corresponding to word breaks were made available, but negative regions corresponding to whole words were not. On the other hand, on the basis of the findings of Experiment 4, this, and previous computational analyses, when line spacing (ls) =8pts, word spacing required should be at most 1pt or a little less (approximately 0.8pts: Experiment 4). This is because the negative regions corresponding to the length of whole words are made available at this word spacing at spatial scale = 6 pixels. When line spacing is wider than normal (ls=16pts), the model suggests that the word spacing required should be between 1–2pts. These predictions are tested in Chapter 6.

6

Visual Processing of Text: Page-Level Effects

In the previous chapter, Computational Analysis 4 showed how the features extracted from pages of text by the model of vision was dependent upon combinative effects of word spacing *and* line spacing. Chapter 6 reports an experiment which aimed to establish whether the human visual processing of text is also affected by the same word and line spacing changes. More importantly, this allows testing of the predictions from the analysis performed in Chapter 5, that the sensitivity of the pattern of text processing performance to varying typographical arrangements can be described by the sensitivity of the model in providing a representation of specific features of the pages of text as a function of the same word spacing and line spacings.

6.1 Experiment 5: Word segmentation as a function of word and line spacing

Experiment 5 examines word segmentation performance as a function of the same word and line spacing combinations as those used in Computational Analysis 4.

6.1.1 Method

The methodology was similar to that employed in Experiment 4. Full details of this can be found in Chapter 4. Experiment 5 specific details are given.

(i) Text Images. Text was generated in Apple Truetype 12pt Times Roman, page-centred to ensure that word spacing was uniform. Text images were created and displayed from a program developed at the University of Stirling by R. J. Watt which generated and displayed stimuli with the required parameters. This obviated the need to first produce a hardcopy of text and digitise it in the previous manner. The probability statistics of the occurrence and position of letters defining each 'word' string were similar to those found in English

(developed from Walker, 1987). As with Experiments 1 and 4, the *reference* page set had a mean word length of 4 letters and the *cue* page set had a mean word length of 7 letters. The standard deviation was 1.5 letters, and the range was 10 letters (1–11). The text had the same 2 parameters as that used in Computational Analysis 4: word spacing and line spacing. Therefore, line spacing (ls), which was defined as the spacing between the bottom of the baseline of one line to the x-line of the line beneath that line had 4 levels: 0, 2, 8 and 16pts. The word spacing (ws) required to perform the task was the dependent measure.

(ii) Procedure. Subjects were presented with 2 500x500 pixel text images which simultaneously appeared at either side of the centre of the display (*i.e.* one text image appeared to the left of centre the other appeared to the right of centre) for 1000msec. At stimulus offset, the text was replaced immediately by a mask of high contrast random noise for 1000msec. If the subject did not make a response by the mask offset the same text was re-displayed once again, and this sequence repeated until a response was made. After a response had been made, a 500msec pause occurred before 2 new images were generated and displayed. In a two alternative forced choice (2AFC) procedure the subject's task was to decide which page of text had the longest mean word length.

The procedure was based on the adaptive method of constant stimuli (APE) used in previous experiments, which selected a range of word spacings over 64 presentations at appropriate points on the psychometric function. The threshold word spacing required to perform the task was defined as the standard deviation of the Gaussian response error distribution determined by probit analysis. APE generated a range of 8 stimulus levels. Thus for a stimulus value of 8, text with a word spacing = 3 pts was displayed and the subject was able to perform at 100% correct. At a value of 0, text with zero word spacing (0 pts) was presented, making the task impossible, and the subject performed at chance levels.

Text was displayed at a range of 256 greylevels on a 21 inch Trinitron display (ProNitron model 80.21) with a refresh rate of 80Hz and a resolution of 72 dpi, controlled by a Macintosh IIcx computer. All other aspects of the viewing conditions are described in Section 3.1.

(iii) Subjects. 3 subjects aged between 21-32 years old participated. All three were experienced psychophysical subjects. SCD and PMC were volunteers drawn from the CCCN and were naive about the purpose of the experiment. PMC had participated in Experiments 1–3. The other, SJE, was the author. PMC was emmetropic. SCD and SJE were corrected hypermetropes.

6.1.2 Results

Inspection of Figure 6.1 clearly shows an interaction between line spacing and the word spacing required for word segmentation. For all 3 subjects, threshold word spacing required for word segmentation decreased as line spacing increased up to a critical line spacing (8pts), above which the word spacing required *increased* with increasing line spacing.

Most importantly, the word spacing required to reach threshold word segmentation performance was within the predictions of the model, made at the end of Chapter 5. At line spacing (ls) =8pts the word spacing required to reach threshold word segmentation performance was less than 1pt (PMC=0.79pts; SCD=0.76pts; SJE=0.86pts), in reasonable agreement with the 0.8–1.0pts that was predicted on the basis of the response of the model in Computational Analyses 3 and 4. When line spacing was very narrow (ls=0–2pt) the word spacing required increased to within that predicted, of above 1pt but at or below 2pts (for ls=0pts: PMC=1.99pts; SCD=1.87pts; SJE=1.45pts).

When line spacing was very wide (ls=16pts), the predictions of the model do not hold so well. It was predicted from Computational Analysis 4 that information required to perform this task was unavailable until above 1pt and was only really fully represented by the time line spacing had reached 2pts. However, subject PMC reaches threshold performance at 0.98pts word spacing. Threshold word segmentation performance for SCD and SJE was more consistent with the predictions of the model, at 1.32pts and 1.4pts respectively.

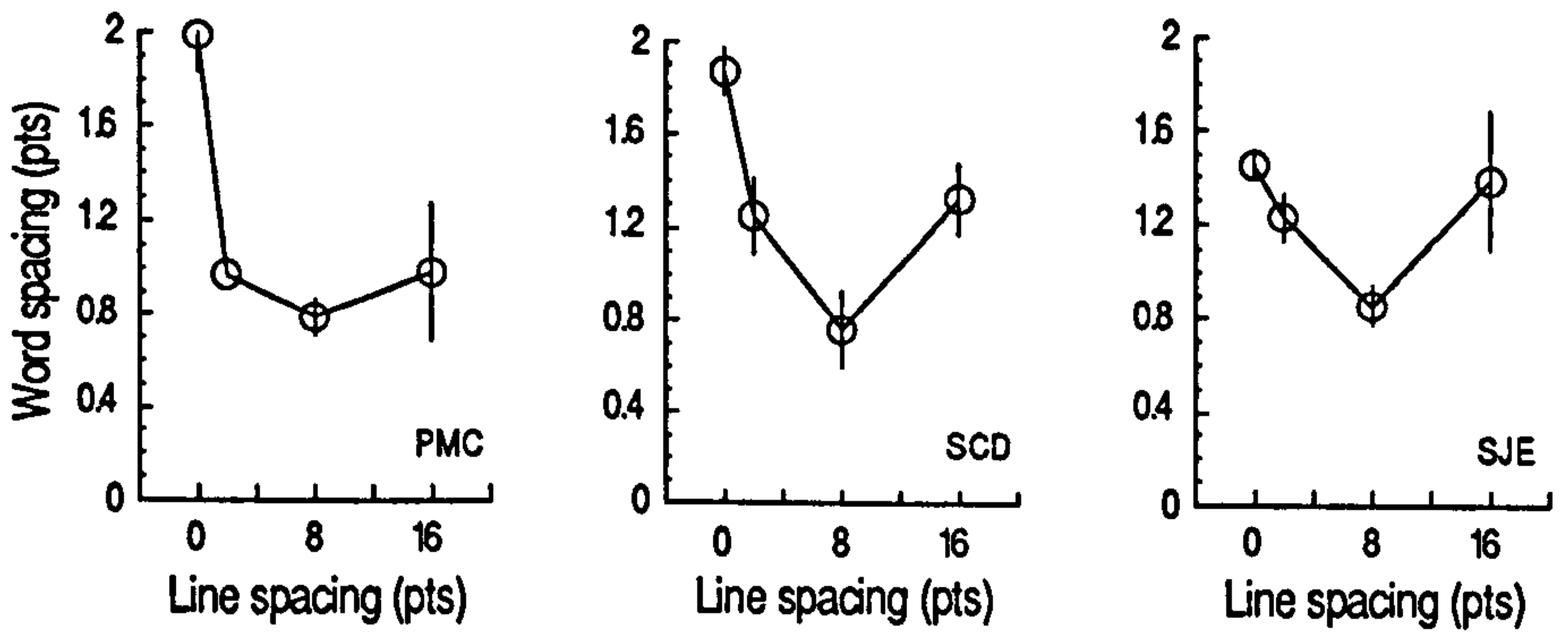


Figure 6.1 Threshold word segmentation performance of 3 subjects (determined in a mean word length estimation task) as a function of word and line spacing. Ordinate is the word spacing (in points) required to reach threshold word segmentation performance (P[c]=83%). Abscissa is the line spacing of the text in points (pts). Each data point is the mean of 3 runs. Each run was the mean of 64 measurements. bars show standard error.

6.2 Discussion

The pattern of word segmentation performance in Experiment 5 was shown to be dependent on the effects of both word spacing and line spacing. Moreover, there was found to be an interaction of effects of both on performance. These findings are important in terms of the predictions generated by the model from Computational Analysis 4 (Chapter 5). In Analysis 4, it was found that both word spacing and line spacing had a determining effect on the features the model was able to extract and make available from text. Furthermore, it also showed that the effects of word spacing and line spacing on the features extracted were not separable. The spatial scale and features made explicit at each scale as a function of word spacing and line spacing combinations led to the prediction that threshold word spacing should decrease from line spacing = 0pts to reach a minimum at line spacing = 8pts and then increase again for line spacing = 16pts. This prediction was borne out by the pattern of text processing performance found in this experiment, thus providing further support for Hypothesis 2, made in Chapter 2.

