

# **Reading from paper versus screens: a critical review of the empirical literature.**

Andrew Dillon  
HUSAT Research Institute  
University of Technology  
Loughborough  
Leics  
UK

---

This item is not the definitive copy. Please use the following citation when referencing this material: Dillon, A. (1992) Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics*, 35(10), 1297-1326.

---

## **Abstract**

The advent of widespread computer use in general and increasing developments in the domain of hypertext in particular have increased awareness of the issue of reading electronic text. To date the literature has been dominated by reference to work on overcoming speed deficits resulting from poor image quality but an emerging literature reveals a more complex set of variables at work. The present review considers the differences between the media in terms of outcomes and processes of reading and concludes that single variable explanations are insufficient to capture the range of issues involved in reading from screens.

Published in *Ergonomics*, 35,10 1297-1326  
Adillon@indiana.edu

## **1 Introduction**

In simple terms, there exist two schools of thought on the subject of electronic texts. The first holds that paper is far superior and will never be replaced by screens. The argument is frequently supported by reference either to the type of reading scenarios that would currently prove difficult if not impossible to support acceptably with electronic text, e.g., reading a newspaper on the beach or a magazine in bed, or the unique tactile qualities of paper. The latter aspect is summed up neatly in Garland's (1982) comment that electronic text may have potential uses:

“but a book is a book is a book. A reassuring, feel-the-weight, take-your-own-time kind of thing...” (cited in Whaller 1987, p. 261).

The second school favours the use of electronic text, citing ease of storage and retrieval, flexibility of structure and saving of natural resources as major incentives. According to this perspective, electronic text will soon replace paper and in a short time (usually ten years hence) we shall all be reading from screens as a matter of habit. In the words of its greatest proponent, Ted Nelson (1987):

“the question is not can we do everything on screens, but when will we, how will we and how can we make it great? This is an article of faith - its simple obviousness defies argument.”

Such extremist positions show no signs of abating though it is becoming clear to many researchers in the domain that neither is particularly satisfactory. Reading from screens is different from paper and there are many scenarios such as those cited that current technology would not support well, if at all. However, technology is developing and electronic text of the future is unlikely to be handicapped by limitations in screen image and portability that currently seem major obstacles. As Licklider pointed out when considering the application of computers in libraries as early as 1965:

"our thinking and our planning need not be, and indeed should not be, limited by literal interpretation of the existing technology" (p.19).

Even so, paper is an information carrier par excellence and possesses an intimacy of interaction that can never be obtained in a medium that by definition imposes a microchip interface between the reader and the text. Furthermore, the millions of books that exist now will not all find their way into electronic form, thus ensuring the existence of paper documentation for many years yet.

The aim of the present review is not to resolve the issue of whether one or other medium will dominate but to examine critically the reported differences between them in terms of use and thereby support reasoned analysis of the paper versus electronic text debate from the perspective of the reader. In so doing it should highlight the crucial issues underlying the usability of a medium.

## 2 The outline of the review

The review will describe the reported differences between the media before examining the attempts at explaining and overcoming them. At the outset it must be stated that drawing any firm conclusions from the literature is difficult. Helander et al (1984) evaluated 82 studies concerning human factors research on VDUs and concluded:

"Lack of scientific rigour has reduced the value of many of these studies. Especially frequent were flaws in experimental design and subject selection, both of which threaten the validity of results. In addition, the choice of experimental settings and dependent and

independent variables often made it difficult to generalize the results beyond the conditions of the particular study." (p. 55.)

Waern and Rollenhagen (1983) point to the frequently narrow scope of experimental designs in such studies. Important factors are either not properly controlled or are simply not reported and most studies use unique procedures and equipment, rendering direct comparison meaningless. The present review is not intended to untangle the methodological knots of other researchers but rather to make sense of the major findings in a general way and indicate where the research needs lie.

A detailed literature already exists on typographical issues related to text presentation on paper (see particularly the work of Tinker) and issues such as line spacing and formatting are well researched. This work will not be reviewed here as much of it remains unreplicated on VDUs and evidence suggests that, even when such factors are held constant, reading differences between the two presentation media remain (see for example Creed et al, 1987).

In the first instance this review examines the nature of the possible differences between the media and draws a distinction here between outcome (section 4) and process (section 5) differences. Following this, a brief overview of the type of research that has been carried out is presented (section 6). This describes the range of issues that have been covered and presents the intended scope of the subsequent review. Experimental comparisons of reading from paper and screen are then reviewed; these are grouped according to the variables they manipulated (sections 7 and 8). A final section highlights the shortcomings of much of this work and indicates the way forward for research in this domain.

### 3 Observed differences: outcome versus process measures

Analysing reading is not a simple task and a distinction has been drawn between assessing reading behaviour in terms of outcome and process measures (Schumacher and Waller 1985). Outcome measures concentrate on what the reader gets from the text and considers such variables as amount of information retrieved, accuracy of recall, time taken to read the text and so forth. Process measures are more concerned with how the reader uses a text and include such variables as where the reader looks in the text and how s/he manipulates it.

In the domain of electronic text outcome measures take on a particular relevance as advocates proclaim increased efficiency and improved performance (i.e., outcomes) with computer presented material (aspects of direct concern to ergonomists). It is not surprising therefore to find that the majority of work comparing the two media has concentrated heavily on such differences. With the emergence of hypertext however, navigation has become a major issue and process measures are gaining increased recognition of importance.

In the following sections a summary of the observed differences between the media in terms of outcomes and processes is presented.

## 4 Outcome Measures

### 4.1 Speed

By far the most common experimental finding is that silent reading from screen is significantly slower than reading from paper (Kak, 1981; Muter et al, 1982; Wright and Lickorish, 1983; Gould and Grischkowsky, 1984; Smedshammar et al 1989). Figures vary according to means of calculation and experimental design but the evidence suggests a performance deficit of between 20% and 30% when reading from screen.

However, despite the apparent similarity of findings, it is not clear whether the same mechanisms have been responsible for the slower speed in these experiments, given the great disparity in procedures. For example, in the study by Muter et al (1982), subjects read white text on a blue background, with the subject being approximately 5 m from the screen. The characters, displayed in teletext format on a television, were approximately 1 cm high, and time to fill the screen was approximately 9 seconds. Even ignoring the unnatural character size and distance from the screen, the authors reported that the experimental room was “well illuminated by an overhead light source”, a factor which by virtue of the possible reflections caused could account for a slow reading speed. Additionally, unless the book used was one of the large format books prepared for the partially sighted, we must assume that the screen text characters were substantially larger than the printed characters.

In comparison, Gould and Grischkowsky (1984) used greenish text on a dark background. Characters were 3 mm high and subjects could sit at any distance from the screen. They were encouraged to adjust the room lighting level and the luminance and contrast of the screen for their comfort. Printed text used 4 mm characters and was laid out identically to the screen text. Wright and Lickorish (1983) give no details of text size other than that it was displayed as white characters on a black 12" screen driven by an Apple II microcomputer with lower case facility. This would suggest that it was closer to Gould's text than Muter's text in appearance. Printed texts were photocopies of printouts of the screen displays produced on an Epson MX-80 dot matrix printer, compared with Gould's 10-point monospace Letter Gothic font.

In contrast to these studies, Switchenko (1984), Askwall (1985) and Cushman (1986) found that reading speed was unaffected by the presentation medium. Askwall attributes this difference in findings to the fact that her texts were comparatively short (22 sentences), and the general lack of experimental detail makes alternative interpretations difficult. Although it is reported that a screen size of 24 rows by 40 columns was used, with letter size approximately 0.5 x 0.5 cm and viewing distance of approximately 30-50 cm, no details of screen colour or image polarity and none of the physical attributes of the printed text are given.

Cushman's primary interest was in fatigue but he also measured reading speed and comprehension using 80-minute reading sessions. Negative and positive image VDU and microfiche presentations were used and most of the 76 subjects are described as having had "some previous experience using microfilm readers and VDUs." On the basis of this study Cushman concluded that there was no evidence of a performance deficit for the VDU presentations compared with printed paper.

As this indicates, the evidence surrounding the argument for a speed deficit in reading from VDUs is less than conclusive. A number of intervening variables, such as the size, type and quality of the VDU may have contaminated the results. As will be consistently demonstrated, this criticism applies repeatedly to most of the evidence on reading from VDUs. However, despite the methodological weaknesses of many of the investigations, evidence continues to mount supporting the case for a general speed decrement. As Gould et al (1987a) noted, many of these experiments are open to interpretation but :

"the evidence on balance...indicates that the basic finding is robust-- people do read more slowly from CRT displays" (p. 269)

#### 4.2 Accuracy

Accuracy of reading could refer to any number of everyday activities such as locating information in a text, recalling the content of certain sections and so forth. In experimental investigations of reading from screens the term accuracy has several meanings too though it most commonly refers to an individual's ability to identify errors in a proofreading exercise. While a number of studies have been carried out which failed to report accuracy differences between VDUs and paper (e.g., Wright and Lickorish, 1983; Gould and Grischkowsky, 1984) recent well controlled experiments by Creed et al (1987) and Wilkinson and Robinshaw (1987) report significantly poorer accuracy for such proofreading tasks on screens.

Since evidence for the effects of presentation media on such accuracy measures often emerges from the same investigations which looked at the speed question, the criticisms of procedure and methodology outlined above apply equally here. The measures of accuracy employed also vary. Gould and Grischkowsky (1984) required subjects to identify misspellings of four types: letter omissions, substitutions, transpositions and additions, randomly inserted at a rate of one per 150 words. Wilkinson and Robinshaw (1987) argue that such a task hardly equates to true proofreading but is merely identification of spelling mistakes. In their study they tried to avoid spelling or contextual mistakes and used errors of five types : missing or additional spaces, missing or additional letters, double or triple reversions, misfits or inappropriate characters, and missing or inappropriate capitals. It is not always clear why some of these error types are not spelling or contextual mistakes but Wilkinson and Robinshaw suggest their approach is more relevant to the task demands of proofreading than Gould and Grischkowsky's.

