

Reading From Paper versus Reading From Screen

A. DILLON, C. MCKNIGHT AND J. RICHARDSON

HUSAT Research Centre, Department of Human Sciences, Loughborough University of Technology, The Elms, Elms Grove, Loughborough, Leics. LE11 1RG

*This paper reviews the literature on reading continuous text from VDUs. The focus is on the reported nature, and potential causes, of reading differences between paper and screens. The first section outlines the scope of the present review. Section 2 discusses the nature of the reported differences between reading from either presentation medium. Five broad differences have been identified suggesting that reading from VDUs is slower, less accurate, more fatiguing, decreases comprehension and is rated inferior by readers. Evidence for the existence of each of these differences is reviewed and conclusions are drawn. In Section 3, ten variables which have been proposed as potential causes of reading differences between paper and screen are reviewed. These include screen dynamics, display polarity, orientation, viewing angle and user characteristics. Recent evidence by Gould *et al.*¹¹ is presented which suggests that the image quality of the screen display is the crucial factor and indicate that positive presentation, high resolution and anti-aliasing interactively affect performance by enhancing the quality of the displayed image. The implications of this work for screen presentation of text are presented.*

Received May 1988

1. INTRODUCTION

The recent proliferation of VDUs following the introduction of computer processed and stored text has resulted in an increased awareness in the inherent difficulties involved in displaying text on screen. In all of the possible applications in which users may find themselves reading from VDUs (e.g. word processors, databases, electronic mail) a large number of the characteristics of text and computers can influence their ability to extract and comprehend the required information.

As a result of this the literature on reading from screens has increased dramatically over the last few years and a general consensus seemed to emerge suggesting that reading from screen was not the same as reading from paper. The major findings suggested that screen reading is slower but some investigators reported that it was also less accurate, more fatiguing and rated inferior to high-quality paper print.

However, drawing any firm conclusions from the literature is difficult. Helander *et al.*¹⁴ 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 Rollenhagen³¹ point to the frequently narrow scope of experimental designs. 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 aim of the present review is not 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 parti-

cularly 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.*⁸). This review will concentrate on identifying the nature of any differences that may exist between reading from paper and screens, their possible causes, and under what conditions, if any, they may be resolved.

2. OBSERVED DIFFERENCES

2.1. Speed

By far the most common experimental finding is that reading from screen is significantly slower than reading from paper.^{15, 19, 34, 12, 18} 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 procedure. For example, in the study by Muter *et al.*,¹⁹ subjects read white text on a blue background, with the subject being approximately 2.5 m from the screen. Displayed in 39 characters per line videotex format on a television, the characters 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 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 bigger than the printed characters.

In comparison, Gould and Grischowsky¹² 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 Lickorish³⁴ 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, Switchenko²⁸, Askwall¹ and Cushman⁹ found that reading speed was unaffected by the presentation medium. Askwall attributes this difference in finding 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 × 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 continued to mount supporting the case for a general speed decrement. As Gould *et al.*¹⁰ 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).

2.2. Accuracy

In experimental investigations of reading on screens, accuracy usually 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 significant accuracy differences between VDU and paper,^{34,12} recent experiments by Creed *et al.*⁸ and Wilkinson and Robinshaw³³ report significantly poorer accuracy for proofreading tasks on VDUs.

Since evidence for the effects of presentation media on accuracy invariably emerges from the same investigations which looked at the speed question, the criticisms of procedure and methodology outlined above apply equally here. Furthermore, the measures of accuracy employed also vary. Gould and Grischkowsky¹² 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 Robinshaw³³ argue that such a task hardly equates to true proof reading 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.*⁸ 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. This error classification was developed in response to what they see 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, a result they put down to the 'display font' and the reading strategy adopted by their subjects.

Regardless of the interpretation that is put on the results of any 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. It would seem that for routine spelling checks, reading from VDUs is not less accurate than reading from paper. However, a performance deficit may occur for more visually or cognitively demanding tasks.

2.3. 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).²¹ In the area of screen reading this has manifested itself in speculation of increased visual fatigue and/or eyestrain when reading from VDUs as opposed to paper.