Finally, the findings raise a further, important, question about the visual processing of text, and one which has particular salience to the reading process. According to MIRAGE, visual processing is a time-course of processing which operates on the basis of a coarse-to-fine spatial scale of visual analysis (Watt, 1987). This component of the model predicts a number of specific consequences for the visual processing of text which are examined in Chapter 7.

7

The Time-Course of the Visual Processing of Text

When the demands of a visual processing task require that spatial relationships of elements in an image be computed, Watt (1987) has argued that the time required is determined by the need to progressively switch-out filters (coarsest first) until the level of detail required by the task is revealed. That is, there is a time-course of visual processing which is based on a coarse-to-fine spatial scale of visual analysis (see Chapter 1, Section 1.4 for further discussion). Given the observed and modelled spatial scale dependency of the processing of text, a time-course of visual processing based on a spatial scale of visual analysis predicts several consequences for the visual processing of text during reading.

This is an aspect of the visual processing of text which is particularly relevant to measures of reading. An important measure of reading is the time it takes to read. The time taken to read is generally considered to be an index of legibility, or a measure of the efficiency of the visual processing of text. What must be meant by this is the component(s) of reading which are determined by the time needed to extract the information required to perform the reading task: the time-course of the visual processing of the text.

There are two obvious and simple instances of the consequences the time-course of visual processing should have for the visual processing of text which can be readily examined. Both are based on the prediction of the model that the time required to perform a text processing task should be dependent upon a time-course of visual processing in which the visual system should have reached a level of visual analysis (based on a scanning from coarse-to-fine spatial scales) at which the information needed to support those tasks (based on the findings of Experiments 1–5 and Computational Analyses 1–4) is represented. Thus, the first predicted consequence is that the time required to perform word segmentation and the time required to identify letter position: two visual processing tasks identified as using information contained at different spatial scales, should be different. Moreover, the time

required to perform these tasks should be consistent with the predictions of the model outlined above. This was examined in Experiments 6a and 6b.

The second instance comes from the finding that changes in the word spacing needed to perform the word segmentation task as a function of line spacing (Experiment 5) was predicted by the model (Computational Analysis 4) on the basis of finding in Computational Analysis 4 that the spatial scale containing the information required to perform the task varied with word spacing and line spacing in a manner which could describe the pattern of psychophysical performance. Thus, the time required to perform the word segmentation task should vary in a manner consistent with a time-course of visual processing based on a coarse-to-fine spatial scale of analysis. This was explored in Experiment 7.

7.1 Experiment 6: time-course of word segmentation (6a) and letter position identification (6b)

Experiments 6a and 6b examined the exposure duration required to reach threshold word segmentation performance and letter position identification performance, respectively.

7.1.1 Method

i) Text Images. For both Experiments 6a and 6b text was generated in Apple Truetype 12pt Times Roman, page-centred to ensure that word spacing was uniform. The probability statistics of the occurrence and position of letters defining each 'word' string were similar to those found in English (developed from Walker, 1987). As with Experiment 5, the *reference* page set had a mean word length of 4 letters and the *cue* page set had a mean word length of 7 letters. The standard deviation was 1.5 letters, and the range was 10 letters (1–11). Line spacing (ls), which was defined as the spacing between the bottom of the baseline of one line to the x-line of the line beneath that line was set to 8pts. Word spacing (ws) was set to 3pts.

In Experiment 6b, mean word length in both pages of text displayed was always the same, at 5 letters. The *reference* text had the probability statistics as the text in Experiments 6a. The *cue* text had a 0.3 greater probability of a vowel occurring in the body of a word than in the reference text. That is, any letter position other than word boundary (first and last letter) position.

(ii) Procedure. Subjects were presented with two 500x500 pixel text images which simultaneously appeared at either side of the centre of the display (*i.e.* one text image appeared to the left of centre the other appeared to the right of centre), initially for a duration of 1000msec. At stimulus offset, the text was replaced immediately by a mask of

high contrast random noise for 1000msec. If the subject did not make a response by the mask offset the same text was re-displayed until a response was made. Two new images were then generated and displayed. In a two alternative forced choice (2AFC) procedure the subject's task in Experiment 6a was to decide which page of text had the greatest mean word length. In Experiment 6b the task was to decide which page of text had the greatest number of vowels in the bodies of the words.

The procedure was based on the adaptive method of constant stimuli (APE) used in previous experiments, which selected a range of exposure durations over 64 presentations to obtain a psychometric function. The threshold exposure duration required to perform the task was defined as the standard deviation of the Gaussian response error distribution determined by probit analysis. APE generated a range of 8 stimulus levels. Thus, for a stimulus value of 8, text was displayed for 1000msec and the subject was able to perform at 100% correct. At a value of 0, exposure duration was 1msec (text was effectively not presented) making the task impossible, and the subject performed at chance levels. Therefore there was a monotonic variation in exposure duration of 125msec (effectively a minimum of one screen refresh).

(iii) Subjects. The same subjects who took part in Experiment 5 also participated in Experiments 6 and 7.

7.1.2 Results and discussion

Mean exposure durations required to segment words and identify letter position are given for each subject in Table 7.1. It shows clearly that letter position identification required a much greater exposure duration than did word segmentation.

Threshold Exposure Duration (msec)

	Word segmentation	Letter position ident.n
PMC	30.0 (8.2)	550.0 (94.0)
SCD	63.0 (4.1)	609.0 (47.0)
SJE	30.8 (0.7)	992.2 (101.0)

Table 7.1 Mean exposure duration (milliseconds) required to perform 2 text processing tasks: word segmentation and letter position identification. Data is for each of 3 subjects: SJE, PMC and SCD. Data represents mean of 3 runs. Each run contained 64 measurements. Figures in brackets are standard errors.

An explanation for the basis of these text processing times, and the differences between them, may be sought by examining data from Watt (1987) concerning the time-course of visual processing. Watt found that although spatial resolution ability was unaffected by

exposure duration, spatial position discrimination depended upon the exposure duration of the stimulus. Watt showed how this could best be explained within the existing Watt & Morgan (1985) model of early vision if the visual system 'switched out' the largest filter active over time (see Chapter 1 or Watt, 1988 for further discussion). Modelling the data in this manner allowed the calculation of the range of filter sizes active in the visual system at any given time after stimulus onset.

This dynamic component to the Watt & Morgan (1985) MIRAGE model based on a coarse-to-fine scale visual analysis over time allows spatial position to be calculated economically (see Chapter 1, Section 1.4 for further explanation of this argument). Word segmentation and letter position identification tasks in these experiments are examples of such visual tasks.

Experiment 1 showed that early visual processing used information to perform word segmentation which was contained at the coarse spatial scale of 6 pixels (filter s.d.). Similarly, Experiment 2 suggested that the information used in identifying letter position was contained at a much finer spatial scale of 1–2 pixels. Given the foregoing argument, exposure duration required for word segmentation and letter position identification should be predicted by different time-courses of processing. On the basis of Watt's (1987) data, after approximately 30–40msec (the threshold for word segmentation) the size of the largest filter active in the visual system (measured as the standard deviation of the filter) should be approximately 6 pixels (30 arcmin), given the viewing distance and size of text. It can be seen from Figure 7.1, which shows the comparison of the data from Experiment 6 with data from Watt (1987) based on the model's predictions, that this is what was found.

To summarise, the results of Experiment 6 are consistent with the predictions of the model. That is, the exposure duration required to perform each task is predicted by a time-course of visual processing based on a coarse-to-fine spatial scale of analysis, in which the time required to reach the spatial scale of analysis containing the information needed to support these tasks would have been reached.

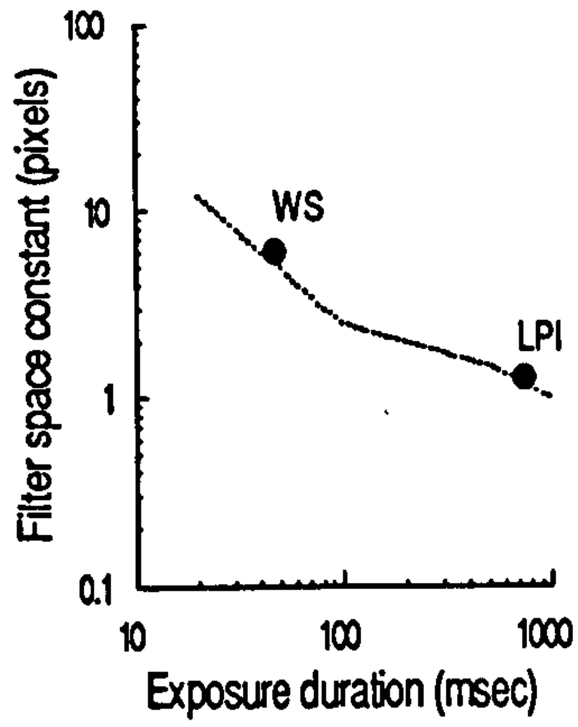


Figure 7.1. Measured exposure duration required to reach threshold performance level on two text processing tasks: word segmentation (WS) and letter position identification (LPI), for subject PMC, shown as solid data points. Data is plotted against data adapted from Watt (1987) of the calculated value of the largest spatial filter (space constant–or scale–in pixels) estimated to be active as a function of time (in milliseconds), shown as a dotted line.

7.2 Experiment 7: Word segmentation duration as a function of word and line spacing

The psychophysical data of Experiment 5 was predicted on the basis of the word and line spacing arrangements required make available, according to the performance of the model in Computational Analysis 4, features, or information, for word segmentation. Word and line spacing arrangements which resulted in changes in performance was consistent with a loss of regions representing whole word length at a coarse spatial scale (filter s.d. = 6 pixels). It was also found that when these regions could not be extracted by the model, the model successfully predicted that threshold performance should be reached, *providing* that word and line spacing still allowed information representing word spaces to be extracted at a finer spatial scale (filter s.d. = 3 pixels).