However Creed et al (1987) distinguished between visually similar errors (e.g., "e" replaced by "c"), visually dissimilar errors (e.g., "e" replaced by "w") and syntactic errors (e.g., "gave" replaced by "given"). They argue that visually similar and dissimilar errors require visual discrimination for identification while syntactic errors rely on knowledge of the grammatical correctness of the passage for detection and are therefore more cognitively demanding. This error classification was developed in response to what they saw as the shortcomings of the more typical accuracy measures which provide only gross information concerning the factors affecting accurate performance. Their findings indicate that visually dissimilar errors are significantly easier to locate than either visually similar or syntactic errors.

In a widely reported study Egan et al (1989) compared students' performance on a set of tasks involving a statistics text presented on paper or screen. Students used either the standard textbook or a hypertext version run on SuperBook, a structured browsing system, to search for specific information in the text and write essays with the text open. Incidental learning and subjective ratings were also assessed. The search tasks provide an alternative to, and more realistic measure of reading accuracy than identifying spelling errors.

The authors report that subjects using the hypertext performed significantly more accurately than those using the paper text. However a closer look at the experiment is revealing. With respect to the search tasks, the questions posed were varied so that their wording mentioned terms contained in the body of the text, in the headings, in both of these or neither. Not surprisingly the largest advantage to electronic text was observed where the target information was only mentioned in the body of text (i.e. there were no headings referring to it). Here it is hardly surprising that the search facility of the computer outperformed humans. When the task was less biased against the paper condition e.g., searching for information to which there are headings, no significant difference was observed. Interestingly the poorest performance of all was for SuperBook users searching for information when the question did not contain specific references to words used anywhere in the text. In the absence of suitable search parameters or look-up terms hypertext suddenly seemed less usable.

McKnight et al (1990) compared reading in two versions of hypertext, a word processor file and a paper copy of a document on winemaking. The measure of accuracy taken was the number of answers correctly made to a set of questions seeking information to be found in the document. Interestingly they report no significant difference between paper and word processor file, but readers in both hypertext conditions were significantly less accurate than readers of the paper document.

Regardless of the interpretation that is put on the results of any one of these studies, the fact remains that investigations of reading accuracy from VDU and paper take a variety of measures as indices of performance. Therefore two studies, both purporting to investigate reading accuracy may not necessarily measure the same events. In summary it would seem that for routine spelling checks reading from VDUs is not less accurate than reading from paper. However, a performance deficit does seem to occur for more visually

or cognitively demanding tasks. Altering the structure of the document as in hypertext applications introduces another level of complexity to the discussion that requires much further research.

### 4.3 Fatigue

The proliferation of information technology has traditionally brought with it fears of harmful or negative side-effects for users who spend a lot of time in front of a VDU (see for example Pearce, 1984). In the area of screen reading this has manifested itself in speculation of increased visual fatigue and/or eyestrain when reading from screens as opposed to paper.

In the Muter et al (1982) study subjects were requested to complete a rating scale on a number of measures of discomfort including fatigue and eyestrain both before and after exposure to the task. There were no significant differences reported on any of these scales either as a result of condition or time. Similarly Gould and Grischkowsky (1984) obtained responses to a 16-item "Feelings Questionnaire" after each of six 45-minute work periods. This questionnaire required subjects to rate their fatigue, levels of tension, mental stress and so forth. Furthermore various visual measurements such as flicker and contrast sensitivity, visual acuity and phoria, were taken at the beginning of the day and after each work period. Neither questionnaire responses nor visual measures showed a significant effect for presentation medium. These results led the authors to conclude that good-quality VDUs in themselves do not produce fatiguing effects, citing Starr et al (1982) and Sauter et al (1983) as supporting evidence.

In a more specific investigation of fatigue Cushman (1986) investigated reading from microfiche as well as paper and VDUs with positive and negative image. He distinguished between visual and general fatigue, assessing the former with the Visual Fatigue Graphic Rating Scale (VFGRS) which subjects use to rate their ocular discomfort, and the latter with the Feeling-Tone Checklist (FTC, Pearson and Byars, 1956). With respect to the VDU conditions, the VFGRS was administered before the session and after 15, 30, 45 and 60 minutes as well as at the end of the trial at 80 minutes. The FTC was completed before and after the session. The results indicated that reading from positive presentation VDUs (dark characters on light background) was more fatiguing than paper and leads to greater ocular discomfort than reading from negative presentation VDUs.

Cushman explained the apparent conflict of these results with the established literature in terms of the refresh rate of the VDUs employed (60 Hz) which may not have been enough to completely eliminate flicker in the case of positive presentation, a suspected cause of visual fatigue. Wilkinson and Robinshaw (1987) also reported significantly higher fatigue for VDU reading and while their equipment may also have influenced the finding they dismiss this as a reasonable explanation on the grounds that no subject reported lack of clarity or flicker and their monitor was typical of the type of VDU that users find themselves reading from. They suggest that Gould and Grischkowsky's (1984) equipment was "too good to show any disadvantage" and that their method of measuring

fatigue was artificial. By gathering information after a task and across a working day Gould and Grischkowsky missed the effects of fatigue within a task session and allowed time of day effects to contaminate the results. Wilkinson and Robinshaw liken the proofreading task used in these studies to vigilance performance and argued that fatigue is more likely to occur within the single work period where there are no rest pauses allowing recovery. Their results showed a performance decrement across the 50-minute task employed, leading them to conclude that reading from typical VDUs at least for periods longer than 10-minutes is likely to lead to greater fatigue.

It is not clear how comparable conclusions drawn from measures of fatigue such as subjective ratings of ocular discomfort are with inferences drawn from performance rates. It would seem safe to conclude that users do not find reading from VDUs intrinsically fatiguing but that performance levels may be more difficult to sustain over time when reading from average quality screens. As screen standards increase over time this problem should be minimised.

#### 4.4 Comprehension

Perhaps more important than the questions of speed and accuracy of reading is the effect of presentation medium on comprehension. Should any causal relationship ever be identified between reading from VDU and reduced comprehension, the impact of this technology would be severely limited. The issue of comprehension has not been as fully researched as one might expect, perhaps in no small way due to the difficulty of devising a suitable means of quantification i.e., how does one measure a reader's comprehension?

Post-task questions about content of the reading material are perhaps the simplest method of assessment, although care must be taken to ensure that the questions do not simply demand recall skills. Muter et al (1982) required subjects to answer 25 multiple-choice questions after two 1 hour reading sessions. Due to variations in the amount of material read by all subjects, analysis was reduced to responses to the first eight questions of each set. No effect on comprehension was found either for condition or question set. Kak (1981) presented subjects with a standardised reading test (the Nelson-Denny test) on paper and VDU. Comprehension questions were answered by hand. No significant effect for presentation medium was observed. A similar result was found by Cushman (1986) in his comparison of paper, microfiche and VDUs. Interestingly however, he noted a negative correlation between reading speed and comprehension, i.e., comprehension tended to be higher for slower readers.

Belmore (1985) asked subjects to read short passages from screen and paper and measured reading time and comprehension. An initial examination of the results appeared to show a considerable disadvantage, in terms of both comprehension and speed, for screen presented text. However, further analysis showed that the effect was only found when subjects experienced the screen condition first. Belmore suggested that the performance decrement was due to the subjects' lack of familiarity with computers and reading from screens - a factor commonly found in this type of study. Very few of the studies reported here attempted to use a sample of regular computer users.



Gould et al (1987a) compared subjects reading for comprehension with proofreading for both media in order to check that typical proofreading tasks did not intrinsically favour a medium that supported better character discrimination. Though only concerned with reading speed (i.e., they took no comprehension measures) they found that comprehension actually exacerbated the differences between paper and screen.

The Egan et al study (1989) described earlier required subjects to write essay type answers to open book questions using paper or hypertext versions of a statistics book. Experts rated the essays and it was observed that users of the hypertext version scored significantly higher marks than users of the paper book. Thus, the authors conclude, the potential of restructuring the text with current technology can significantly improve comprehension for certain tasks.

The most recently published study covering this issue is by Muter and Maurutto (1991) who asked readers to answer questions about a short story read either on paper or screen immediately after finishing the reading task. They reported no significant comprehension difference between readers using either medium.

It seems therefore that comprehension of material is not negatively affected by presentation medium and under some circumstances may even be improved. However, a strong qualification of this interpretation of the experimental findings is that suitable comprehension measures for reading material are difficult to devise. The expert rating used by Egan et al is ecologically valid in that it conforms to the type of assessment usually employed in schools and colleges but the sensitivity of post-task question and answer sessions to subtle cognitive differences caused by presentation medium is debatable. Without evidence to the contrary though, it would seem as if reading from VDUs does not negatively affect comprehension rates though it may affect the speed with which readers can attain a given level of comprehension.

#### 4.5 Preference

Part of the folklore of human factors research is that naive users tend to dislike using computers and much research aims at encouraging user acceptance of systems through more usable interface design. Given that much of the evidence cited here is based on studies of relatively novice users it is possible that the results are contaminated by subjects' negative predispositions towards reading from screen. On the basis of a study of 800 VDU operators' comparisons of the relative qualities of paper and screen based text, Cakir et al (1980) report that high quality typewritten hardcopy is generally judged to be superior. Preference ratings were also recorded in the Muter et al (1982) study and despite the rather artificial screen reading situation tested, users only expressed a mild preference for reading from a book. They expressed the main advantage of book reading to be the ability to turn back pages and re-read previously read material, mistakenly assuming that the screen condition prevented this.

Starr (1984) concluded that relative subjective evaluations of VDUs and paper are highly dependent on the quality of the paper document, though one may add that the quality of

the VDU display probably has something to do with it too. Egan et al (1989) found a preference for hypertext over paper amongst subjects in their study of a statistics text where the electronic copy was displayed on a very high quality screen. Recent evidence from Muter and Mauretto (1991) revealed that approximately 50% of subjects in their comparative studies of reading from paper and current screens expressed a preference for screen, lending some support to the argument that preferences are shifting as screen technology improves.

What seems to have been overlooked as far as formal investigation is concerned is the natural flexibility of books and paper over VDUs, e.g., paper documents are portable, cheap, apparently "natural" in our culture, personal and easy to use. The extent to which such "common-sense" variables influence user performance and preferences is not yet well-understood.