In the Muter *et al.* study,¹⁹ 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 Grischkowsky¹² 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.*²⁷ and Sauter *et al.*²⁴ as supporting evidence.

In a more specific investigation of fatigue, Cushman⁹ 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).²² 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 eliminate flicker completely in the case of positive presentation, a suspected cause of visual fatigue. Wilkinson and Robinshaw³³ 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's¹² 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 proof-reading 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 far measures of fatigue such as subjective ratings of ocular discomfort are comparable 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.

2.4. 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 the degree of comprehension achieved by the reader?

Post-task questions about content of the reading material are perhaps the simplest method of assessing comprehension, although care must be taken to ensure that the questions do not simply demand recall skills. Muter *et al.*¹⁹ 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. Kak¹⁵ 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 Cushman⁹ 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.

Belmore⁵ 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 CRTs—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.

It seems that comprehension of material is not affected by presentation medium. However, a strong qualification of this interpretation of the experimental findings is that suitable comprehension measures for reading material are difficult to devise. 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 that reading from VDUs does not affect comprehension.

2.5. Preference

Part of the folklore of human factors research is that naïve 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.*⁷ report that high-quality typewritten hardcopy is generally judged to be superior. Preference ratings were also recorded in the Muter *et al.*¹⁹ 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.

Starr²⁶ 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. 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. books are portable, cheap, apparently 'natural' in our culture, personal and easy to use. The extent to which such 'common-sense' variables influenced user performance and preferences is not yet well understood. Additionally, when judging the importance of early studies it must be remembered that general screen quality has improved considerably in recent years.

2.6. Summary

Empirical investigations of the area have suggested five possible 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 is 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. With respect to reader preference, top-quality hardcopy seems to be preferred to screen displays, which is not altogether surprising.

It must be noted that the type of task performed in many of these studies represents a very limited subset of what is labelled 'reading'. Proofreading or visual scanning and searching would probably not rank very highly on a detailed list of most frequent reading activities. Browsing, light reading and formal studying are probably more frequent interactions with written material. Creed *et al.*⁸ 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 have yet to be addressed.

3. CAUSES OF DIFFERENCE

While the precise nature and extent of the differences between reading from either presentation medium have not been completely defined, attempts to identify possible causes of any difference have 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.¹⁸

More relevant work has been carried out on possible causes for the observed 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. The aim of this work was to attempt 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 differences between reading from paper and reading from VDUs.

3.1. Orientation

One of the advantages of paper over VDUs is that it can be picked up and oriented to suit the reader. VDUs present the reader with text in a reasonably rigid vertical orientation, though thanks to ergonomic design principles some flexibility to alter orientation is available in many systems. Gould *et al.*¹⁰ investigated the hypothesis that differences in orientation may account for differences in reading performance. Subjects were required to read three articles, one on VDU, one on paper-horizontal and the other on paper-vertical. 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 material³⁰ it does not explain the observed reading differences in the comparisons reported here.

3.2. Eye movements

Mills and Weldon¹⁸ 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. Indeed Tinker²⁹ reports on how certain characteristics affect eye movements and Kolers *et al.*¹⁶ have employed measures of eye movement to investigate the effect of text density on ocular work and reading efficiency. Obviously if reading from VDUs is slower than paper then noticeable effects in eye movement patterns should be found and these may serve to indicate the causes of any differences in reading between paper and screen.

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.*¹⁰ report on 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 refixated more when reading from VDUs. Studying eye movement patterns therefore does not appear to offer any real insight into the causes of observed differences.

3.3. Visual angle

Gould *et al.*¹⁰ hypothesised that due to the usually longer line lengths on VDUs the visual angle subtended by lines on 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 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.

He found significantly greater reading distances for VDUs. In a more controlled follow-up study, Gould and Grischowsky¹³ 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. 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.

3.4. Aspect ratio

The term aspect ratio refers to the relationship of width to height. Typical paper sizes are higher than they are wide, 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 *et al.*¹⁰ 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 aspect ratio.