If interpretation of the data from these and the other, earlier, experiments and computational analyses has been correct, then the exposure duration required for word segmentation should vary as word and line spacing varies in a manner which is consistent with different time-courses of visual processing required to reach the different spatial scales of analysis in which appropriate features used to support word segmentation are made available. Specifically, from the findings of Computational Analysis 4, the following would be expected:

(i) When line spacing =0–2pts, providing word spacing is between 1–4pts the required information is available at spatial scale =3 pixels. Therefore, exposure duration required should be in the order of 100–150msec.

(ii) When line spacing =0–2pts and word spacing = 8pts, the required information is available at spatial scale = 4 pixels. Therefore, exposure duration required should be in the order of 80msec.

(iii) When line spacing =8–16pts and word spacing = 1–4pts, the required information is available at a spatial scale = 6 pixels. Therefore the exposure duration required should be approximately 30msec.

(iv) When line spacing = 16pts and word spacing = 8pts the required information is available at spatial scales = 6–8 pixels. Therefore the exposure duration required should be in the order of 20–30msec.

These predictions are tested in this experiment.

7.2.1 Method

This was the same as Experiment 6 (see Section 7.1.1) except that the text had exactly the same characteristics as Experiment 5 (see Section 6.1.1).

7.2.2 Results

The mean exposure duration required to reach performance threshold on the word segmentation task as a function of word and line spacing is shown for each observer in Figure 7.2. Three features of the data warrant comment. First, threshold exposure duration required varied as a function of word spacing. Exposure duration required for a given line spacing decreased as word spacing increased between 1–4pts for each observer. The exception to this was the pattern of performance found for ws = 8pts and ls = 8pts, where exposure duration required was slightly higher than that required for ws=4pts and ls=8pts.

Second, exposure duration required varied as line spacing varied. The pattern of performance shown in Figure 7.2 shows that as line spacing increased from 0–8pts, threshold exposure duration for a given word spacing declined, reaching a minimum for text with line spacing =8pts, word spacing =4pts (mean threshold exposure duration: SJE=30.8msec; PMC=33.6msec; SCD=44.6msec). As line spacing continued to increase to 16pts, the exposure duration required also increased.

The third feature to note is that the pattern of data shows that the exposure duration required to reach threshold performance level was not determined solely by word spacing or line spacing but was determined by effects of both word *and* line spacing. That is, the effects of word and line spacing were not separable.

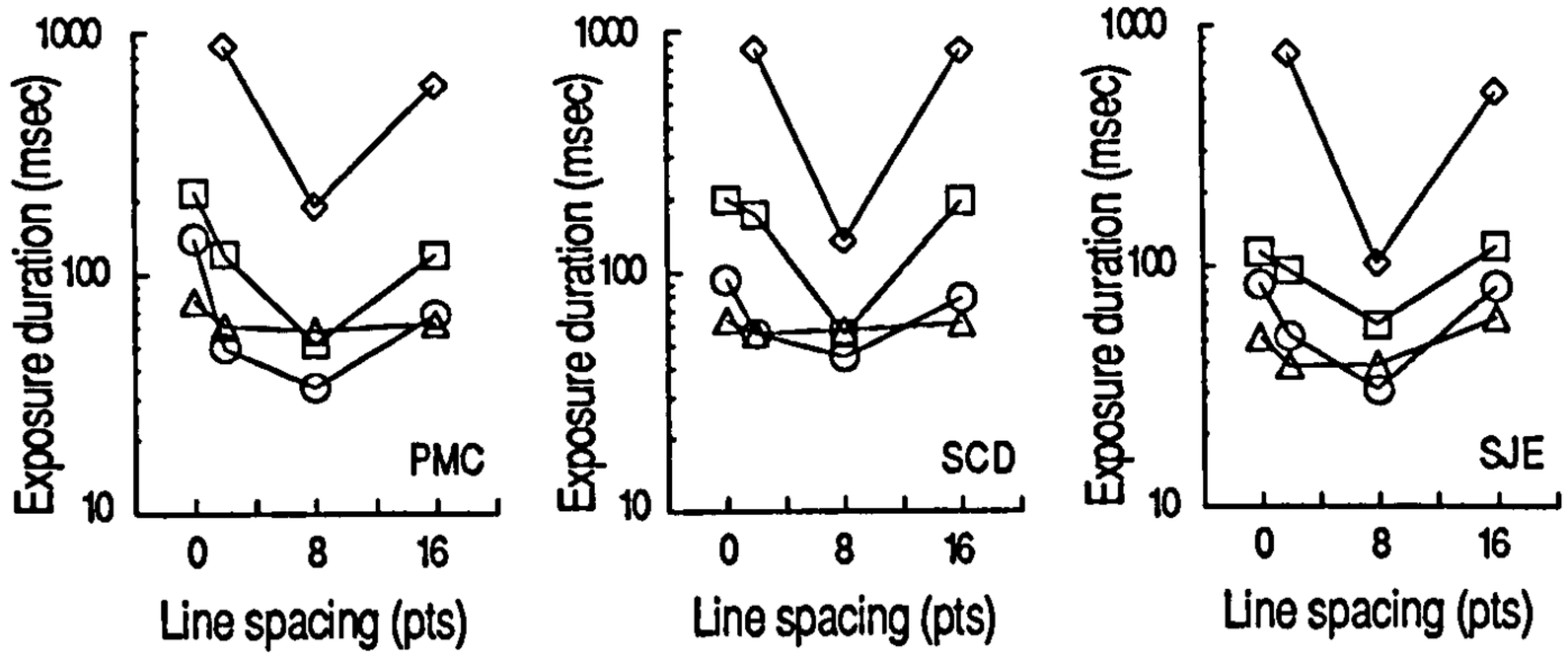


Figure 7.2. Exposure duration in milliseconds (ordinate) required to reach threshold word segmentation performance as a function of word spacing (different symbols) and line spacing (abscissa) in points, obtained in mean word length estimation task. Each data point is the mean of 3 runs. Each run is the probability of 83% correct performance estimated from 64 measurements. Standard error is not shown for reasons of clarity. It was typically less than 10% of the mean exposure duration and always less than 20%. Diamonds: ws=1pt; squares: ws=2pts; circles: ws=4pts; triangles: ws=8pts.

7.2.3 Discussion

In pursuit of an explanation of the pattern of results found in Experiment 7, it is interesting to note first, that the pattern of performance varied as a function of word and line spacing in a very similar way to the pattern of performance as a function of word spacing and line spacing obtained in Experiment 5. This is worth noting because the threshold word spacing required for word segmentation found in Experiment 5 was predicted by the word spacing at which particular features suggested by the modelling to be used for word segmentation were made available as word and line spacing varied. More importantly for the present discussion is that this information was found to be contained at different spatial scales as a function of word spacing and line spacing.

A comparison of Experiments 5 and 7, illustrated in Figure 7.3, shows how the changes in the word spacing required for word segmentation as a function of line spacing and the exposure duration required for word segmentation both vary similarly with line spacing.

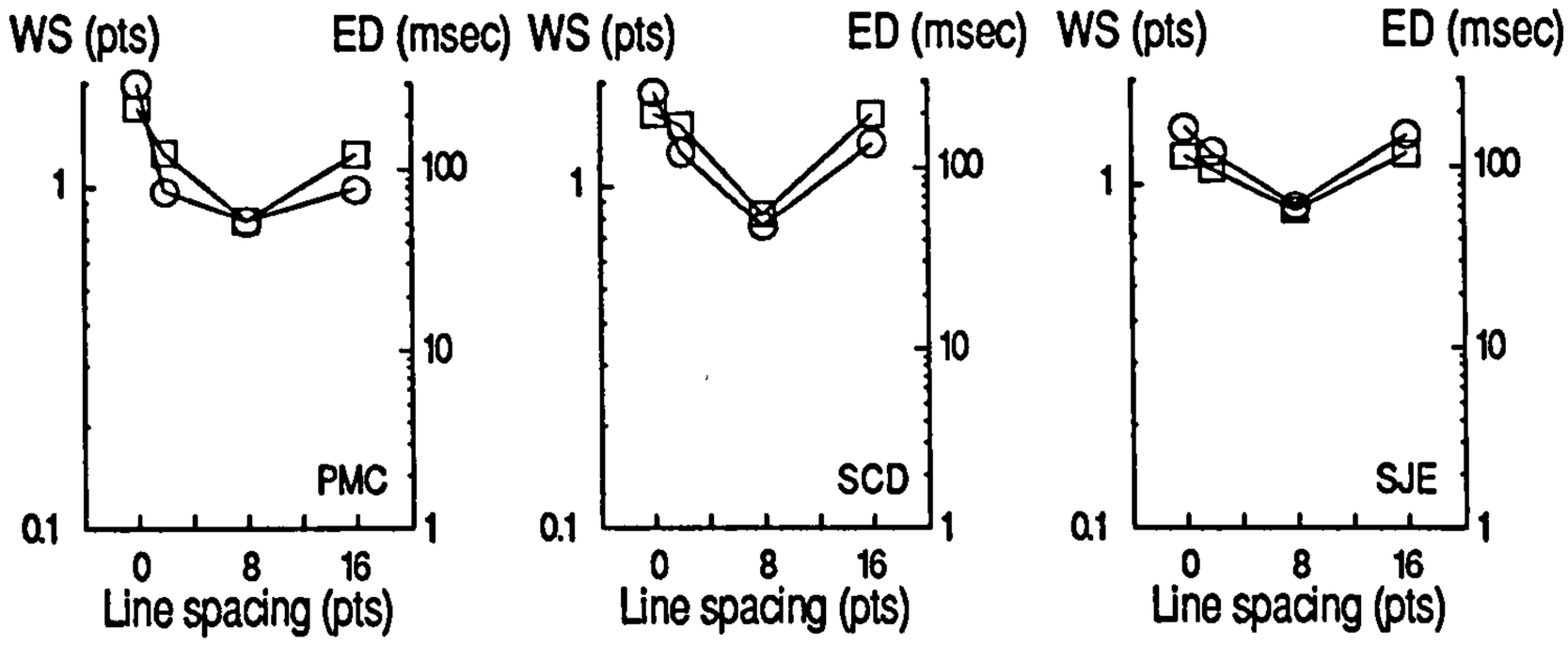


Figure 7.3. Comparison of Experiment 5 (circles) and Experiment 7, word spacing = 2pts (squares) data for 3 subjects, PMC, SCD and SJE. Left ordinate (WS pts) is word spacing (in points) required to reach threshold word segmentation performance. Right ordinate (ED msec) is exposure duration (in milliseconds) required to reach threshold performance level on same task.