#### 4.6 Summary

Empirical investigations of the area have suggested five possible outcome differences between reading from screens and paper. As a result of the variety of methodologies, procedures and stimulus materials employed in these studies, definitive conclusions cannot be drawn. It seems certain that reading speeds are reduced on typical VDUs and accuracy may be lessened for cognitively demanding tasks. Fears of increased visual fatigue and reduced levels of comprehension as a result of reading from VDUs would however seem unfounded though the validity of separating accuracy and comprehension into two discrete outcomes is debatable. With respect to reader preference, top quality hardcopy seems to be preferred to screen displays, which is not altogether surprising.

#### 5 Process Measures

Without doubt, the main obstacle to obtaining accurate process data is devising a suitable, non-intrusive observation method. While techniques for measuring eye-movements during reading now exist, it is not at all clear from eye-movement records what the reader was thinking or trying to do at any time. Furthermore, use of such equipment is rarely non-intrusive, often requiring the reader to remain immobile through the use of head restraints, bite bars etc., or read the text one line at a time from a computer display — hardly equatable to normal reading conditions!

Less intrusive methods such as the use of light pens in darkened environments to highlight the portion of the text currently viewed (Whalley and Fleming 1975) or modified reading stands with semisilvered glass which reflect the readers' eye movements in terms of current text position to a video camera (Pugh 1979) are examples of the lengths researchers have gone to in order to record the reading process. However, none of these are ideal as they alter the reading environment, sometimes drastically, and only the staunchest advocate would describe them as non-intrusive.

Verbal protocols of people interacting with texts require no elaborate equipment and can be elicited wherever a subject normally reads. In this way they are cheap, relatively

naturalistic and physically non-intrusive. However, the techniques have been criticised for interfering with the normal processing involved in task performance and requiring the presence of an experimenter to sustain and record the verbal protocol (Nisbett and Wilson 1977).

Although a perfect method does not yet exist it is important to understand the relative merits of those that are available. Eye-movement records have significantly aided theoretical developments in modeling reading (see e.g., Just and Carpenter 1980) while use of the light-pen-type techniques have demonstrated their worth in identifying the effects of various typographic cues on reading behaviour (see e.g., Waller 1984). Verbal protocols have been effectively used by researchers to gain information on reading strategies (see e.g., Olshavsky 1977).

Nevertheless, such techniques have rarely been employed with the intention of assessing the process differences between reading from paper and from screen. Where paper and hypertext are directly compared, although process measures may be taken with the computer and or video cameras, the final comparison often rests on outcome measures (e.g., McKnight et al 1990).

Despite this, it is widely accepted that the reading process with screens is different than that with paper regardless of any outcome differences. The following sections outline three of the most commonly cited process differences between the media. In contrast to the outcome differences it will be noted that, for the reasons outlined above, these differences are less clearly empirically demonstrated.

## 5.1 Eye movements

Mills and Weldon (1986) argue that measures of eye movements reflect difficulty, discriminability and comprehensibility of text and can therefore be used as a method of assessing the cognitive effort involved in reading text from paper or screen. Indeed Tinker (1958) reports on how certain text characteristics affect eye movements and Kolars et al (1981) employed measures of eye movement to investigate the effect of text density on ocular work and reading efficiency. Obviously if reading from screen is different than paper then noticeable effects in eye movement patterns might be found indicating possible causes and means of improvement.

Eye movements during reading are characterised by a series of jumps and fixations. The latter are of approximately 250 msec. duration and it is during these that word perception occurs. The 'visual reading field' is the term used to describe that portion of foveal and parafoveal vision from which visual information can be extracted during a fixation and in the context of reading this can be expressed in terms of the number of characters available during a fixation. The visual reading field is subject to interference from text on adjacent lines, the effect of which seems to be a reduction in the number of characters available in any given fixation and hence a reduction in reading speed.

Gould et al (1987a) report an investigation of eye movement patterns when reading from either medium. Using a photoelectric eye movement monitoring system, subjects were required to read two 10-page articles, one on paper, the other on screen. Eye movements typically consisted of a series of fixations on a line, with re-fixations and skipped lines being rare. Movement patterns were classified into four types: fixations, undershoots, regressions and re-fixations. Analysis revealed that when reading from VDU subjects made significantly more (15%) forward fixations per line. However this 15% difference translated into only 1 fixation per line. Generally, eye movement patterns were similar and no difference in duration was observed. Gould explained the 15% fixation difference in terms of image quality variables. Interestingly he reports that there was no evidence that subjects lost their place, "turned-off" or re-fixated more when reading from VDUs.

It seems therefore that gross differences in eye movements do not occur between screen and paper reading. However, given the known effect of typographic cueing on eye movements with paper and the oft-stated non-transferability of paper design guidelines to screens, it is possible that hypertext formats might influence the reading process at this level in a manner worth investigation.

## 5.2 Manipulation

Perhaps the most obvious difference between reading from paper and from screens is the ease with which paper can be manipulated and the corresponding difficulty of so doing with electronic text. Yet manipulation is an intrinsic part of the reading process for most tasks. Manipulating paper is achieved by manual dexterity, using fingers to turn pages, keeping one finger in a section as a location aid, or flicking through tens of pages while browsing the contents of a document, activities difficult or impossible to support electronically (Kerr 1986).

Such skills are acquired early in a reader's life and the standard physical format of most documents means these skills are transferable between all document types. With electronic text this does not hold. Lack of standards means that there is a bewildering range of interfaces to computer systems and mastery of manipulation in one application is no guarantee of an ability to use another. Progressing through the electronic document might involve using a mouse and scroll bar in one application and function keys in another; one might require menu selection and "page" numbers while another supports touch-sensitive "buttons". With hypertext, manipulation of large electronic texts can be rapid and simple while other systems might take several seconds to refresh the screen after the execution of a "next page" command.

Such differences will almost certainly affect reading. Waller (1986) suggests that as readers need to articulate their needs in manipulating electronic texts (i.e., formulate an input to the computer to move the text rather than directly and automatically performing the action themselves) a distraction of cognitive resources required for comprehension could occur. Richardson et al., (1988) report that subjects find text manipulation on screen awkward compared to paper, stating that the replacement of direct manual interaction with an input device deprived users of much feedback and control.

It is obvious that manipulation differences exist and that electronic text is usually seen as the less manipulable medium. Current hypertext applications however, support rapid movement between various sections of text which suggests that innovative manipulations might emerge that, once familiar with them, convey advantages to the reader of electronic texts. This is an area for further work.

### 5.3 Navigation

When reading a lengthy document the reader will need to find their way through the information in a manner that has been likened to navigating a physical environment (Dillon et al 1990a). There is a striking consensus among many researchers in the field that this process is the single greatest difficulty for readers of electronic text. This is particularly (but not uniquely) the case with hypertext where frequent reference is made to “getting lost in hyperspace” (e.g., Conklin 1987, McAleese 1989) which is described, in the oft-quoted line of Elm and Woods (1985), as:

“the user not having a clear conception of the relationships within the system or knowing his present location in the system relative to the display structure and finding it difficult to decide where to look next within the system” (p.927).

With paper documents there tends to be at least some standards in terms of organisation. With books for example, contents pages are usually at the front, indices at the back and both offer some information on where items are located in the body of the text. Concepts of relative position in the text such as ‘before’ and ‘after’ have tangible physical correlates. No such correlation holds with hypertext and such concepts are greatly diminished in standard electronic text.

There is some direct empirical evidence in the literature to support the view that navigation can be a problem. Edwards and Hardman (1989) for example, describe a study which required subjects to search through a specially designed hypertext. In total, half the subjects reported feeling lost at some stage (this proportion is inferred from the data reported). Such feelings were mainly due to “not knowing where to go next” or “not knowing where they were in relation to the overall structure of the document” rather than “knowing where to go but not knowing how to get there” (descriptors provided by the authors). Unfortunately, without direct comparison of ratings from subjects reading a paper equivalent we cannot be sure such proportions are solely due to using hypertext.

McKnight et al (1990) compared navigation for paper, word processor and two hypertext documents by examining the number of times readers went to index and contents pages/sections, inferring that time spent here gave an indication of navigation problems. They reported significant differences between paper and both hypertext conditions (the latter proving worse), with word processor users spending about twice as long as paper readers in these sections (a statistically non-significant difference however).

Indirect evidence comes from the numerous studies which have indicated that users have difficulties with a hypertext (Monk et al 1988, Gordon et al 1988). Hammond and Allinson (1989) speak for many when they say:

“Experience with using hypertext systems has revealed a number of problems for users..... First, users get lost... Second, users may find it difficult to gain an overview of the material... Third, even if users know specific information is present they may have difficulty finding it” p294.

There are a few dissenting voices. Brown (1988) argues that:

“although getting lost is often claimed to be a great problem, the evidence is largely circumstantial and conflicting. In some smallish applications it is not a major problem at all” (p. 2) .

This quote is telling in several ways. The evidence for navigational difficulties is often circumstantial, as noted above. The applications in which Brown claims it is not a problem at all, are, to use his word, “smallish” and this raises a crucial issue with respect to electronic text research that is taken up later, how much faith can we place in evidence from studies involving very short texts. However, the evidence that we currently possess seems to indicate that navigation is a reading process issue worthy of further investigation.

#### 5.4 Summary

The reading process is affected by the medium of presentation though it is extremely difficult to quantify and demonstrate such differences empirically. The major differences appear to occur in manipulation which seems more awkward with electronic texts and navigation which seems to be more difficult with electronic and particularly hypertexts. Eye movement patterns do not seem to be significantly altered by presentation medium. Further process issues may emerge as our knowledge and conceptualisation of the reading process improves.

#### 6 Explaining the differences: A classification of issues

While the precise nature and extent of the differences between reading from either medium have not been completely defined, attempts to identify possible causes of any difference have frequently been made. A significant literature exists on issues dealing with display characteristics such as line length and spacing. It is not the aim of this review to detail this literature fully except where it relates to possible causes for reading differences between paper and screen. Experimental investigations which have controlled such variables have still found performance deficits on VDUs, thus suggesting that the root cause of observed differences lies elsewhere. For a comprehensive review of these issues see Mills and Weldon (1985).