3.5. Dynamics

Detailed work has been carried out on screen filling style and rates^{6, 16, 25} 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.*¹⁰ had subjects read from paper, VDU and good quality photographs of the VDU material which eliminated any dynamics. Results provided little in the way of firm evidence to support the dynamics hypothesis. Subjects again read consistently faster from paper compared to both other presentation media, which did not differ significantly from each other. Creed *et al.*⁸ 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.

3.6. 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 it is effectively refreshing the screen contents. 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.*² suggested from their experimental data that a refresh rate of 93 Hz is necessary for 99% of subjects to perceive a positive presentation (dark characters on light background) display as flicker free.

This effect of luminance (which is itself not independent of screen size) suggests that the attempt by Creed *et al.*⁸ to control the effect of flicker using photographs was not really valid. Their use of a dark green background with light green characters at a likely* refresh rate of 50 Hz means that there should have been no noticeable flicker in the VDU condition. Gould *et al.*¹⁰ admit that the photographs used in their study were of professional quality but appeared less clear than the actual screen display. Hence, for a variety of reasons it appears that using photographs to control for the effects of flicker may not be a suitable method.

The extent to which flicker may have been an important variable in many studies is unknown as details of screen luminance and refresh rates are often not included in publications. Hence, flicker may play some part in explaining the differences between the two media, particularly in studies where the luminance could be expected to have been high, i.e. those which use dark characters on a light background. However, since some workers³² consider conventionally-driven fluorescent light to be effectively providing intermittent illumination with a frequency of 100 Hz, attempts to compare screen and paper should proceed with caution.

3.7. 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

* Not reported directly by Creed *et al.*⁸ but inferred from the description of equipment used.

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 Radl²³ reported increased performance on a data input task for dark characters, and Bauer and Cavonius³ reported a superiority of dark characters on various measures of typing performance and operator preference.

With regard to reading from screens, Cushman⁹ 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.*¹¹ 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.*¹¹ conclude that to the extent that polarity makes a difference it favours faster reading from dark characters on a light background. Furthermore they cite Tinker³⁰ who reported that polarity interacted with type size and font when reading from paper. The findings of Bauer *et al.*² 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.

3.8. 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. Traditionally 5×7 matrices have been 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.*⁷ recommend a minimum 7×9 matrix. Pastoor *et al.*²⁰ 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 9×13 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.⁴ 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.*¹⁹ compared reading speeds for text displayed with proportional or non-proportional spacing and found no effect. In an experiment intended to appreciate the possible effect of such font characteristics on the performance differences between paper and screen reading, Gould *et al.*¹¹ found no evidence to support the case for proportionally spaced text.

Kolers *et al.*¹⁶ 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 significance. On the other hand Kruk and Muter¹⁷ found that single spacing produced 10.9% slower reading than double spacing. Once more the results appear inconclusive.

Obviously much work needs to be done before a full understanding is achieved of the relative advantages and disadvantages of particular formats and types of display. In a discussion of the role of display fonts in explaining any of the observed differences between screen and paper reading, Gould *et al.*¹¹ state that on the basis of their investigations there is strong evidence 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.

3.9. 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.*¹¹ 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 non-anti-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.

3.10. User characteristics

It has been noted that many of the studies reported in this review employed relatively naïve 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.*¹⁰ Experienced users were described as 'heavy, long-time daily users of IBM 3277 CRT displays'. Inexperienced users had no experience of reading from computers. No significant differences were found between these groups, both reading slower from screen.

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.

3.11. The interaction of display variables: the work of Gould

Despite many of the findings reported thus far, it appears that reading from screens can in certain circumstances be as fast and as accurate as reading from paper. Gould *et al.*¹¹ have empirically demonstrated that under the right conditions the 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 × 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 (60 Hz). 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 section. After a series of experimental manipulations aimed at identifying those variables responsible for the improved performance Gould *et al.*¹¹ 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/CRT display reading rate difference.

Gould *et al.*¹¹ conclude that the explanation of the differences 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 alone such as critical flicker frequency, font or polarity are responsible for any differences.