A Pearson's product-moment correlation between the data of Experiments 5 and 7 (word spacing =2pts; the most appropriate word spacings of Experiment 7 since word spacing required in Experiment 5 was between 1–2pts) confirms the goodness of fit between the two sets of data ($r = 0.910$, n = 12, $p < 0.0005$) (an association between threshold exposure duration at ws = 4pts [Experiment 7] and threshold word spacing [Experiment 5] was also found [$r = 0.808$, n = 12, $p < 0.005$]). This suggests that the visual information which determined the pattern of word segmentation performance in Experiment 5 is likely to be the same as that which determined the exposure duration required to perform the same task, found in Experiment 7.

To try to discover the basis of the pattern of performance found in Experiment 7 it is necessary to compare the data of Experiment 7 with the predictions of the model made in Section 7.1. Figure 7.4 shows the comparison between the model and the psychophysical data. The model data is the exposure duration needed to reach the spatial scale of analysis at which the model (from Computational Analysis 4) revealed information hypothesised to be used for word segmentation (on the basis of the findings of Experiments 1 and 4) as a function of word spacing and line spacing. This exposure duration is based on an estimation of the time-course of visual processing based on a coarse-to-fine spatial scale of analysis, taken from the data of Watt (1987).

Figure 7.4 shows this comparison. There is a reasonable agreement between the model's predictions and the psychophysical data of Experiment 7, at least for line spacings between 0–8pts. The values of r in a Pearson's product-moment correlation suggests that there is a

reasonable fit between the data and the model for each word spacing, but because of the very low *n*, none of the fits are statistically significant. However, there is a tendency for the data at small word spacings (1–2pts) to depart from the predictions at line spacing = 16pts. It is difficult to account for this finding. Speculation might favour an explanation in terms of the smaller sample size (fewer words in each text image) from which word length estimation was made, making the psychophysical decision unreliable, requiring additional inspection time, but this would not explain why the departure from the predictions is not so great at wider word spacings. Further work is required to examine this problem properly.

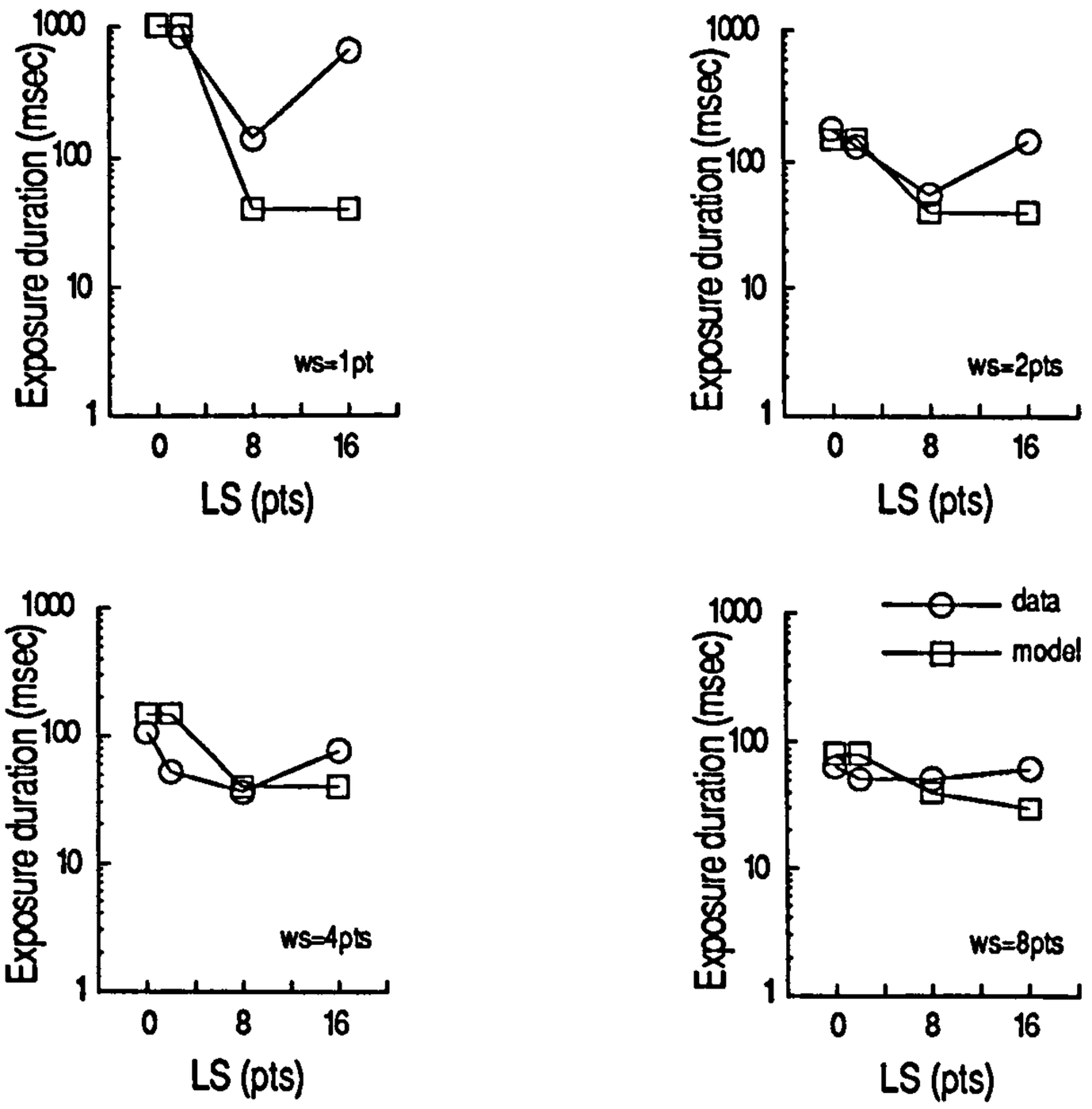


Figure 7.4. Comparison of Experiment 7 mean of 3 subjects data with the time-course predictions of the MIRAGE model, based on data from Watt (1987). Model predictions are estimated scale 'active' based on a coarse-to-fine spatial scale of visual analysis over time. Ordinate is exposure duration required to reach word segmentation performance threshold. Abscissa (LS pts) is line spacing in points.

To summarise, the exposure duration required to perform a visual processing of text task was able to be described by the model in terms of a time-course of visual processing based on a coarse-to-fine spatial scale of analysis. The findings of Chapter 7 therefore finds support for Hypothesis 3 (Chapter 2, Section 2.3): that the visual processing of text will have a time-course which is predicted by the model.

The findings of the comparison between the data of Experiment 7 and the response of the model of vision from Computational Analysis 4 are particularly relevant to measures of reading. By definition, reading speed must be determined, at least partly, by the time-course of visual processing of text. It is of interest therefore, to consider whether reading performance can be accounted for by the time required to perform one visual text processing task: word segmentation. This is the subject of Chapter 8.

8

The Visual Processing of Text and Reading

Reading performance (measured typically by reading speed) must be determined at least in part by the time-course of visual processing of text. It is this component which is attempted to be determined in reading rate experiments (*e.g.*, Tinker, 1965; Fisher, 1975; Bock, Monk & Hulme, 1993). However, as was discussed in Chapter 1, there are many processes which are involved in reading, only some of which are visual processes. There are others, such as linguistic processes (see Coltheart, 1987 for a discussion). In measuring reading rate alone, it is not possible to account for and explain the contribution of each of these processes in the effect they have on this measure. This was one of the reasons for having been critical in Chapter 1 about the use of measures of reading rate to study the visual processing of text in reading.

However, this chapter reports the findings of an experiment which examined how reading speed varied as a function of the same typographical parameters–word spacing and line spacing–manipulated in Experiment 7. It is argued that this is now worthwhile for two reasons. The first reason is that, because it is not possible to otherwise separate out the contribution of visual from other, non-visual, factors affecting reading, any measure of reading is only useful in terms of its relationship to a separately determined measure of the visual processing of individual tasks in reading. The second reason stems from the finding that at least one text processing task: word segmentation could (generally) be explained by the model of visual processing of text in terms of a time-course of visual analysis. This suggests that such a measure of the visual processing of text may now have been obtained. It might therefore be possible to estimate how the visual processing of text required for word segmentation influences the speed or efficiency of reading.

Experiment 8 examines how silent reading speed varies as word spacing and line spacing varies. The findings of this experiment are then compared to a model of the visual

processing task of word segmentation to establish how the model of visual processing can account for the visual processing of text in reading.

Experiment 8: Reading

The general procedure was to obtain a measure of silent reading rate for a range of word spacing and line spacing combinations identical to those used in Experiments 5 and 7. Silent reading was used because a pilot study revealed differences in reading rates between oral and silent reading; silent reading rate being more sensitive to changes in spacings. Note that this difference is contrary to previous findings by, for example, Legge *et al.* (1985) who found no difference between oral and silent rates.

8.1 Method

8.1.1 Text

The same text which was used in Computational Analysis 4 was used in Experiment 8. See Section 5.1 for details.

