Available online at www.sciencedirect.com
SCIENCE DIRECT®
Int. J. Human-Computer Studies 64 (2006) 395–404
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

The influence of font type and line length on visual search and information retrieval in web pages

Jonathan Ling*, Paul van Schaik

School of Psychology, Keele University, Keele, Staffordshire ST5 5BG, UK

Received 14 April 2005; received in revised form 22 August 2005; accepted 28 August 2005
Available online 11 October 2005
Communicated by A. Cockburn

Abstract

Most web sites are heavily text-based. Previous research has indicated that the way in which this text is presented may have a significant impact on usability. This paper reports findings from two experiments that explored the influence of font type and line length on a range of performance and subjective measures. Experiment 1 used a visual search task and Experiment 2 examined information retrieval. Overall, there was little impact of font on task performance, although the effect of line length was significant, with longer line lengths facilitating better scanning (Experiment 1) and shorter line lengths leading to better subjective outcomes (Experiments 1 and 2). Implications of these results for the design of web pages are discussed and recommendations given.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Font; Line length; Web; Visual search; Information retrieval

1. Introduction

Most web sites rely on text to present their content. While there has been recognition of the importance of how this text is presented (see e.g. Nielsen, 2002), designers still need to make their web sites more usable (Trewin et al., 2004). It is imperative that web sites are constructed to enable a high level of usability for all users (Shneiderman, 2000), particularly because many users of the Internet may have a degree of visual impairment or dyslexia, or be lower-literacy users (Nielsen, 2005), all of whom will need clear, uncluttered pages. Consideration of human factors should be an important part of the design process; poorly designed layouts can quickly lead to fatigue, with a resultant lowering of speed and accuracy of task performance (Streveler and Wasserman, 1984). Aesthetic considerations are also important for usability (Lindgaard and Dudek, 2003; van Schaik and Ling, 2003) as these may be closely linked to individuals' motivation and satisfaction (Moneta and Csikszentmihalyi, 1996).

Typography offers many alternatives for presenting online text which are likely to impact on the usability of web pages. There are many ways of presenting letters and words, using different combinations of font types, weights, slants, sizes, line lengths, and justification. What web designers need to know is the impact of their design decisions about text presentation upon usability. This is important regardless of whether usability is measured by performance or preference, although the latter may have a significant impact on the use of commercial sites.

Several authors have argued against presenting large amounts of text on screen as users will not read it (e.g. Morkes and Nielsen, 1997). This view has been enshrined in a range of web-authoring guidelines that encourage web designers to use text sparingly or to break it up into small chunks to avoid scrolling (Bradley, 2002; Briem, 2002), this is despite the fact that navigating between pages may take more time (Ingraham and Bradburn, 2003).

Although not based on empirical work, List (2001) argues that the best way to improve readability is to avoid distracting readers in any way, 'so that the text comes through without interference'. Both readers and content developers can, List suggests, employ strategies to enhance reading from a computer screen, such as through adjusting

*Corresponding author. Tel.: +44 0 1782 583736; fax: +44 0 1782 583387.
E-mail address: j.ling@tees.ac.uk (J. Ling).

1071-5819/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2005.08.015

the environment or the browser software or improving web page design. There is general agreement (Ingraham and Bradburn, 2003) that there are three basic elements that must be considered when developing a highly readable web page: typeface, spacing (whether down or across the page), and colour. These have all been the subject of research, however this paper focuses on the first two of these attributes, which have not previously been investigated empirically in web pages.

Usability research has focused on various design parameters, including font type (Davidov, 2002) and line length (Dyson and Haselgrove, 2001). However, a problem with many fonts is that they are designed to be printed and so may be less easy to read when presented on screen. Unfortunately, the results of investigations into font preference have often presented web site designers with difficult choices. For example, Bernard et al. (2003) compared serif and sans serif fonts in 12- or 14-point size in a task where participants had to detect substituted words in text. They found that 14-point fonts were more legible, led to faster reading, and were preferred to the 12-point fonts. However, they also found that even though participants performed tasks more quickly with serif fonts, they still preferred sans serif fonts. Any such tension between aesthetics and performance can cause problems for designers of web pages (Schmidt et al., 2003): should designers choose a font that users like but that interferes with the speed at which they can locate information on a site and perhaps cause them to 'bail out' of a site? Or should they choose a font that risks alienating potential users right from the start?

Information retrieval is a key activity in using the web. As this activity invariably involves visual search, research on the processes involved in visual search is essential. Two basic stages are involved in visual search (Scott, 1993; Wolfe, 1994). Initially all screen locations are processed in parallel, but only a limited amount of information is extracted. Subsequently, a second stage restricts itself to areas of the visual field that are of interest. At this stage more complex tasks can be carried out and more information extracted. However, this stage has limited capacity and information can only be extracted from one or, perhaps, a few spatial locations at a time. Visual conspicuity is one factor that influences the selection process. This conspicuity can be defined as a combination of an object's properties, relative to its background, which attracts attention and is therefore attended to. Conspicuity can be based on colour, shape, size, orientation and other properties of the display. If a target object is made visually conspicuous, search time can be up to 83% faster (Nygren, 1996).