Examining the last 15 years of Human Factors research in this area it is possible to distinguish three types of investigation. Dillon (1990) for example, has loosely categorised these as levels, depending on their concern with: broad or narrow issues (e.g., cognition or perception); size of text (e.g., one page or multi-page document) and specificity of prediction that can be made from this work (e.g., the nature of the difference between media or the likely existence of a difference).

Initial (or first level) work concentrated on what could be termed basic ergonomics such as screen angle, image polarity and so forth. This work continues to some extent today. Concerned with perceptual or physical rather than mainly cognitive issues, this work has been carried out mainly on proofreading short texts and has produced detailed results on the likely performance deficits for certain screen types. As technology developed and user interfaces afforded more sophisticated interaction with electronic texts, second level issues to do with document manipulation, such as scrolling versus paging, came to the fore. These involved work with larger texts and more cognitively demanding tasks than proofreading. This is still an area of concern for many researchers. The third level in this scheme has resulted from the explosion of hypertext systems and concerns issues such as navigation and information models grouped under the heading information structuring.

In a very real sense all these areas are inter-related. Hypertext, by necessity involves reading from screens and manipulating electronic text and therefore research at the basic ergonomic level has relevance to the information structuring work, if only as a reminder of necessary but insufficient preconditions to effective reading from screens. Given the major concern of this review is with empirical literature, a form mainly lacking in much of the hypertext area, the following sections cover only the issues of basic and visual ergonomics as well as those of document manipulation. The issues concerned with information structuring are sufficiently detailed to warrant a paper of their own which would be different in granularity from the present one by virtue of poor level of empiricism involved. However a paper dealing with those issues and relating them to the present areas is currently in preparation by the present author. Readers concerned primarily with navigation in electronic documents are referred to Dillon et al. (1990a)

## 7 Basic Ergonomic Issues

An electronic text is physically different from a paper one. Consequently, many researchers have examined these aspects of the medium in an attempt to explain the performance differences. An exhaustive programme of work conducted by Gould and his colleagues at IBM between 1982 and 1987 represents probably the most rigorous and determined research effort. They tried to isolate a single variable responsible for observed differences. The following sections review this work and related findings in the search for an explanation of the observed performance differences between reading from paper and reading from VDUs.

## 7.1 Orientation

One of the advantages of paper over VDUs is that it can be picked up and orientated to suit the reader. VDUs present the reader with text in a relatively fixed vertical orientation, though thanks to more ergonomic designs some flexibility to alter vertical orientation is now available in many systems. Gould et al (1987a) investigated the hypothesis that differences in orientation may account for differences in reading performance. Subjects were required to read three articles, one on a vertically positioned VDU, one on paper-horizontal and the other on paper-vertical (paper attached via copy-holder to equivalent VDU). Both paper conditions were read significantly faster than the VDU and there were no accuracy differences. While orientation has been shown to affect reading rate of printed material (Tinker, 1963) it does not explain the observed reading differences in the comparisons reported here.

## 7.2 Visual angle

Gould (1986) hypothesised that due to the usually longer line lengths on VDUs the visual angle subtended by lines in each medium differs and that people have learned to compensate for the longer lines on VDUs by sitting further away from them when reading. In an initial crude experiment of reading differences Gould (1986) visited the offices of 26 people who were reading either from VDU or paper and measured reading distance from both media with a metre stick. They found significantly greater reading distances for VDUs. Further work has confirmed that preferred viewing distance for screens is greater than that for paper (Jaschinski-Kruza 1990).

In a more controlled follow-up study Gould and Grischkowsky (1986) had 18 subjects read twelve different three-page articles for misspellings. Subjects read two articles at each of six visual angles: 6.7, 10.6, 16.0, 24.3, 36.4 and 53.4 degrees, varied by maintaining a constant reading distance while manipulating the image size used. Results showed that visual angle significantly affected speed and accuracy. However the effects were only noticeable for extreme angles, and between a range of 16.0 to 36.4 degrees, which covers typical VDU viewing, no effect for angle was found.

## 7.3 Aspect ratio

The term aspect ratio refers to the relationship of width to height. Typical paper sizes are higher than they are wider, while the opposite is true for typical VDU displays. Changing the aspect ratio of a visual field may affect eye movement patterns sufficiently to account for some of the performance differences. Gould (1986) had eighteen subjects read three 8-page articles on VDU, paper and paper-rotated (aspect ratio altered to resemble screen presentation). The results however showed little effect for ratio.



## 7.4 Dynamics

Detailed work has been carried out on screen filling style and rates (e.g., Bevan, 1981; Kolers et al, 1981; Schwartz et al, 1983) and findings suggest that variables such as rate and direction of scrolled text do influence performance and subjective ratings. In order to understand the role of dynamic variables such as scrolling, "jittering" and screen filling in reading from VDUs, Gould et al (1987a) had subjects read from paper, VDU and good quality photographs of the VDU material which maintained the screen image but eliminated any possible dynamics. Results provided little in the way of firm evidence to support the idea of dynamics causing problems. Subjects again read consistently faster from paper compared to both other presentation media, which did not differ significantly from each other. Creed et al (1987) also compared paper, VDU and photos of the screen display on a proofreading task with thirty subjects. They found that performance was poorest on VDU but photographs did not differ significantly from either paper or VDU in terms of speed or accuracy, though examination of the raw data suggested a trend towards poorer performance on photos than paper. It seems unlikely therefore that much of the cause for differences between the two media can be attributed to the dynamic nature of the screen image.

## 7.5 Flicker

Characters are written on a VDU by an electron beam which scans the phosphor surface of the screen, causing stimulated sections to glow temporarily. The phosphor is characterised by its persistence, a high-persistence phosphor glowing for longer than a low-persistence phosphor. In order to generate a character that is apparently stable it is necessary to rescan the screen constantly with the requisite pattern of electrons. The frequency of scanning is referred to as the refresh rate. Since the characters are in effect repeatedly fading and being regenerated it is possible that they appear to flicker rather than remain constant. The amount of perceived flicker will obviously depend on both the refresh rate and the phosphor's persistence; the more frequent the refresh rate and the longer the persistence, the less perceived flicker. However refresh rate and phosphor persistence alone are not sufficient to predict whether or not flicker will be perceived by a user. It is also necessary to consider the luminance of the screen. While a 30 Hz refresh rate is sufficient to eliminate flicker at low luminance levels, Bauer et al (1983) suggested that a refresh rate of 93 Hz was necessary in order for 99% of subjects to perceive a display of dark characters on a light background (i.e., positive presentation, see 7.6.) as flicker free.

If flicker was responsible for the large differences between reading from paper and VDU it would be expected that studies such as Creed et al's (1987) which employed photographs of screen displays would have demonstrated a significant difference between reading from photos and VDUs. However the extent to which flicker may have been an important variable in many studies is unknown as details of screen persistence and refresh rates are often not included in publications. Gould et al (1987a) admit that the photographs used in their study were of professional quality but appeared less clear than the actual screen display. It is likely that using photos to control flicker may not be a

suitable method and flicker may play some part in explaining the differences between the two media.

## 7.6 Image polarity

A display in which dark characters appear on a light background (e.g., black on white) is referred to as positive image polarity or negative contrast. This will be referred to here as positive presentation. A display on which light characters appear on a dark background (e.g., white on black) is referred to as negative image polarity or positive contrast. This will be referred to here as negative presentation. The traditional computer display involves negative presentation, typically white on black though light green on dark green is also common.

Since 1980 there has been a succession of publications concerned with the relative merits of negative and positive presentation. Several studies suggest that, tradition notwithstanding, positive presentation may be preferable to negative. For example Radl (1980) reported increased performance on a data input task for dark characters and Bauer and Cavonius (1980) reported a superiority of dark characters on various measures of typing performance and operator preference.

With regards to reading from screens Cushman (1986) reported that reading speed and comprehension on screens was unaffected by polarity, though there was a non-significant tendency for faster reading of positive presentation. Gould et al (1987a) specifically investigated the polarity issue. Fifteen subjects read 5 different 1000 word articles, 2 negatively presented, 2 positively presented and one on paper (standard positive presentation). Further experimental control was introduced by fixing the display contrast for one article of each polarity at a contrast ratio of 10:1 and allowing the subject to adjust the other article to their own liking. This avoided the possibility that contrast ratios may have been set which favoured one display polarity. Results showed no significant effect for polarity or contrast settings, though 12 of the 15 subjects did read faster from positively presented screens, leading the investigators to conclude that display polarity probably accounted for some of the observed differences in reading from screens and paper.

In a general discussion of display polarity Gould et al (1987b) state that:

"to the extent that polarity makes a difference it favours faster reading from dark characters on a light background." (p.514)

Furthermore they cite Tinker (1963) who reported that polarity interacted with type size and font when reading from paper. The findings of Bauer et al (1983) with respect to flicker certainly indicate how perceived flicker can be related to polarity. Therefore the contribution of display polarity in reading from screens is probably important through its interactive effects with other display variables.

## 7.7 Display characteristics

Issues related to fonts such as character size, line spacing and character spacing have been subjected to detailed research. However the relationship of much of the findings to reading continuous text from screens is not clear.

Character size on VDUs is closely related to the dimension of the dot matrix from which the characters are formed. In the sixties 5x7 matrices were used but they offer little opportunity for representing lower-case ascenders and descenders, and consequently produce poor legibility. The dramatic increase in computer processing power now means that there is little cost in employing larger matrices and Cakir et al (1980) recommend a minimum of 7x9. Pastoor et al (1983) studied the relative suitability of four different dot-matrix sizes and found reading speed varied considerably. On the basis of these results the authors recommended a 9 x13 character size matrix. However their study was concerned with television screens and their tasks included isolated word reading and column searching. In short, the optimum character size for reading from screens appears to be contingent on the task performed.

Considerable experimental evidence exists to favour proportionally rather than non-proportionally spaced characters (e.g., Beldie et al 1983). Once more though, the findings must be viewed cautiously. In the Beldie et al study for example, the experimental tasks did not include reading continuous text. Muter et al (1982) compared reading speeds for text displayed with proportional or non-proportional spacing and found no effect. In an experiment intended to identify the possible effect of such font characteristics on the performance differences between paper and screen reading, Gould et al (1987a) found no evidence to support the case for proportionally spaced text.