8.1.2 Procedure

Prior to the experiment start, subjects were informed that they were to read 40 passages of text, and that each page differed in its word spacing and line spacing. They were instructed to read each passage as quickly as possible, but carefully, as questions about the topic would be asked on half of all the passages; they would not be warned which passages.

Subjects were given 6 practice trials with pages differing in word and line spacing combinations to familiarise them with the type of layouts of pages they could expect. At the start of the experiment, the subject was presented with a page of text (order of presentation was varied randomly between subjects). The subject read silently for 10 seconds, at which time a tone signalled to the subject to start reading aloud. This procedure was used so that it could be determined where in the text the subject had read to, and therefore how many words had been read. On a randomly determined 50% of trials, the subject was asked one question about the topic per sentence read, unless they had not managed to read up to the end of the first sentence, in which case they were asked to recall verbatim what they had read. The criterion level for accepting a subject's data was that 2 out of 3 of all questions asked had to be answered correctly. No subject 's data was rejected on this basis.

Text was read from platform angled at 45 degrees to the table at which the subject sat, and was positioned 50cm away from the subject.

The mean number of words read in each typographical combination was calculated from each of 2 passages for each combination. This data was then converted to a reading rate, in words per minute, by multiplying the mean number of words read in each passage by 6.

8.1.3 Subjects

Twelve subjects participated. Ten were Psychology undergraduates who took part to fulfil a course requirement. The remaining two were CCCN members who took part voluntarily. All had normal or corrected-to-normal vision.

8.2 Results

Figure 8.1 shows the mean reading rate of 12 subjects as a function of word and line spacing, expressed in words read per minute. Both word spacing and line spacing had an effect on reading rate. For all word spacings apart from ws=0pt, increases in word spacing resulted in increases in reading rate at any line spacing, only up to the 'optimum' word spacing of 4pts. A further increase in word spacing to 8pts resulted in a decline in reading rate.

Increases in line spacing resulted in faster reading rates for every word spacing (except for ws=0pts) up to the 'optimum' line spacing of 8pts. A further increase in line spacing to 16pts resulted in a decline in reading rate. Thus, the conventional ('optimum') page-level typographical arrangement (ws=4pts; ls=8pts) resulted in the fastest reading rate [408 wpm]. A 2–way analysis of variance shows that there was a significant main effect of word spacing ($F_{(4,209)} = 166.2; p < 0.001$) and line spacing ($F_{(3,209)} = 160.94; p < 0.001$) on reading rate.

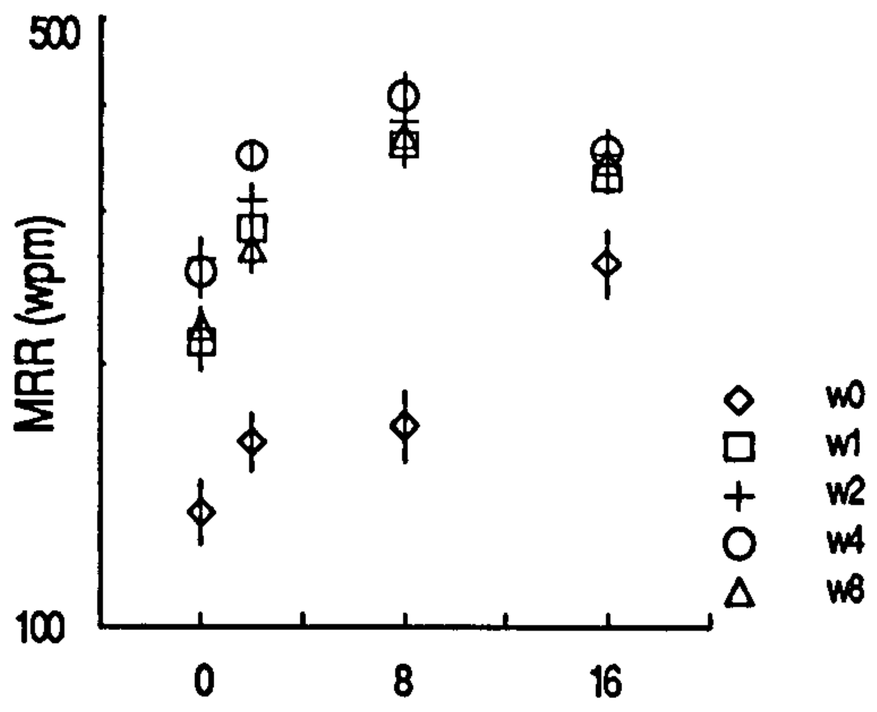


Figure 8.1 Mean silent reading rate (MRR) in words per minute (wpm) as a function of word spacing and line spacing. Each data point is the mean of 12 reader's data. Each reader's data is the mean of 2 trials per each word-line spacing combination. Different symbols represent each word spacing, from 0(w0) to 8 (w8)pts. Bars show standard error.

Furthermore, the effects of word spacing and line spacing were not separable: there was a significant interaction between word spacing and line spacing on reading rate ($F_{(12,209)} = 8.10$; $p < 0.001$).

8.3 Discussion: modelling the visual processing of text in reading

It is possible to compare and model the relationship between the time-course of word segmentation performance observed in Experiment 7 and the mean reading time per word found in Experiment 8. Reading time per word (RTPW) may be expressed as:

$$RTPW = at + b$$

where RTPW is in msec, '*t*' is an estimate of the word segmentation component of reading (how long it takes to segment each word in reading) in the reading process based on the data of Experiment 7, 'a' is a constant, and 'b' is "everything else" in the reading process (it is not possible at this stage to separate out any other text processing tasks or non-visual components). The data is modelled given the following parameters: *t* (in all ws) = mean exposure duration (msec) to reach threshold word segmentation performance in Experiment 7; a=1 (ws=2, ws=4, ws=8) and 0.25 (ws=1); b =90msec (ws=1), 70msec (ws=2), 110msec (ws=4) and 150msec (ws=8).

This model provides a surprisingly good fit to the actual RTPW found in Experiment 8 data, as shown in Figure 8.2. A Pearson's product-moment correlation confirms the goodness of fit between the model and observed RTPW ($r = 0.672$, $n = 16$; $p < 0.005$).

It is difficult to provide an adequate explanation of why the 'everything else' component of the model (b) was required to vary between 70–150msec, simply because there will be so many variables which contribute to 'b'. The actual times of the 'everything else' component of the model, at between 70–150msec are in agreement with other findings that when segmentation of words in a page of text is not necessary, as in rapid serial visual presentation (RSVP) techniques, the time required to read is in the range 50–150msec (Gilbert, 1959; Ward & Juola, 1982; Rubin & Turano, 1992). Both of these aspects of the model do, however, require further work to provide a more robust explanation of them.

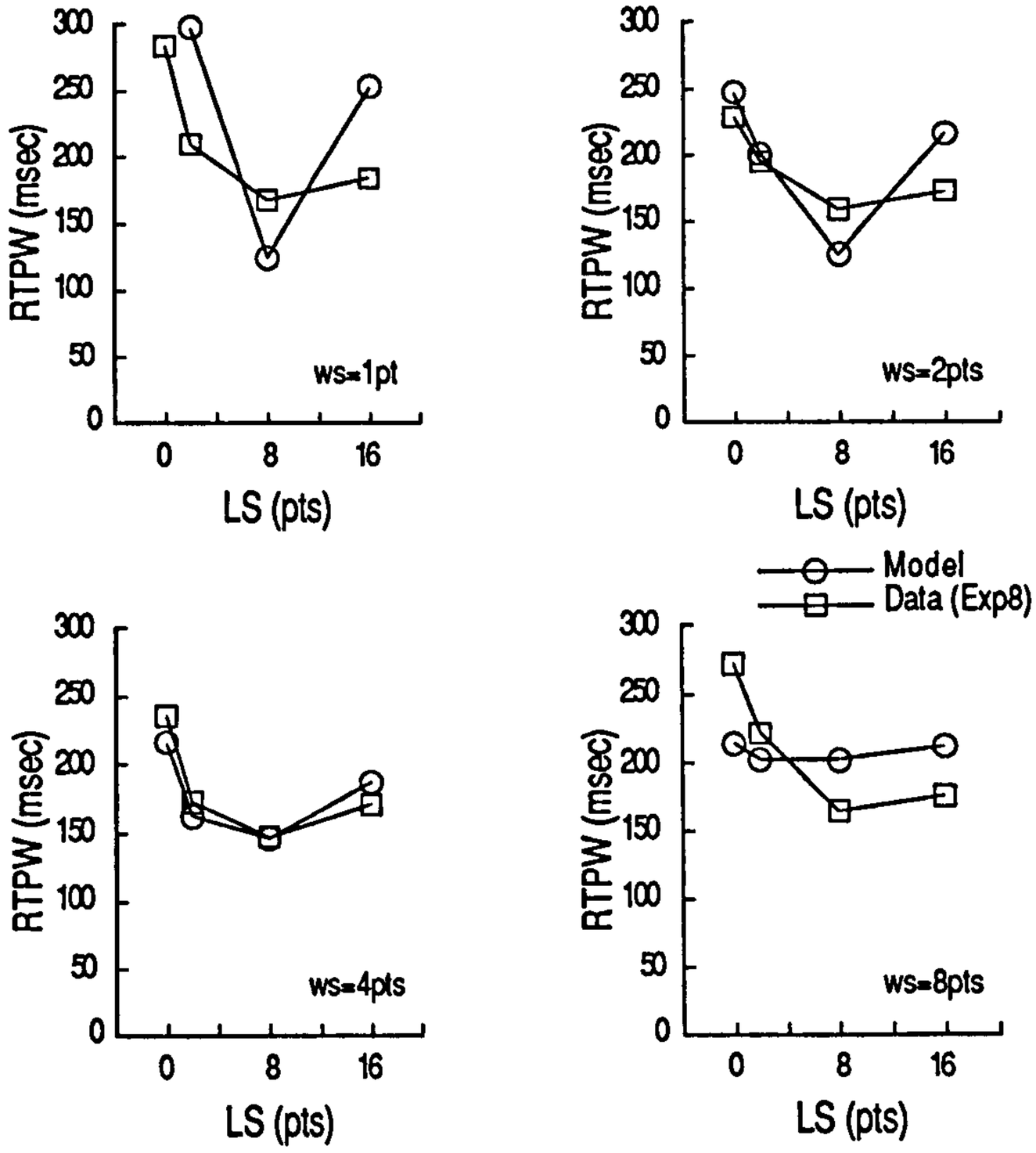


Figure 8.2 Comparison of mean reading time per word (RTPW) in milliseconds (msec) for actual RTPW data of Experiment 8 (squares) and a model of the RTPW based on data from Experiment 7 which makes explicit the word segmentation component (circles). Abscissa (LS) is line spacing in pts. Each graph shows the actual and modelled RTPW for each one of the word spacing (ws) and line spacing combinations.