The Gould *et al.*¹¹ findings are of tremendous importance. They would suggest that the results of the many studies reported earlier can be explained in terms of the quality of screen image presented to the subjects. Muter *et al.*,¹⁹ for example, employed television screens with negative presentation in their investigation. Wilkinson and Robinshaw³³ also used negative presentation and a screen described by themselves as 'of average quality'. In fact none of the studies reporting performance deficits that are cited in this review can claim to have presented screen images of the quality employed in the Gould *et al.*¹¹ studies. If Gould's findings are taken seriously it can be argued that there is no performance deficit for screen reading as such, and any reported differences can be attributed to the quality of screen image employed.

3.12. Conclusions and future directions

Although reading from computer screens may be slower and sometimes 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 (except the cost of high quality) nor the reader are causal factors. Reading from screens can be as fast and as accurate as reading from paper. Invariably it is the quality of the image presented to the reader which is crucial. Tinker³⁰ reports dramatic interaction effects of image quality variables on paper and according to Gould *et al.*¹¹ it is likely that these occur on screen too. Positive presentation combined with a high screen resolution and refresh rate to avoid flicker can produce good images, and with the addition of anti-aliased characters in an appropriate font 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 of poorer quality than those used by Gould and his associates to overcome the performance deficit. In this sense it is true to say that reading from screen is not the same as reading from paper. Until screen standards are raised sufficiently these differences are likely to remain. Furthermore the studies by Gould reported here were concerned only with speed and accuracy. 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 differences such as fatigue and reader preference 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.

Although Gould's work answered many of the questions regarding use of screens, there are two areas which we feel have yet to be tackled. The first of these concerns the task and the second concerns the display.

With regard to the task, it was noted above that practically all studies have used proofreading. However, this is perhaps not as relevant a task for most users of text as reading, say, journal articles or even books. The reading of such documents raises questions relating not only to comprehension but also to issues such as manipulation and navigation.

Navigation issues raise further questions about the presentation of contextual information, and it is clear that more of this information can be visible on a large screen than on a small screen. Hence, work is required on

the effects of screen size on the comprehension of extended texts. It is in these two complementary directions that future research would seem to be required.

Acknowledgement

This work was funded by the British Library Research and Development Department as part of Project Quartet. We are grateful to Professor Brian Shackel for commenting on an earlier draft of this paper.