Kolers et al (1981) studied interline spacing and found that with single spacing significantly more fixations were required per line, fewer lines were read and the total reading time increased. However the differences were small and were regarded as not having any practical significance. On the other hand Kruk and Muter (1984) found that single spacing produced 10.9% slower reading than double spacing, a not inconsiderable difference.

Muter and Maurutto (1991) attempted various “enhancements” to screen presented text to see if they could improve reading performance. These included double spacing between lines, proportional spacing within words, left justification only and positive presentation. “Enhanced” text proved to be read no differently from more typical electronic text (i.e., basically similar to paper) which the authors state may be due to one or two of their “enhancements” having a negative and therefore neutralising effect on others or some “enhancements” interacting negatively. Unfortunately, their failure to manipulate such variables systematically means firm conclusions cannot be drawn.

Obviously much work needs to be done before a full understanding of the relative advantages and disadvantages of particular formats and types of display is achieved. In a discussion of the role of display fonts in explaining any of the observed differences

between screen and paper reading Gould et al (1987a) conclude that font has little effect on reading rate from paper (as long as the fonts tested are reasonable). They add that it is almost impossible however to discuss fonts without recourse to the physical variables of the computer screen itself e.g., screen resolution and beam size, once more highlighting the potential cumulative effect of several interacting factors on reading from screens.

### 7.8 Anti-aliasing

Most computer displays are raster displays typically containing dot matrix characters and lines which give the appearance of "staircasing" i.e. edges of characters may appear jagged. This is caused by undersampling the signal that would be required to produce sharp, continuous characters. The process of anti-aliasing has the effect of perceptually eliminating this phenomenon on raster displays. A technique for anti-aliasing developed by IBM accomplishes this by adding variations in grey level to each character.

The advantage of anti-aliasing lies in the fact that it improves the quality of the image on screen and facilitates the use of fonts more typical of those found on printed paper. To date the only reported investigation of the effects of this technique on reading from screens is that of Gould et al (1986). They had 15 subjects read three different 1000 word articles, one on paper, one on VDU with anti-aliased characters and one on VDU without anti-aliased characters. Results indicated that reading from anti-aliased characters did not differ significantly from either paper or aliased characters though the latter two differed significantly from each other. Although the trend was present the results were not conclusive and no certain evidence for the effect of anti-aliasing was provided. However the authors report that 14 of the 15 subjects preferred the anti-aliased characters, describing them as clearer and easier to read.

### 7.9 User characteristics

It has been noted that many of the studies reported in this review employed relatively naive users as subjects. The fact that different types of users interact with computer systems in different ways has long been recognised and it is possible that the differences in reading that have been observed in these studies result from particular characteristics of the user group involved.

Most obviously, it might be assumed that increased experience in reading from computers would reduce the performance deficits. A direct comparison of experienced and inexperienced users was incorporated into a study on proofreading from VDUs by Gould et al (1987a). Experienced users were described as "heavy, daily users.....and had been so for years". Inexperienced users had no experience of reading from computers. No significant differences were found between these groups, both reading slower from screen.

Smedshammar et al (1989) report that post-hoc analysis of their data indicate that fast readers are more adversely affected by VDU presentation than slow readers. However, their classification of reading speed is based on mean performance over three conditions

in their experiment rather than controlled, pre-trial selection suggesting caution in drawing conclusions. Smith and Savory (1989) report an interaction effect between presentation medium, reading strategy and susceptibility to external stress measured by questionnaire suggesting that working with VDUs may exaggerate some differences in reading strategy for individuals with high stress levels. Caution in interpretation of these results is suggested by the authors.

No reported differences for age or sex can be found in the literature. Therefore it seems reasonable to conclude that basic characteristics of the user are not responsible for the differences in reading from these presentation media.

#### 7.10 The interaction of display variables: the work of Gould et al.

Despite many of the findings reported thus far, it appears that reading from screens can at least be as fast and as accurate as reading from paper. Gould et al (1987b) have empirically demonstrated that under the right conditions such differences between the two presentation media disappear. In a study employing sixteen subjects, an attempt was made to produce a screen image that closely resembled the paper image i.e., similar font, size, colouring, polarity and layout were used. Univers-65 font was positively presented on a monochrome IBM 5080 display with an addressability of 1024 x 1024. No significant differences were observed between paper and screen reading. This study was replicated with twelve further subjects using a 5080 display with an improved refresh rate (60Hz). Again no significant differences were observed though several subjects still reported some perception of flicker.

On balance it appears that any explanation of these results must be based on the interactive effects of several of the variables outlined in the previous sections. After a series of experimental manipulations aimed at identifying those variables responsible for the improved performance Gould et al (1987b) suggested that the performance deficit was the product of an interaction between a number of individually non-significant effects. Specifically, they identified display polarity (dark characters on a light, whitish background), improved display resolution, and anti-aliasing as major contributions to the elimination of the paper/screen reading rate difference.

Gould et al (1987b) conclude that the explanation of many of the reported differences between the media is basically visual rather than cognitive and lies in the fact that reading requires discrimination of characters and words from a background. The better the image quality is, the more reading from screen resembles reading from paper and hence the performance differences disappear. This seems an intuitively sensible conclusion to draw. It reduces to the level of simplistic any claims that one or other variable such as critical flicker frequency, font or polarity are responsible for any differences. As technology improves we can expect to see fewer speed deficits at least for reading from screens. Recent evidence from Muter and Maurutto (1991) using a commercially available screen has shown this to be the case, although other differences remain.

## 7.11 Conclusion

Although reading from computer screens may be slower and occasionally less accurate than reading from paper, no one variable is likely to be responsible for this difference. It is almost certain that neither inherent problems with the technology nor the reader are causal factors. Invariably it is the quality of the image presented to the reader which is crucial. Tinker (1963) reports dramatic interaction effects of image quality variables on paper and according to Gould et al (1987a) it is likely that these occur on screen too. Positive presentation combined with a high screen resolution to avoid flicker can produce good images and with the addition of anti-aliased characters it becomes possible to provide a screen display that resembles the print image and thereby facilitates reading. It must be remembered however that typical computer displays present images that are still of poorer quality than those used by Gould and his associates to overcome the performance deficit. Until screen standards are raised sufficiently these differences are likely to remain.

A major shortcoming of the studies by Gould et al is that they only address limited outcome variables: speed and accuracy. Obviously speed is not always a relevant criterion in assessing the output of a reading task. Furthermore, the accuracy measures taken in these studies have been criticised as too limited and further work needs to be carried out to appreciate the extent to which the explanation offered by Gould is sufficient. It follows that other observed outcome differences such as fatigue, reader preference and comprehension should also be subjected to investigation in order to understand how far the image quality hypothesis can be pushed as an explanation for reading differences between the two media.

A shortcoming of most work cited in this section is the task employed. Invariably it was proofreading which hardly constitutes normal reading for most people. Thus the ecological validity of many of these studies is low. Beyond this, the actual texts employed were all relatively short (Gould's for example averaged only 1100 words but many other researchers used even shorter texts). As a result, it is difficult to generalise these conclusions beyond the specifics of task and texts employed to the wider class of activities termed "reading". Creed et al (1987) defend the use of proofreading on the grounds of its amenability to manipulation and control. While this desire for experimental rigour is laudable one cannot but feel that the major issues involved in using screens for real-world reading scenarios are not addressed by such work. With this in mind, the following section considers the literature on research concerned with the manipulation facilities where of necessity, lengthy texts need to be employed.

## 8 Manipulation Facilities

It is clear that the search for the specific ergonomic variables responsible for differences between the media has been insightful. However, few readers of electronic texts would be satisfied with the statement that the differences between the media are visual rather than cognitive. This might explain absolute speed and accuracy differences on limited tasks

but hardly accounts for the range of process differences that are found as described earlier.

Once the document becomes too large to display on a single screen other factors than image quality immediately come into play. At this stage readers must start to manipulate the document and thus be able to relate current to previously-displayed material. In such a situation other factors such as memory for text and its location, ability to search for items and speed of movement through the document come into play and the case for image quality as the major determinant of performance is less easy to sustain. Several researchers have pinned their hopes on improved manipulation facilities with electronic texts removing many of the differences between the media. In this section, research into variables affecting such issues is reviewed.

### 8.1 Scrolling versus paging

The manner in which a reader moves through a document is distinctly different in either medium and even within the electronic medium, various techniques are employed for displaying sections of the text. Scrolling (the facility to move the text up or down on the screen smoothly by a fixed increment to reveal information currently out of view) and paging (the facility to move the text up or down in complete screensful - akin to page turning with paper texts) are two of the most common.

There is evidence to suggest that readers establish a visual memory for the location of items within a printed text based on their spatial location both on the page and within the document (Rothkopf, 1971; Lovelace and Southall, 1983). This memory is supported by the fixed relationship between an item and its position on a given page. A scrolling facility is therefore liable to weaken these relationships and offers the reader only the relative positional cues that an item has with its immediate neighbours.

On the basis of a literature review, Mills and Weldon (1986) report that there is no real performance difference between scrolling and paging though Schwartz et al. (1983) found that novices tend to prefer paging (probably based on its close adherence to the book metaphor) and Dillon et al (1990b) report that a scrolling mechanism was the most frequently cited improvement suggested by subjects assessing their reading interface.

Scrolling has also been investigated in conjunction with direction (vertical or horizontal—Sekey and Tietz, 1982), rate (self-paced or machine-paced—Kolers et al., 1981) and display size (Duchnicky and Kolers, 1983). With reference to direction and rate, all seem to conclude that ideally, lengthy texts should be presented vertically and at the reader's choice of rate. Even so, Kolers et al. (1981) report that forcing readers to increase their rates by 10-20% does not lead to loss of comprehension and actually appears to increase efficiency of eye-movements as measured by rate and length of fixation.

It seems therefore that scrolling is a popular form of text manipulation with more experienced users probably due to its speed even if there are theoretical grounds for

doubting its superiority over paging. There is no firm evidence that either facility significantly affects reading performance compared to paper.