Although this model is likely to provide a rather over simplistic account of the reading process, it nevertheless provides a useful means of examining the relationship between the visual processing aspects of the processing of text in reading to reading performance. Indeed, the comparability of the reading time per word (RTPW) of Experiment 8 data and the model of the RTPW has shown how changes in reading performance might be explained by a model of the visual processing of text.

9

Summary and Conclusions

9.1 Summary

In introducing the topic of this thesis, consideration was first given to what a further study could contribute to this area. Examination of the research into the role of visual processing in reading led to the conclusion that, in fact, very little is understood about the nature of the visual processing of text in reading. It was argued that this lack of understanding was due, in large part, to two problems. The first problem was the inability of methodologies to separate the visual aspects of text processing in reading from the other, non-visual, aspects, or to identify and examine the effects of the visual processing of individual text processing tasks. The second problem was the inadequate account in reading research of the operations performed by early visual processing itself. It was suggested that these problems had led, inevitably, to a number of assumptions being made, in both research and theory, about what the role of vision in reading must consequently be limited to.

An outline of what the operations performed by early vision might be was then given, and the case made for a computational approach to visual processing. This discussion led to the conclusion that a profitable way forward in gaining an understanding of the visual processing of text in reading might be made by applying a computational model of vision to text with the aim of providing an account of the visual processing of text by modelling this process.

The experimental and computational work of the thesis began with the initial step of applying an implementation of the (Watt, 1988) MIRAGE model of early vision to pages of text. The model was adopted because of its ability to account for a wide range of psychophysical findings, and its compatibility with neurophysiological data.

Analysis of the result of this process showed how the modelled representation of text images extracted a structured set of features, or "regions", of the information contained in a page of text which was organised across a range of spatial scales spanning those found in

human vision. It was suggested that the regions extracted were at least capable of supporting a range of visual tasks performed in the processing of text in reading. The manner in which these features were organised across spatial scales led to the grouping, or more precisely, the 'un-grouping' of regions of information in the image. Thus, 'line' regions represented at very coarse spatial scales were un-grouped into 'word' regions at a less coarse spatial scale. At progressively finer spatial scales, these word regions were un-grouped into letter regions, and finally, at the finest spatial scale, letter regions were un-grouped into letter stroke regions. An important feature of the model is that grouping is defined by the properties of the visual system itself. It is worth noting that if the visual processing of text is indeed based on a coarse-to-fine spatial scale of visual analysis, as the model suggests, this would be the antithesis of the way in which the visual processing of text, or even single words, is assumed to proceed by currently influential models of word recognition in reading (*e.g.*, McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; McClelland, 1986). These models either necessitate, or assume, a visual representation at a letter feature-level *before* letter-level and word-level information are represented by "higher-level" processing.

From this initial stage of describing the information available to process text by a model of early visual processing, a number of hypotheses were made from the predictions of the model.

Hypothesis 1, stated that the visual processing of text should be spatial scale dependent. Furthermore, if the human visual processing of text is to be described by the operations of the model, the sensitivity of visual text processing behaviour to changes in physical parameters of text should be explicable in terms of the sensitivity of the model to the same changes. This was *Hypothesis* 2.

Hypothesis 3 was based on the predictions of a time-course of early visual processing from a coarse-to-fine spatial scale of visual analysis by the model. This stated that the visual processing of text should be consistent with a time-course of visual processing which is based on a coarse-to-fine spatial scale of image analysis.

Chapter 3 examined the first question needing to be asked: Whether the visual system has access to, and uses, information which is spatial scale dependent to perform different text processing tasks of the sort performed in reading. A series of psychophysical experiments (Experiments 1–3) revealed that the information in a text image used to perform a number of such tasks was contained in different spatial scales. Experiment 1 showed that word segmentation performance was found to be dependent, mainly, upon information contained at coarse spatial scale, while Experiment 2 found that letter position identification

relied on information contained at finer spatial scales. Experiment 3 found that sentence boundary location was reliant on information contained at a very coarse spatial scale. The data of Experiment 1 was compared to the results of a second computational analysis in which a method was devised to model the psychophysical discriminability of information in the text images used in Experiment 1. Computational Analysis 2 found that the model was able to provide an initial account of the basis of the spatial scale dependency of the visual processing of text, providing support for Hypothesis 1.

Hypothesis 2 was addressed in Chapter 4. Experiment 4 showed how word segmentation performance varied as word spacing varied. Computational Analysis 3 showed how this pattern of word segmentation performance as a function of word spacing could be described by the information represented by the model as a function of word spacing, as predicted. From the comparison of the findings of Experiment 1 and Computational Analysis 2, the prediction of the model was that word segmentation performance should be based on extracting a set of negative regions at coarse spatial scales describing whole word lengths. This is what was found. This provided support for Hypothesis 2: that the sensitivity of visual processing of text to changes in the characteristics of the text should be described by predictable changes in the pattern of 'information' represented by the model.

Computational Analysis 4 showed how the availability of features, or regions, from which text processing performance had been modelled, was dependent on the relationship between 'page-level' (word spacing and line spacing) typographical parameters. This generated a further prediction of the model: that word segmentation performance should be dependent on particular word spacing and line spacing relationships, based on the ability of the model to extract regions in text images required to perform this visual processing task. The predictions of the model were borne out in Experiment 5. Thus, providing further support for Hypothesis 2.

The pattern of the visual processing of text found in Experiment 5 led to a consideration of the time-course of the visual processing of text. A further, profitable, line of enquiry towards an understanding of the visual processing of text was considered to be one in which the time required to perform a text processing task was established and compared to the predictions of the model in terms of its time-course of visual processing.

In Chapter 7, Experiment 6 found that different text processing tasks (word segmentation and letter position identification) relying on information contained at different spatial scales (from the findings of Experiments 1 and 2) required different exposure durations to reach a required level of performance. Experiment 7 found that the time required to perform a

word segmentation task varied as word and line spacing varied. Importantly, this was in a direction generally consistent with the modelled time required for the visual system to reach the spatial scale of analysis which would reveal the information required to perform the task, based on a coarse-to-fine spatial scale of analysis. Experiments 6 and 7 thus provided support for Hypothesis 3.

Finally, Chapter 8, which reported the findings of a reading experiment (Experiment 8), made an initial attempt to explain how the pattern of visual processing of text might influence the pattern of reading performance. The interest in Experiment 8 was not so much in the finding that reading time varied as a function of word and line spacing, but was in the discovery of how reading time was related to the time required to perform one visual text processing task: word segmentation. By modelling the contribution the visual processing task of word segmentation made to the reading time, it was suggested that the model of the visual processing of text adopted in this thesis might be able to provide an adequate account of the visual processing of text in reading.

The findings of this thesis can be seen to have a number of implications for both models of reading and models of vision. These are now outlined and discussed.

9.2 Conclusions

The issues addressed in this thesis span several disciplines. It is most appropriate to discuss the conclusions which can be drawn from the research, and the implications which arise from it, in each of these areas.

9.2.1 Reading

The findings of this thesis suggest that early vision can be seen to be capable of supporting a range of tasks of the sort required in reading which, as was outlined in Chapter 1, are often ascribed to other, non-visual, processes. Indeed, the findings suggest a very different account of the role of early vision in reading to that proposed by any current model of reading. In the interactive-activation model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; McClelland, 1986) vision operates at only a single (fine) spatial scale. Even making the assumption that the other levels in the model might correspond in some, unspecified, way to an analysis at different spatial scales, it is difficult to see how the model would not still require operation from fine-to-coarse spatial scales. In the limiting case of single words presented in isolation, the work of this thesis (Experiments 1–3 and Computational Analysis 2) suggests that early visual processing would provide a representation of information first at a coarse scale corresponding to overall shape of the word, then at progressively finer scales in which some information about the internal letters

of the word, particularly boundary letters, then finally about each letter and its features (strokes etc.) is represented. Both this and previous work (Watt, 1987) indicates that the visual system operates on a coarse-to-fine scale of visual analysis.

The findings also suggest that proposals based on the need to appeal to an attentional mechanism as a necessary processing stage in reading might well be explained within an early visual processing account. The computational and psychophysical data obtained here supports a view of the visual processing of text in which the information required to process text is defined and represented (extracted) by the state of the visual system at a given time, rather than the direction of attention to any of a number of levels which contain that information (*e.g.*, McConkie & Zola, 1987; Bock, Monk & Hulme, 1993).

Finally, the work of this thesis makes it clear that there is a need to develop a new account of the visual processing of text in reading. As was outlined initially in Chapter 1, theories and models of reading have been limited and constrained by the lack of data concerning the nature of the early visual representation of the information used to perform and support text processing in reading. The findings of this thesis provides a new account of the early visual processing of text, which needs to be incorporated into a new framework of its role in the reading process. A proposal for a new account of the early visual processing of text in reading is outlined in Section 9.3.