REFERENCES

1. S. Askwall, Computer supported reading vs reading text on paper: a comparison of two reading situations. *International Journal of Man-Machine Studies* **22**, 425-439 (1985).
2. D. Bauer, M. Bonacker and C. R. Cavonius, Frame repetition rate for flicker-free viewing of bright VDU screens. *Displays*, January, 31-33 (1983).
3. D. Bauer and C. R. Cavonius, Improving the legibility of visual display units through contrast reversal. In *Ergonomic Aspects of Visual Display Terminals*, edited E. Grandjean and E. Vigliani. London: Taylor and Francis (1980).
4. I. P. Beldie, S. Pastoor and E. Schwarz, Fixed versus variable letter width for televised text. *Human Factors* **25** (3), 273-277 (1983).
5. S. Belmore, Reading computer presented text. *Bulletin of the Psychonomic Society* **23** (1), 12-14 (1985).
6. N. Bevan, Is there an optimum speed for presenting text on VDUs? *International Journal of Man-Machine Studies* **14**, 59-76 (1981).
7. A. Cakir, D. J. Hart and T. F. M. Stewart, *Visual Display Terminals*. Chichester: John Wiley and Sons (1980).
8. A. Creed, I. Dennis and S. Newstead, Proof-reading on VDUs. *Behaviour and Information Technology* **6** (1), 3-13 (1987).
9. W. H. Cushman, Reading from microfiche, VDT and the printed page: subjective fatigue and performance. *Human Factors* **28** (1), 63-73 (1986).
10. J. D. Gould, L. Alfaro, V. Barnes, R. Finn, N. Grischkowsky and A. Minuto, Reading is slower from CRT displays than from paper: attempts to isolate a single-variable explanation. *Human Factors* **29** (3), 269-299 (1987).
11. J. D. Gould, L. Alfaro, R. Finn, B. Haupt and A. Minuto, Reading from CRT displays can be as fast as reading from paper. *Human Factors* **29** (5), 497-517 (1987).
12. J. D. Gould and N. Grischkowsky, Doing the same work with hard copy and cathode-ray tube (CRT) computer terminals. *Human Factors* **26** (3), 323-337 (1984).
13. J. D. Gould and N. Grischkowsky, Does visual angle of a line of characters affect reading speed? *Human Factors* **28** (2), 165-173 (1986).
14. M. G. Helander, P. A. Billingsley and J. M. Schurick, An evaluation of human factors research on visual display terminals in the workplace. *Human Factors Review*, Chapter 3, 55-129 (1984).
15. A. V. Kak, Relationships between readability of printed and CRT-displayed text. *Proceedings of Human Factors Society - 25th Annual Meeting*, 137-140 (1981).
16. P. A. Kolars, R. L. Duchnick and D. C. Ferguson, Eye movement measurement of readability of CRT displays. *Human Factors* **23** (5), 517-527 (1981).
17. R. S. Kruk and P. Muter, Reading continuous text on video screens. *Human Factors* **26** (3), 339-345 (1984).
18. C. B. Mills and L. J. Weldon, *Reading Text from Computer Screens*. Centre for Automation Research, Human-Computer Interaction Laboratory, University of Maryland, Maryland 20742 (1985).
19. P. Muter, S. A. Latrémouille, W. C. Treurniet and P. Beam, Extended reading of continuous text on television screens. *Human Factors* **24** (5), 501-508 (1982).
20. A. Pastoor, E. Schwarz and I. P. Beldie, The relative suitability of four dot-matrix sizes for text presentation on colour television screens. *Human Factors* **25** (3), 265-272 (1983).
21. B. G. Pearce (ed.), *Health Hazards of VDUs?* Chichester: John Wiley and Sons (1984).
22. R. G. Pearson and G. E. Byars, The development and validation of a checklist for measuring subjective fatigue. Report TR-56-115, San Antonio, TX: USAF School of Aviation Medicine (1956).
23. G. W. Radl, Experimental investigations for optimal presentation mode and colours of symbols on the CRT screen. In *Ergonomic Aspects of Visual Display Terminals*, edited E. Grandjean and E. Vigliani. London: Taylor and Francis (1980).
24. S. L. Sauter, M. S. Gottlieb, K. M. Rohrer and V. N. Dodson, The well-being of video display terminal users: an exploratory study. Report 210-79-0034. U.S. Department of Health and Human Services (1983).
25. E. Schwarz, I. P. Beldie and S. Pastoor, A comparison of paging and scrolling for changing screen contents by inexperienced users. *Human Factors* **25** (3), 279-282 (1983).
26. S. J. Starr, Effects of video display terminals in a business office. *Human Factors* **26**, 347-356 (1984).
27. S. J. Starr, C. R. Thompson and S. J. Shute, Effects of video display terminals on telephone operators. *Human Factors* **24**, 699-713 (1982).
28. D. M. Switchenko, Reading from CRT versus paper: the CRT disadvantage hypothesis re-examined. *Proceedings of Human Factors Society, 28th Annual Meeting*, 429-430 (1984).
29. M. A. Tinker, Recent studies of eye movements in reading. *Psychological Bulletin* **55**, 215-231 (1958).
30. M. A. Tinker, *Legibility of Print*. Ames, Iowa: Iowa State University Press (1963).
31. Y. Waern and C. Rollenhagen, Reading text from visual display units (VDUs). *International Journal of Man-Machine Studies* **18**, 441-465 (1983).
32. A. Wilkins, Intermittent illumination from visual display units and fluorescent lighting affects movements of the eyes across text. *Human Factors* **28** (1), 75-81 (1986).
33. R. T. Wilkinson and H. M. Robinshaw, Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. *Behaviour and Information Technology* **6** (2), 125-133 (1987).
34. P. Wright and A. Lickorish, Proof-reading texts on screen and paper. *Behaviour and Information Technology* **2** (3), 227-235 (1983).