## 8.2 Display size

Display size is a much discussed but infrequently studied aspect of human-computer interaction in general and reading electronic text in particular. Popular wisdom suggests that “bigger is better” but empirical support for this edict is sparse. Duchnicky and Kolars (1983) investigated the effect of display size on reading constantly scrolling text and reported that there is little to be gained by increasing display size to more than 4 lines either in terms of reading speed or comprehension. Elkerton and Williges (1984) investigated 1,7,13, and 19-line displays and reported that there were few speed or accuracy advantages between the displays of 7 or more lines. Similarly, Neal and Darnell (1984) report that there is little advantage in full page over partial page displays for text-editing tasks.

These results seem to suggest that there is some critical point in display size, probably around 5 lines, above which improvements are slight. Intuitively this seems implausible. Few readers of paper texts would accept presentations of this format. Experiences with paper suggest that text should be displayed in larger units than this. Furthermore, loss of context is all too likely to occur with lengthy texts and the ability to browse and skim backward and forward is much easier with 30 or so lines of text than with 5 line displays. Of the experiments cited, only the Duchnicky and Kolars study was concerned with reading for comprehension and their passages were never longer than 300 words. Thus their findings on window size seem to bear little relevance to reading of lengthy texts.

Deliberately examining this, Richardson et al (1989) had subjects perform 10 information location tasks using an electronic book with a display size of 20 or 40 lines. Though they observed no performance differences between conditions they did report a significant preference effect favouring the larger display. Similarly Dillon et al (1990b) investigated screen sizes of 20 and 60 lines for reading an electronic version of an academic article. Interestingly they found a manipulation effect for screen size that could not be explained by the fact that to read a complete text on a small screen necessitates more manipulations than seeing it on a large one. They reported that when such simple manipulations are discounted and attention is paid only to changes in direction or jumps of 2 or more “pages”, readers using the small screen still manipulated the text more. They proposed that the likeliest explanation was that readers like to re-read large parts of texts or jump about when using articles and that the smaller screen condition required more manipulations to observe the same amount of text as the bigger screen. As in the Richardson et al study, the authors report a preference effect favouring the larger display.

As with many variables, the task being performed is likely to be a deciding factor. Small screens pose problems for readers wishing to browse through lengthy texts but are likely to be more acceptable for tasks requiring a straight perusal of short material such as a letter or memo. Significantly, many applications now allow the user to change window size within the constraints of the overall screen size which may accommodate some



preference differences but does not resolve issues to do with optimum screen size for particular tasks.

It is likely that many of the effects of screen size are too subtle to be assessed by gross outcome measures such as speed and accuracy. Larger screens might suit better spatial memory formation or browsing, variables that are not usually measured by investigators. As concluded in the basic ergonomic research, it is likely that the interaction of size with other manipulation variables is important.

### 8.3 Text splitting across screens

A related issue to display size and scrolling/paging is the splitting of paragraphs mid-sentence across successive screens. In this case, which is more likely to occur in small displays, the reader must manipulate the document in order to complete the sentence. This is not a major issue for paper texts such as books or journals because the reader is usually presented with two pages at a time and access to previous pages is normally easy. On screen however, access rates are not so fast and the break between screens of text is likely to be more critical.

Research into reading has clearly demonstrated the complexity of the cognitive processing that occurs. The reader does not simply scan and recognise every letter in order to extract the meaning of words and then sentences. Comprehension is thought to require inference and deduction, and the skilled reader probably achieves much of his/her smoothness by predicting probable word sequences (Chapman and Hoffman, 1977 though see Mitchell 1982). The basic units of comprehension in reading that have been proposed are propositions (Kintsch, 1974), sentences (Just and Carpenter, 1980) and paragraphs (Mandler and Johnson, 1977). Splitting sentences across screens is likely to disrupt the process of comprehension by placing an extra burden on the limited capacity of working memory to hold the sense of the current conceptual unit while the screen is filled. Furthermore, the fact that between 10-20% of eye movements in reading are regressions to earlier fixated words and that significant eye movement pauses occur at sentence ends (Ellis, 1983) would suggest that sentence splitting is also likely to disrupt the reading process and thereby hinder comprehension.

In the Dillon et al (1990b) study cited earlier, the role of text splitting on performance was also examined. They found that splitting text across screens caused readers to return to the previous page to re-read text significantly more often than when text was not split. Though this appeared to have no effect on subsequent comprehension of the material being read, they concluded that it was remarked upon by the subjects sufficiently often to suggest that it would be a nuisance to regular users. In this study however the subjects were reading from a paging rather than scrolling interface where the effect of text splitting was more likely to cause problems due to screen-fill delays. With scrolling interfaces text is always going to split across screen boundaries but there is rarely a perceptible delay in image presentation to disrupt the reader. It would seem therefore that to the extent to which such effects are likely to be noticeable, text splitting should be avoided for paging interfaces.

## 8.4 Window format

It has become increasingly common to present information on computer screen via windows i.e., sections of screen devoted to specific groupings of material. Current technology supports the provision of independent processes within windows or the linking of inputs in one window with the subsequent display in another, the so called “co-ordinated windows” approach (Shneiderman 1987).

Such techniques have implications for the presentation of text on screen as they provide alternatives to the straightforward listing of material in “scroll” form or as a set of “pages”. For example, while one window might present a list of contents in an electronic text, another might display whole sections of it according to the selection made. In this way, not only is speed of manipulation increased but the reader can be provided with an overview of the document’s structure to aid orientation while reading an opened section.

The use of such techniques is now commonplace in hypertext applications. GUIDE for example, uses windows in one instance to present short notes or diagrams as elaborations or explanations of points raised in the currently viewed text, rather like sophisticated footnotes. The concept of hypertext as non-linear text is, in a very real sense, derived from such presentation facilities.

Tombaugh et al (1987) investigated the value of windowing for readers of lengthy electronic texts. They had subjects read two texts on single or multi-window formats before performing 10 information location tasks. They found that novices initially performed better with a single-window format but subsequently observed that, once familiar with the manipulation facilities, the benefits of multi-windowing in terms of aiding spatial memory became apparent. They highlight the importance of readers acquiring familiarity with a system and the concept of the electronic book in order to accrue the benefits of such facilities.

Simpson (1989) compared performance with a similar multi-window display, a “tiled” display (in which the contents of each window were permanently visible) and a ‘conventional’ stack of windows (in which the windows remained in reverse order of opening). She reported that performance with the conventional window stack was poorest but that there was no significant difference between between the “tiled” and multi-window displays. She concluded that for information location tasks, the ability to see a window's contents is not as important as being able to identify a permanent location for a section of text.

Stark (1990) asked people to examine a hypertext document in order to identify appropriate information for an imaginary client and manipulated the scenario so that readers had to access information presented either in a ‘pop-up’ window which appeared in the top right hand corner of the screen or a ‘replacement’ window which overlaid the information currently being read. Though no significant task performance or navigation effects were observed, subjects seemed more satisfied with pop-ups than replacements.

Such studies highlight the impact of display format on readers' performance of a standard reading task: information location. Spatial memory seems important and paper texts are good at supporting its use through permanence of format. Windowing, if deployed so as to retain order can be a useful means of overcoming this inherent weakness of electronic text. However, studies examining the problems of windowing very long texts (where more than five or six stacked windows or more frequent window manipulations are required) need to be performed before any firm conclusions about the benefits of this technique can be drawn.

## 8.5 Search facilities

Electronic text supports word or term searches at rapid speed and with total accuracy and this is clearly an advantage for users in many reading scenarios e.g. checking references, seeking relevant sections, etc. Indeed it is possible for such facilities to support tasks that would place unreasonable demands on users of paper texts e.g., searching a large book for a non-indexed term or several volumes of journals for references to a concept.

Typical search facilities require the user to input a search string and choose several criteria for the search such as ignoring certain text forms (e.g., all uppercase words) but sophisticated facilities on some database systems can support specification of a range of texts to search. The usual form for search specification is Boolean, i.e., users must input search criteria according to formal rules of logic employing the constructs 'either', 'or' as well as 'and', which when used in combination support powerful and precise specifications. Unfortunately most end-users of computer systems are not trained in their use and while the terms may appear intuitive, they are often difficult to employ successfully.

In current electronic text facilities a simple word search is most common but users still seem to have difficulties. Richardson et al (1988) reported that several subjects in their experiment displayed a tendency to respond to unsuccessful searches by increasing the specificity of the search string rather than lessening it. The logic appeared to be that the computer required precision rather than approximation to search effectively. While it is likely that such behaviour is reduced with increased experience of computerised searching, a study by McKnight et al (1989) of information location within text found other problems. Here, when searching for the term "wormwood" in an article on wine making, two subjects input the search term "woodworm", displaying the intrusion of a common sense term for an unusual word of similar sound and shape (a not uncommon error in reading under pressure due to the predictive nature of this act during sentence processing). When the system correctly returned a "Not Found" message, both users concluded that the question was an experimental trick.

Thus it seems as if search facilities are a powerful means of manipulating and locating information on screen and convey certain advantages impossible to provide in the paper medium. However, users may have difficulties with them in terms of formulating accurate search criteria. This is an area where research into the design of search facilities

and increased exposure of users to electronic information can lead to improvements resulting in a positive advantage of electronic text over paper.

## 8.6 Input device

Over the last 15 years numerous input devices have been designed and proposed as optimal for users e.g., trackball, mouse, function keyboard, joystick, light pen etc. Since Card et al's (1978) claim that the speed of text selection via a mouse was constrained only by the limits of human information processing, this device has assumed the dominant position in the market.

It has since become clear that, depending on the task and users, other input devices can significantly outperform the mouse (Milner 1988). For example, when less than ten targets are displayed on screen and the cursor can be made to jump from one to the next, cursor keys are faster than a mouse (Shneiderman 1987). In the electronic text domain, Ewing et al (1986) found this to be case with the HyperTIES application, though there is reason to doubt their findings as the mouse seems to have been used on less than optimal surface conditions.

Though 'direct manipulation' (Shneiderman 1984) might be a common description of an interface, it seems that its current manifestations leave much to be desired when it comes to manipulating text. Obviously practice and experience will play a considerable part here. Expertise with an input device affords the user a high level of control and breeds a sense of immediacy between selection and action.