9.2.2 Vision

An important aspect of the work of this thesis is the conclusions which can be drawn about human vision from its findings, and the implications this might have for a better understanding of visual processing.

Much of vision research has been conducted using stimuli which may have little to do with natural supra-threshold visual processing. This is in part due to the fact that while there are many benefits to be gained from using a psychophysical approach to studying vision, it is hard to use psychophysical procedures to study real visual tasks. One of the reasons for conducting this work was to address this problem. Text is an ideal stimulus with which to study vision, first, because it has been designed around how vision operates, and second, the visual processing of text is a 'real' visual task.

The most important point to make is that the findings of the experimental and computational work have demonstrated that the model of early vision adopted in the thesis, the MIRAGE model (Watt, 1988) has been able to predict and describe some aspects of both the spatial and temporal (dynamic) the human visual processing, and the relationship between the two. The work thus provides further support for this model of early visual processing.

The findings of the thesis do, however, raise an intriguing aspect of both the model of vision, and of human visual processing. It concerns the use of Laplacian of Gaussian filters as an image processing operation representing an initial step in MIRAGE. Using a computational model of early vision in which the image is convolved with Laplacian of Gaussian filters might be considered a limitation of this work, for two important reasons. The first is that it is well established (*e.g.*, Blakemore & Campbell, 1969; Movshon & Blakemore, 1973; Phillips & Wilson, 1984; Georgeson & Harris, 1984) that early human visual mechanisms exist which are orientation selective. The second is that text is composed almost entirely of vertical and horizontal features, from which might be expected an optimal response from vertically and horizontally selective mechanisms. Given this, the finding that this implementation of the model was able to account for the experimental findings without recourse to orientation selective mechanisms, is interesting.

This finding becomes more intriguing by preliminary investigations by the author using an oriented filter implementation of MIRAGE, which indicated that the modelled representation of text was certainly no better, and appeared on initial inspection at least, to be worse in some instances than that delivered by a representation using Laplacian of Gaussian filters. On this basis, it would seem unlikely that the visual system "throws away" the representation of information delivered by Laplacian of Gaussian mechanisms. It is encouraging then, to discover that although neurophysiological evidence demonstrates the existence of oriented receptive fields in the primate striate cortex and later stages of cortical processing (*e.g.*, DeValois *et al*, 1982), there is some evidence that non-oriented, circular-symmetric receptive fields are still represented in these cortical areas (Spillman & Werner, 1990).

However, there is a further aspect of visual processing which requires consideration in the context of the foregoing argument. What Hubel & Wiesel (1968) originally termed "hypercomplex cells" in the striate cortex are more likely to be so-called 'end-stopped' or 'end-inhibited' orientation selective mechanisms (*e.g.*, Maffei, 1985), or more plausibly, given the organisation of the striate cortex into "hypercolumns" (*e.g.*, Hubel *et al.*, 1977), mechanisms which receive cross-orientation inhibition. It would be of interest to discover whether such mechanisms could provide a better, more detailed, representation of information from a page of text than that provided from convolutions using Laplacian of Gaussian filters. This approach may be a fruitful way forward for subsequent research into the visual processing of text.

9.2.3 Typography

The introduction to this thesis argued that conventional typographical practice is not determined by an understanding of the visual basis of reading. Nevertheless, the work of this thesis suggests that such a basis exists. In this respect, typographical principles might be seen in much the same light as Gestalt 'principles' of grouping, and importantly, ones which are based on both spatial, and dynamic aspects of early visual processing. Indeed, the findings of this thesis showing how spatial (typographical) arrangements determined the text elements which are 'grouped' together by the model of visual processing suggest that both typographical practice and Gestalt principles may share the same visual basis. That is, the information used to process text is determined by the visual consequences of the spatial arrangement of text. Any importance of the present work for typography will come from the ability of the findings to further be able to define and predict what the basis of these consequences will be for any text. Such an explanation, might enable the specification of optimum typographical arrangements in the design of documents for particular reading tasks.

9.2.4 Human-Computer Interaction

The type of specification just outlined above would be particularly applicable to the building of computer-based type-design systems and document-design and specification. Indeed, related to the preceding discussion is an issue last mentioned in the introduction to this thesis. It is the issue of the legibility of text presented electronically on a visual display terminal (VDT). It is well established, as outlined briefly in Chapter 1, that reading from a VDT is generally slower and more error prone than reading from paper (*e.g.*, Wright & Lickorish, 1983; Gould *et al.*, 1987). Also pointed out in the introduction was that this phenomenon was not understood, despite much research into this subject (Dillon, 1992). However, the findings of the experiments conducted in this thesis have shown that even very small changes in the typographical arrangements within a page of text can have consequences for the speed and accuracy of visual processing of text.

It is possible therefore that even text which "looks" acceptable (to a user-interface designer, software engineer or user), may have some typographical arrangement which differs slightly from that of an optimally designed page and which may be responsible for any observed difference in reading speed. Seen in this light, it seems unsurprising, in retrospect, that reading from a VDT might often be slower than reading from paper.

Work which examines this topic specifically will serve to establish the extent to which this explanation is able to account for the differences found in reading from VDTs and from paper.

9.3 Towards a computational theory of early visual processing in reading

The findings of the work conducted in this thesis make it possible to finally, and tentatively, propose an outline of a new account of the early visual processing in reading. This outline–and it should be stressed that it is *only* an outline–is based on the findings of this thesis, the view of visual processing adopted here (based on Watt, 1988) and sources already discussed elsewhere in this thesis.

Early vision provides a representation of a text image which is a set of symbolic descriptors specifying the statistics of the set of positive and negative regions (corresponding to 'ON' and 'OFF' mechanisms in visual pathway) of information extracted from the text image at a range of spatial scales (Watt, 1988). This symbolic description, and the early visual image processing operations generating this representation, are used to control and perform the necessary text processing tasks during reading; the number and nature of these text processing tasks being dependent on the reading task requirements. The scale-based time-course of visual processing proceeds asynchronously across the visual field, enabling these various stages of text processing to be performed simultaneously.

On initially coming to a page, all the filters in early vision are active, enabling the reader to determine the orientation and scale of the text, as well as 'landmarks' such as headings, which are represented at a coarse spatial scale. In addition to this, a statistical representation based on information at a range of spatial scales is made available, providing some information about the overall appearance of the page. As the time-course of visual processing progresses, and the largest active filter is progressively switched out, the consequent un-grouping of elements, or information, in the text by the largest active filter results in regions corresponding to whole words being extracted. At this, or an adjacent scale, positive regions corresponding to word spacing may also be extracted, depending on the spatial (typographical) arrangement of the text. This process preserves positional information, relative to the position calculated for the region found at the previous spatial scale, providing the necessary information to program and make a saccade. This level of un-grouping thus serves as a word segmentation step. At this stage, the information provided by early vision about word length and other supraletter features (Monk & Hulme, 1983), along with a partial word definition based on a statistical representation of its features may identify the word, depending on the visual and narrative context. If identification is not possible, the information is used to make a saccade to its centroid or 'centre-of-gravity' in order to fixate the word (Vitu, 1991).

On fixating the word, the finer spatial scale representation now reached as a consequence of the coarse-to-fine scale time course of processing un-groups the negative word-level

regions to provide a finer description of letters and their position in the word that uniquely specifies that word. In circumstances where letters themselves are still not recognised, the extraction of letter features at the finest spatial scale ultimately serves to uniquely specify each letter.

This brief outline of a theory of the early visual processing of text in reading has obvious omissions (and, undoubtedly, errors). The most obvious is the issue of the initial selection of spatial scale. It is quite possible that in reading, as with other visual tasks (*e.g.*, Duncan, 1984), task requirements may determine the most appropriate scale at which to start visual processing operations. How, and on what basis, the visual system determines this has not been given consideration here. This is partially because there is no reason to believe that such a mechanism would alter in any significant way the sequence of every other aspect of the visual processing of text explored here. Notwithstanding this, it is certainly worth examining.

The final point to make about the thesis is that because this thesis represents a new approach to the issue of the visual processing of text, in many respects, the work is essentially a preliminary study. Indeed the work has really only scratched the surface of a potentially very promising and exciting new approach to understanding and exploring the visual processing of text.

It is hoped that both the work of this thesis, and subsequent work emerging from both its findings, and the questions it raises, will serve to illuminate further the nature of the visual processing of text, the reading process itself, the visual basis of typography and last, but not least, the mechanisms of early human visual processing.

References

Blakemore, C. B., & Campbell, F. W. (1969). On he existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology (London), 203,* 237-260.

Bock, J. M., Monk, A. F., & Hulme, C. (1993). Perceptual grouping in visual word recognition. *Memory & Cognition , 21,* 81-88.

Brady, M. (1981). Towards a computational theory of early visual processing in reading. *Visible Language, 15 (2),* 183-214.

Bruce, V., & Green, P. R. (1990). *Visual Perception: Physiology, Psychology and Ecology. 2nd Edition.* Lawrence Erlbaum Associates Ltd. Hove. U.K.

Campbell, F. W., & Robson, J. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London), 197,* 551-566.

Cattell, J. M. (1886). The time taken up by cerebral operations. *Mind, 11,* 220-242.

Coltheart, M. (1987). *Attention and Performance XII, The Psychology of Reading.* 40-61. Lawrence Erlbaum Associates Ltd, Hove. UK.

DeValois, R. L., Yund, E. W., & Helper, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research, 22,* 531-544.

Dillon, A. (1992). Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics, 35,* 1297-1326.

Duncan, J. (1984). Selective attention and the organisation of visual information. *Journal of Experimental Psychology: General, 113,* 501-517.

Duncan, J. (1987). Attention and reading: Wholes and parts in shape recognition - A tutorial review. In: M. Coltheart (Ed.) *Attention and Performance XII, The Psychology of Reading.* 40-61. Lawrence Erlbaum Associates Ltd, Hove. U.K.

Emmott, S. J., & Watt, R. J. (1992). Psychophysical and computational studies of early visual processing in reading. *Perception, 21,* 58.

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London), 187,* 517-552.

Eriksen, C. W., & Yeh, Y. Y. (1975). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 583-597.

Findlay, J. M. (1982). Global visual processing for saccadic eye movements. Vision *Research, 22,* 1033-1045

Fisher, D. F., & Shebilske, W. L. (1985). There is more than meets the eye in the eye-mind assumption. In: R. Groner, G. McConkie, & C. Menz (Eds.) *Eye Movements and Human Information Processing.* Amsterdam: North-Holland.

Georgeson, M. A., & Harris, M. G. (1984). Spatial selectivity of contrast adaptation: Models and data. *Vision Research, 24,* 729-741.

Gilbert, L. C. (1959). Saccadic movements as a factor in visual perception in reading. *Journal of Educational Psychology, 55,* 15-19.

Gould, J. D., Alfaro, L., Barnes, V., Finn, R., Grischkowsky, N., & Minuto, A. (1987). Reading is slower from CRT displays than from paper: Attempts to isolate a single-variable explanation. *Human Factors, 29,* 269-299.

Haber, R. N., & Haber, L. R. (1981). The shape of a word can specify its meaning. *Reading Research Quarterly, XVI,* 334-345.

Henning, G. B., Hertz, B. G., & Broadbent, D. E. (1975). Some experiments bearing on the hypothesis that the visual system analyses patterns in independent bands of spatial frequency. *Vision Research, 15,* 887-899.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195,* 215-243.

Hubel, D. H., Wiesel, T. N., & Stryker, M. P. (1977). Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. *Nature, 269,* 328-330.

Jamar, J. H. T., & Koenderinck, J. J. (1985). Contrast detection and detection of contrast modulation for noise gratings. *Vision Research, 25,* 511-521.

Johnson, J. C., & McClelland, J. L. (1973). Visual factors in word perception. *Perception and Psychophysics, 14,* 365-370.

Johnson, J. C., & McClelland, J. L. (1980). Experimental tests of a hierarchical model of word identification. *Journal of Verbal Learning and Verbal Behaviour, 19,* 503-524.

Julez, B. (1980). Spatial frequency channels in one-, two- and three-dimensional vision: variations on an auditory theme by Bekesy. In C.S. Harris (Ed.), *Visual Coding and Adaptability.* Hillsdale, N.J. LEA Associates Inc.

Koffka, K. (1935). *Principles of Gestalt Psychology.* New York: Harcourt Brace.

Köhler, W. (1947). *Gestalt Psychology: An introduction to new concepts in modern psychology.* New York: Liveright Publishing Corporation.

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading - I. Normal vision. *Vision Research, 25,* 239-252.

Maffei, L. (1985). Complex cells control simple cells. In: D. Rose & V. G. Dobson (Eds.) Models of the Visual Cortex, New York: Wiley.

Marr, D. (1982) *Vision.* San Francisco. Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. B207* , 187-217.

McClelland, J. L. (1986). The programmable blackboard model of reading. In: J. L. McClelland & D. E. Rumelhart, *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 2: Psychological and Biological Models.* Bradford Books, MIT Press, Cambridge, Massachusetts.

McClelland, J. L. (1987). The case for interactionism in language processing. In: Coltheart, M. (Ed.) *Attention and Performance XII, The Psychology of Reading.* 40-61. Lawrence Erlbaum Associates Ltd, Hove. U.K.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375-407.

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

McConkie, G.W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics, 25,* 221-224.

McConkie, G.W., & Zola, D. (1987). Visual attention during eye fixations while reading. In: M. Coltheart (Ed.) *Attention and Performance XII, The Psychology of Reading.* 385-401. Lawrence Erlbaum Associates Ltd, Hove. UK.

Monk, A. F. (1985). Theoretical note: Co-ordinate systems in visual word recognition. *The Quarterly Journal of Experimental Psychology, 37A,* 613-625.

Monk, A. F., & Hulme, C. (1983). Errors in proofreading: Evidence for the use of word shape in word recognition. *Memory & Cognition, 11,* 16-23.

Movshon, J. A., & Blakemore, C. (1973). Orientation specificity and spatial selectivity in human vision. *Perception, 2,* 53-60.

Nachmias, J., & Rogovitz, B. E. (1983). Masking by spatially modulated gratings. *Vision Research, 23,* 1621-1629.

Nothdurft, H. C. (1991). Different effects from spatial frequency masking in texture segregation and texton detection tasks. *Vision Research, 31,* 299-320.

Nyman, G. (1990). Sequential vision and reading. In: C. von Euler, I Lundberg, and G. Lennerstrand (Eds.), *Brain and Reading.* Macmillan.

Paap, K. R., Newsome, S. L., & Noel, R. W. (1984). Word shape's in poor shape for the race to the lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 413-428.

Pavel, M., Sperling, G., Riedl, T., & Vanderbeek, A. (1987). Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American Sign Language. *Journal of the Optical Society of America, (A), 4,* 2355-2365

Perry, V. H., & Silveira, L. C. L. (1988). Functional lamination in the ganglion cell layer of the macaque's retina. *Neuroscience, 25,* 217-224.

Philips, G. C., & Wilson, H. R. (1984). Orientation bandwidths of spatial mechanisms measured by masking. *Journal of the Optical Society of America A, 1,* 226-232.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition, 18,* 1-63.

Rayner, K., & Betera, J. H. (1979). Reading without a fovea. *Science, 206,* 468-469.

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

Rayner, K., McConkie, G. W., & Erhlich, S. F. (1978). Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 529-544.

Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading.* Prentice Hall International, USA.

Rayner, K., & Pollatsek, A. (1987). Eye movements in reading: A tutorial review. In: M. Coltheart (Ed.) *Attention and Performance XII, The Psychology of Reading.* Lawrence Erlbaum Associates Ltd, Hove. UK.

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

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception II: The contextual enhancement effects and some tests and extensions to the model. *Psychological Review, 89,* 60-94.

Rubin, G. S., & Turano, K. (1992). Reading without saccadic eye movements. *Vision Research, 32,* 895-902.

Smith, F. (1969). Familiarity of configuration vs. discriminability of features in the visual identification of words. *Psychnomic Science, 14,* 261-263.

Spencer, H. (1968). The Visible Word. Macmillan Press, UK.

Spillman, L., & Werner, J. S. (1990). *Visual Perception: The neurophysiological foundations.* Academic Press.

Southall, R. (1988). Visual structure and the transmission of meaning. In J. C. van Vliet (Ed.), *Document Manipulation and Typography,* Cambridge University Press, pp. 35-45.

Stromeyer, C. F., & Julez, B. (1972). Spatial frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America, 62,* 1221-1232.

Tinker, M. A. (1955). *Tinker speed of reading test.* University of Minnesota Press, Minneapolis, USA.

Tinker, M. A. (1965). *Bases for Effective Reading.* Minnesota Press. Minneapolis USA.

Triesman, A. M., & Souther, J. (1986). Illusory words: The roles of attention and of top-down constraints in conjoining letters to form words. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 3-17.

Underwood, G., Bloomfield, R. & Clews, S. (1988). Information influences the pattern of eye fixations during sentence comprehension. *Perception, 17,* 267-278.

Vitu, F. (1991). The existence of a centre of gravity effect during reading. *Vision Research, 31,* 1289-1313.

Walker, P. (1987). Word shape as a cue to the identity of a word: An analysis of the Kucera and Francis (1967) word list. *The Quarterly Journal of Experimental Psychology, 39A,* 675-700.

Ward, J. N., & Juola, J. F. (1982). Reading with and without eye movements: Reply to Just, Carpenter and Wolley. *Journal of Experimental Psychology (General), 111,* 239-241.

Wässle, H. (1988). Sampling of visual space by retinal ganglion cells. *Investigative Ophthalmology & Visual Science, Supplement, 29,* 117.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after stimulus offset. *Journal of Optical Society of America, A4,* 2006-2021.

Watt, R. J. (1988). *Visual Processing: Computational, Psychophysical and Cognitive Research.* Lawrence Earlbaum Associates Ltd. Hove, UK.

Watt, R. J. (1991). *Understanding Vision.* Academic Press, London.

Watt, R. J. (1993). The visual analysis of pages of text. In R. Sassoon (Ed.), *Computers and Typography* pp 178-201. Intellect, Oxford.

Watt, R. J., & Andrews, D. P. (1981). APE: adaptive probit estimation of psychometric functions. *Current Psychological Research, 1,* 205-214.

Watt, R. J., Bock, J., Thimbleby, H., & Wilkins, A. (1990). Visible Aspects of Text. In: *Applying Visual Psychophysics to User Interface Design.* BT conference proceedings, Lavenham, 1990.

Watt, R. J., & Morgan, M. J. (1983). The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Research, 23,* 1457-1477.

Watt, R. J., & Morgan, M. J. (1984). Spatial filters and the localisation of luminance changes in human vision. *Vision Research, 24,* 1387-1397

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25,* 239-252.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II. *Psychologische Forschung, 4,* 301-350. Translated as: Laws of organisation in perceptual forms. In: W. D. Ellis (1955). *A source book of Gestalt Psychology.* London: Routledge & Kegan Paul.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1,* 59-85.

Wilson, H. R. (1983). Psychophysical evidence for spatial channels. In: O. J. Braddick & A. C. Sleigh (Eds.), *Physical and biological processing of images.* Berlin: Springer-Verlag.

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

Wright, P., & Lickorish, A. (1983) Proof-reading texts on screen and paper. *Behaviour and Information Technology, 2*, 227-235.