It is important to realise that the whole issue of input device cannot be separated from other manipulation variables such as scrolling or paging. For example, a mouse that must be used in conjunction with a menu for paging text will lead to different performance characteristics than one used with a scroll bar. For the moment however the mouse appears dominant and as the "point and click" concept becomes integrated with the "look and feel" of hypertext it will prove difficult to replace, even if convincing experimental evidence against its use, or an innovative credible alternative should emerge.

## 8.7 Icon design

In aiding the manipulation of documents electronically, icons have become popular in many hypertext applications. GUIDE, for example, uses such forms as boxes, arrows and circles when the cursor moves over an actionable area of the document, while HyperCard provides numerous "button" shapes that cause different document manipulations to occur. Used in conjunction with a mouse such facilities can support rapid, easy manipulations of the text and allow the user to access the document through numerous routes – giving rise to the notion of non-linearity in hypertext.

Icons are also used to represent a document in situations where the user might be selecting one of several texts. While it is easy enough to convey an image of book or

other text type iconically few systems attempt to provide the range of cues available with paper such as size, age, level of usage and so forth.

There are sound theoretical grounds for supporting iconic representation. Being language independent icons convey information by pictographic means and should thus support use by individuals unfamiliar with the terminology of operating systems and command languages. Further advantages of iconic representations are that they utilise little display space and render syntax errors obsolete (Gittens 1986)

On the negative side, icons can be confusing if their form provides no immediate clue to their action. Arrows, trashcans and folders might be intuitive but this is not always the case (the “home” icon on HyperCard is a picture of a little house and naive users have failed to appreciate the intended reference [McKnight et al 1989]). Designing icons to convey less obvious actions than “goto” is not a simple task. Some designers even provide icons with textual descriptors to provide clues to their use which seems to defeat the purpose.

Stammers et al (1989) reported that icons are most useful when they represent concrete rather than abstract actions which while intuitively sensible, suggests ultimate limitations on their use as many computer functions are highly abstract in nature. Brems and Whitten (1987) found that icons were more appropriate for experienced than novice users which is ironic given the stated benefits of icons.

Generalising such findings to the electronic text domain is difficult at present. A reasonable conclusion seems to be that icons have a role, particularly for simple or repetitive actions such as “go there” or “look at this in more detail” but are less applicable for conveying information of abstract actions. For manipulation purposes the basic range of actions is always likely to be limited therefore it is conceivable that standard designs for such actions might appear soon. Obviously this is an area for further research.

## 8.8 Conclusion

Manipulating electronic text is considered to be more difficult than manipulating paper. Research suggests that factors such as non-splitting of text, rapid response and increased display size can improve matters and that facilities such as searching and multi-windowing might even offer benefits to electronic text over paper.

As with the basic ergonomic issues reviewed earlier the interaction of several of these variables is likely to be crucial. Small displays limit windowing facilities and may increase text-splitting causing manipulation differences with paper that might not emerge with large, multi-windowed displays. Furthermore, as Tombaugh et al (1987) pointed out familiarity with the facilities is vital. It is not always clear from the literature how this variable has been controlled in many studies.

The range of tasks used for such investigations is much wider and often more ecologically valid than those used in the basic ergonomic work reviewed. However, the

increased variability in both text size and task range mean that comparisons between studies are more difficult than for studies concerned with visual ergonomics. For example, the Dillon et al (1990b) investigated screen size effects by asking subjects to read an academic text for comprehension purposes, allowing them to manipulate the text by a paging mechanism while Duchnicky and Kolers (1983) investigated the same variable using different window sizes, short test texts, different comprehension techniques, with subjects using a knob to control scrolling rate. Obviously in such situations, comparisons are difficult.

As an explanation of the differences between the media, manipulation must be incomplete. Even if combined with good image quality, optimum manipulation facilities are unlikely to remove all the problems associated with electronic text. This is becoming obvious from much of the recent work on hypertext that is concerned with structuring information and has shown that even with high quality screens and supposedly optimum input devices such as a mouse, paper may still prove more usable than screen presented text for some tasks (e.g., McKnight et al 1990). In other words, even by making images clear, and supporting readers manipulating the text, we are still missing something else. Unfortunately, empirical data on reading from paper and screen largely stops here and we enter the realm of conjecture and theorising about “information structures” and “hyperspace” and out of the experimental data domain that is of concern in this review.

## 9. General Conclusion

At the outset it was stated that reading can be assessed in terms of outcome and process measures. To date however, most experimental work has concentrated on the former and in particular, has been driven by a desire to identify a single variable to account for the significant reading speed differences that have been reported. The present review sought to examine the experimental literature with a view to identifying all relevant issues and show how single variable explanations are unlikely to offer a satisfactory answer.

While substantial progress has been made in terms of understanding the impact of image quality on reading speed, it is clear that ergonomists are still a long way from understanding fully the effect of presentation medium on reading. While it is now possible to draw up recommendations on how to ensure no speed deficit for proofreading short texts on screen, changes in task and text parameters mean such advice has less relevance.

One is struck in reviewing this literature by the rather limited and often distorted view of reading that ergonomists seem to have. Most seem to concern themselves with the control of so many variables that the resulting experimental task bears little resemblance to the activities most of us routinely perform under the banner “reading”. It is perhaps no coincidence that the major stumbling block of reader preference has been so poorly investigated beyond the quick rating of screens and test documents in post-experimental surveys.

The assumption that overcoming speed or accuracy differences in proofreading is sufficient to claim, as some authors have, that “there is no difference” between the media (Osborne and Holton 1988) is testimony to the limitations of some ergonomists’ views of human activities such as reading. Other tasks, such as reading to comprehend, to learn or for entertainment are less likely to require readers to concern themselves with speed. These are the sort of tasks people will regularly wish to perform and it is important to know how electronic text can be designed to support them. Such tasks will also of necessity involve a wide variety of texts, differing in length, detail, content-type and so forth— issues that have barely been touched upon to date by researchers.

The findings on image quality and the emerging knowledge of manipulation problems should not be played down however. Knowing what makes for efficient visual processing and control of electronic text can serve as a basis for future applications. As Muter and Maurutto (1991) demonstrated, a typical high quality screen with effective manipulation facilities can provide an environment that holds its own in speed, comprehension and preference terms with paper, at least over the relatively constrained reading scenarios found in the researchers’ laboratory. But if our desire is to create systems that improve on paper rather than just matching it in performance and satisfaction terms (as it should be) then much more work and a more realistic conceptualisation of human reading is required.

## References

Askwall, S. (1985) Computer supported reading vs reading text on paper: a comparison of two reading situations. *International Journal of Man-Machine Studies*, 22, 425-439.

Bauer, D. & Cavonius, C. R. (1980) Improving the legibility of visual display units through contrast reversal. In E. Grandjean and E. Vigliani (Eds.) *Ergonomic Aspects of Visual Display Terminals*. London: Taylor and Francis.

Bauer, D., Bonacker, M. and Cavonius, C.R. (1983) Frame repetition rate for flicker-free viewing of bright VDU screens. *Displays*, January, 31-33.

Beldie, I. P., Pastoor, S. & Schwarz, E. (1983) Fixed versus variable letter width for televised text. *Human Factors*, 25(3), 273-277.

Belmore, S. (1985) Reading computer presented text. *Bulletin of the Psychonomic Society*, 23(1), 12-14.

Bevan, N. (1981) Is there an optimum speed for presenting text on VDUs. *International Journal of Man-Machine Studies*, 14, 59-76.

Brems, D. and Whitten W. (1987) Learning and preference for icon-based interfaces. In: Proc. of the 31st Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Soc. 125-129.

Brown, P. (1988). Hypertext: the way forward. In J.C. Van Vliet (ed.) Proceedings of EP88. Cambridge University Press

Cakir, A., Hart, D. J. & Stewart, T. F. M. (1980) Visual Display Terminals. Chichester: John Wiley and Sons.

Conklin, J. (1987) Hypertext: an introduction and survey. Computer, September, 17-41.

Card, S., English, W. and Burr, B (1978) Evaluation of mouse, rate-controlled isometric joystick, step keys and text keys for text selection on a CRT. Ergonomics, 21, 601-613.

Chapman, L. J. and Hoffman, M. (1977) Developing Fluent Reading. Milton Keynes: Open University Press.

Creed, A., Dennis, I. & Newstead, S. (1987) Proof-reading on VDUs. Behaviour and Information Technology, 6(1), 3-13.

Cushman, W. H. (1986) Reading from microfiche, VDT and the printed page: subjective fatigue and performance. Human Factors, 28(1), 63-73.

Dillon, A. (1990) The human factors of hypertext. International Forum on Information and Documentation 15(4) 32-36.

Dillon, A., Richardson, J. and McKnight, C. (1990a) Navigation in Hypertext: a critical review of the concept. In D.Diaper, D.Gilmore, G.Cockton and B.Shackel (eds.) INTERACT'90. North Holland: Amsterdam.

Dillon, A., Richardson, J. and McKnight, C. (1990b) The effect of display size and paragraph splitting on reading lengthy text from screen. Behaviour and Information Technology 9 (3) 215-227

Duchnicky, R.L. and Kolers P.A. (1983) Readability of text scrolled on a visual display terminal as a function of window size. Human Factors, 25(6), 683-692.

Edwards, D. and Hardman, L. (1989) "Lost in Hyperspace": Cognitive Mapping and Navigation in a Hypertext Environment. In R. McAleese (ed.) Hypertext: Theory into Practice, Oxford: Intellect.

Egan, D., Remde, J., Landauer, T., Lochbaum, C. and Gomez, L. (1989) Behavioural evaluation and analysis of a hypertext browser. Proceedings of CHI'89, ACM: New York, 205-210



- Elkerton, J. and Williges, R. (1984) Information retrieval strategies in a file search environment. *Human Factors*, 26(2), 171-184.
- Ellis, A. (1983) *Reading, Writing and Dyslexia*. London: Lawrence Erlbaum Associates.
- Elm, W. and Woods, D. (1985) Getting lost: a case study in interface design. *Proceedings of the Human Factors Society 29th Annual Meeting*, 927-931.
- Ewing, J., Mehrabanzad, S. Sheck, S., Ostroff, D. and Shneiderman, B. (1986) An experimental comparison of a mouse and arrow-jump keys for an interactive encyclopedia. *International Journal of Man-Machine Studies*, 24, 1, 29-45.
- Garland, J. (1982) Ken Garland and Associates: Designers-20 years work and play. Cited in, Waller, R. (1987) *The typographic contribution to language: towards a model of typographic genres and their underlying structures*. PhD Thesis, Dept. of Typography and Graphic Communication, University of Reading.
- Gittens, D. (1986) Icon based human-computer interaction. *International Journal of Man-Machine Studies* 24, 519-543.
- Gordon, S. Gustavel, J., Moore, J. and Hankey, J. (1988) The effects of hypertext on reader knowledge representation, *Proceedings of the Human Factors Society 32nd Annual Meeting*
- Gould, J. D. & Grischkowsky, N. (1984) Doing the same work with hard copy and cathode-ray tube (CRT) computer terminals. *Human Factors*, 26(3), 323-337.
- Gould, J. D. & Grischkowsky, N. (1986) Does visual angle of a line of characters affect reading speed? *Human Factors*, 28(2), 165-173.
- Gould, J.D., Alfaro, L. Barnes, V., Finn, R., Grischkowsky, N. and Minuto, A. (1987a) Reading is slower from CRT displays than from paper: Attempts to isolate a single variable explanation. *Human Factors*, 29(3)269-299.
- Gould, J.D., Alfaro, L., Finn, R., Haupt, B. and Minuto, A. (1987b) Reading from CRT displays can be as fast as reading from paper. *Human Factors* 29(5), 497-517.
- Hammond, N. and Allinson, L. (1989) Extending hypertext for learning: an investigation of access and guidance tools. In: A. Sutcliffe and L. Macaulay (eds.) *People and Computers V*. Cambridge: Cambridge University Press, 293-304.
- Helander, M. G., Billingsley, P. A. & Schurick, J. M. (1984) An evaluation of human factors research on visual display terminals in the workplace. *Human Factors Review*, Chapter 3, 55 - 129.

Jaschinski-Kruza, W. (1990) On the preferred viewing distances to screen and document at VDU workplaces. *Ergonomics*, 33(8),1055-1063.

Just, M.A. and Carpenter, P. (1980) A theory of reading: from eye movements to comprehension. *Psychological Review*, 87 (4), 329-354.

Kak, A. V. (1981) Relationships between readability of printed and CRT-displayed text. *Proceedings of Human Factors Society - 25th Annual Meeting*, 137 - 140.

Kerr S. T.(1986) Learning to use electronic text: an agenda for research on typography, graphics, and interpanel navigation. *Information Design Journal*.4, 3. 206-211.

Kintsch, W. (1974) *The Representation of Meaning in Memory* Hillsdale, N.J.: Lawrence Earlbaum Associates.

Kolers, P. A., Duchnicky, R. L. & Ferguson, D. C. (1981) Eye movement measurement of readability of CRT displays. *Human Factors*, 23(5), 517-527.

Kruk, R. S. & Muter, P. (1984) Reading continuous text on video screens. *Human Factors*, 26(3), 339-345.

Licklider, J. (1965) *Libraries of the Future* Cambridge MA: MIT Press

Lovelace, E. A. and Southall, S. D. (1983) Memory for words in prose and their locations on the page. *Memory and Cognition*, 11 (5), 429-434.

McAleese, R. (1989) Navigation and browsing in Hypertext. In R. McAleese (ed.) *Hypertext:Theory into Practice*. Oxford: Intellect.

McKnight, C., Dillon, A. and Richardson, J. (1990) A comparison of linear and hypertext formats in information retrieval. In: R. McAleese and C. Green (eds.) *Hypertext: State of the Art*. Oxford:Intellect.

Mandler and Johnson (1977) Remembrance of things parsed: story structure and recall. *Cognitive Psychology*, 9, 111-151.

Mills, C.B. and Weldon, L.J. (1986) Reading text from computer screens. *ACM Computing Surveys*, 19, 329-358.

Milner,N. (1988)A review of human performance and preference with different input devices to computer systems, in: D.Jones and R.Winder (eds.) *People and Computers IV*, Cambridge: Cambridge University Press.

Mitchell, D. (1982) *The Process of Reading*. Chichester: Wiley.

Monk,A., Walsh, P. and Dix,A. (1988) A comparison of hypertext, scrolling, and folding as mechanisms for program browsing, in: D.Jones and R.Winder (eds.) People and Computers IV, Cambridge: Cambridge University Press.

Muter, P., Latremouille, S. A., Treurniet, W. C. & Beam, P. (1982) Extended reading of continuous text on television screens. *Human factors*, 24(5), 501-508.

Muter, P. and Maurutto, P. (1991) Reading and skimming from computer screens and books: the paperless office revisited? *Behaviour and Information Technology*, 10(4) 257-266.

Neal, A. and Darnell, M. (1984) Text editing performance with partial line, partial page and full page displays. *Human Factors* 26(4), 431-441.

Nelson, T., (1987) *Literary Machines Abridged Electronic Version* 87.1

Nisbett, R. and Wilson, T. (1977) Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84, 231-259.

Olshavsky, J. (1977) Reading as problem solving: an investigation of strategies. *Reading Research Quarterly*, 4, 654-674.

Oborne,D. and Holton, D. (1988) Reading from screen versus paper: there is no difference. *International Journal of Man-Machine Studies*, 28, 1,1-9.

Pastoor, S., Schwarz, E. and Beldie, I. P. (1983) The relative suitability of four dot-matrix sizes for text presentation on colour television screens. *Human Factors*, 25(3), 265-272.

Pearce, B. (Ed) *Health Hazards of VDUs?* Chichester: John Wiley and Sons.

Pugh, A. (1979) Styles and strategies in adult silent reading. In: P.Kolers, M.Wrolstad, and H. Bouma (eds.) *Processing of Visible Language* 1. Plenum Press: London.

Radl, G.W. (1980)Experimental investigations for optimal presentation mode and colours of symbols on the CRT screen. In E. Grandjean and E. Vigliani (Eds.) *Ergonomic Aspects of Visual Display Terminals*. London: Taylor and Francis.

Richardson,J., Dillon,A. and McKnight, C. and Saadat-Samardi, M. (1988) The manipulation of screen presented text: experimental investigation of an interface incorporating a movement grammar. HUSAT Memo #431

Richardson, J., Dillon, A. and McKnight, C. (1989) The effect of window size on reading and manipulating electronic text. In E. Megaw (ed.) *Contemporary Ergonomics* 1989. London: Taylor and Francis.

Rothkopf, E. Z. (1971) Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behaviour*, 10, 608-613.

Sauter, S., Gottlieb, M., Rohrer, K. and Dodson, V. (1983) The well-being of video display terminal users: an exploratory study. Report No: 210-79-0034. Cincinnati, OH: US Dept. of Health and Human Sciences.

Schwartz, E., Beldie, I. and Pastoor, S. (1983) A comparison of paging and scrolling for changing screen contents by inexperienced users. *Human Factors*, 25, 279-282.

Sekey, A. and Tietz, J. (1982) Text display by "saccadic scrolling". *Visible Language*, 17, 62-77.

Shneiderman, B. (1984) The future of interactive systems and the emergence of direct manipulation, in: Y. Vassiliou (ed.) *Human Factors and Interactive Computer Systems*, Norwood, N.J.: Ablex.

Shneiderman, B. (1987) *Designing the User Interface: strategies for effective human-computer interaction* San Francisco: Addison Wesley.

Schumacher, G. and Waller, R. (1985) Testing design alternatives: a comparison of procedures. In: T. Duffy and R. Waller (eds.) *Designing usable texts*, Orlando, FL: Academic Press.

Shulman, H. Shute, S. and Weissmann, O. (1985) Icons versus names as command designators in text editing. *Proceedings of the International Conference on Cybernetics and Society*. New York: IEEE, 268-272.

Simpson, A. (1989) Navigation in hypertext: design issues. Paper presented at International OnLine Conference '89, London, December.

Smedshammar, H., Frenckner, K., Nordquist, C. and Romberger, S. (1989) Why is the difference in reading speed when reading from VDUs and from paper bigger for fast readers than for slow readers? Paper presented at WWDU 1989, Second International Scientific Conference, Montreal.

Smith, A. and Savory, M. (1989) Effects and after-effects of working at a VDU: Investigation of the influence of personal variables. In: E.D. Megaw (ed.) *Contemporary Ergonomics 1989*. London: Taylor and Francis.

Stammers, R., George, D. and Carey, M. (1989) An evaluation of abstract and concrete icons for a CAD package. *Contemporary Ergonomics 1989*, London: Taylor and Francis.

Starr, S.J. (1984) Effects of video display terminals in a business office. *Human Factors*, 26, 347-356.

Switchenko, D. M.(1984) Reading from CRT versus paper: the CRT disadvantage hypothesis re-examined. Proceedings of Human Factors Society, 28th Annual Meeting, Santa Monica, CA: Human Factors Society, 429-430.

Tinker, M.A. (1958) Recent studies of eye movements in reading. Psychological Bulletin, 55, 215-231.

Tinker, M. A. (1963) Legibility of Print. Ames, Iowa: Iowa State University Press.

Tombaugh, J. Lickorish, A. and Wright, P. (1987) Multi-window displays for readers of lengthy texts, International Journal of Man-Machine Studies 26,5, 597-616

Waern, Y. & Rollenhagen, C. (1983) Reading text from visual display units (VDUs). International Journal of Man-Machine Studies, 18, 441-465.

Waller, R. (1984) Designing government forms: a case study Information Design Journal 4, 36-57.

Waller, R. (1986) What electronic books will have to be better than. Information Design Journal 5, 72-75.

Whalley, P. and Fleming, R. (1975) An experiment with a simple recorder of reading behaviour, Programmed Learning and Educational Technology, 12, 120-124.

Wilkinson , R.T. and Robinshaw, H.M. (1987) Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. Behaviour and Information Technology, 6(2), 125-133.

Wright, P. and Lickorish, A. (1983) Proof-reading texts on screen and paper. Behaviour and Information Technology, 2, (3), 227-235.