Although several researchers have suggested that screens should be structured in an organised manner to improve visual search and therefore usability (Graf and Krueger, 1989; Tullis, 1997) the effect of line length in web page design has received little empirical focus. Investigations of the effect of line length in both printed and on-screen formats have generated differing results. For example for printed words, it is generally accepted that line lengths should not exceed 70 characters per line (cpl; Spencer, 1968). The situation is unfortunately less clear for on-screen presentations. For example, in a comprehension task, Duchnicky and Kolers (1983) found that blocks of text of 75–100 cpl were read more quickly than when the text was presented in shorter lines. This finding contrasts with the more complex comprehension task used by Dyson and Haselgrove (2001) who found that varying the amount of text presented between 55 and 100 cpl had no effect on reading speed. However, both of these tasks were based on reading blocks of text presented on-screen and not in the context of a realistic web browser window.

Davidov (2002) suggests web designers should use 60–65 c pl, rather than the average line lengths of 80 or 100 characters seen in books, with a lower limit of 40 characters. Disappointingly for those reading online, List (2001) notes that most browsers are set to 100 characters, far in excess of the width that can comfortably be taken in by the eye.

According to Briem (2002), line lengths should be kept short for two reasons. First, unrestrained (i.e. undefined in HTML) lines will turn paragraphs into single lines when presented on large monitors. Second, shorter line lengths are easier to scan than longer ones. Briem recommends that a typical 520 pixels wide screen should be divided into 120 pixels for the navigation area, and 390 pixels for the content area, with a 10 pixel gutter separating the two areas. When using 10 or 12 point fonts, this would equate to line lengths of 12 and 39 or 10 and 32 characters for navigation (on the left) and content (on the right) areas respectively, much narrower than that recommended by either Davidov (2002) or Dyson and Haselgrove (2001). The suggestions for the optimal number of cpl in web pages made by Davidov (2002) and Briem, although plausible, were not supported by empirical research. Dyson and Haselgrove's recommendations although based on empirical work, did not use realistic web browsers and used a reading rather than a visual search or information retrieval task. Therefore there is a need for research to explore these parameters fully, within the context of a realistic web environment.

The research reported in this paper investigated the effects of font type and line length on online behaviour and preferences as a means of generating clear design guidelines for the production of web pages. Two experiments were conducted, the first using a visual search task and the second using an information retrieval task.

2. Experiment 1

2.1. Method

2.1.1. Experimental Design

The experiment used a two-factor mixed measures design. The within-subjects factor was line length with

four levels: 55, 70, 85 and 100 cpl. The between-subjects factor was the font used to display text on the web pages. This had two levels: Arial 10 point and Times New Roman 12 point. Dependent variables were accuracy and speed of visual search as well as subjective measures.

2.1.2. Participants

There were 72 participants, 57 of whom were female. Of these, 39 took part in the Arial condition, and 33 took part in the Times condition. Sample sizes were unequal because participants were an opportunity sample who completed the tasks during lab classes. A majority of participants (61%) were 25 or under; those remaining were aged between 26 and 50. All participants used the web and 94% had been doing so for more than a year. Frequency of using the web varied from more than once a day to less than once a month, with a majority (76%) using the web at least once a day. All participants had normal or corrected-to-normal visual acuity.

2.1.3. Materials and apparatus

Mock web pages were created in HTML and displayed using Microsoft Internet Explorer web browser version 6.0. The pages were captured and saved as bitmap files. The web pages consisted of two frames, a larger frame (85%, the content area) and a smaller frame (15%, the navigation area) containing five hypertext links, all of which were matched for length in each of the trials (Fig. 1). The navigation area always appeared on the left of the screen. In the content area, links appeared in the text in five locations, distributed evenly down the web page from top to bottom; hyperlinks of the same length were used on each of the screens. Location of the target link (when present) was counterbalanced across the trials, so links appeared in all the possible locations an equal number of times. Hypertext links were presented in bold and underlined, and were always coloured blue (colour code #0000FF). The rest of the text of the web pages was always black (colour code #000000). The background in both areas was white (colour code #FFFFFF). The font used for both normal text and hyperlinks was either all Arial 10 point or all Times 12 point, depending on the experimental condition; the fonts were different point sizes in order to make them as close in physical size to each other as possible so that the same amount of text appeared on pages regardless of the font used. Text presented in the content area appeared in one of four line lengths: 55, 70, 85 or 100 cpl. All text was left aligned, and both text and hyperlinks were the same in all conditions. The text used for the content area of the web pages was taken from a source, the doctoral thesis of one of the authors, with which all participants were unfamiliar. Each web page was created in two versions with identical content: one with Arial 10 point and another with Times 12 point.

In the experiment, the images were presented using SuperLab experiment generator software (Cedrus Corporation, 2002) which ran on personal computers (Intel Pentium, 333 MHz, 64 Mb RAM, Microsoft NT4 operating system, 14 inch monitors). The screen dimensions were 800×600 pixels. In order to ensure maximum clarity, contrast was set to maximum level and brightness to minimum.

Four items from Tractinsky et al.'s (2000) aesthetics scale (previously used by van Schaik & Ling, 2003, for

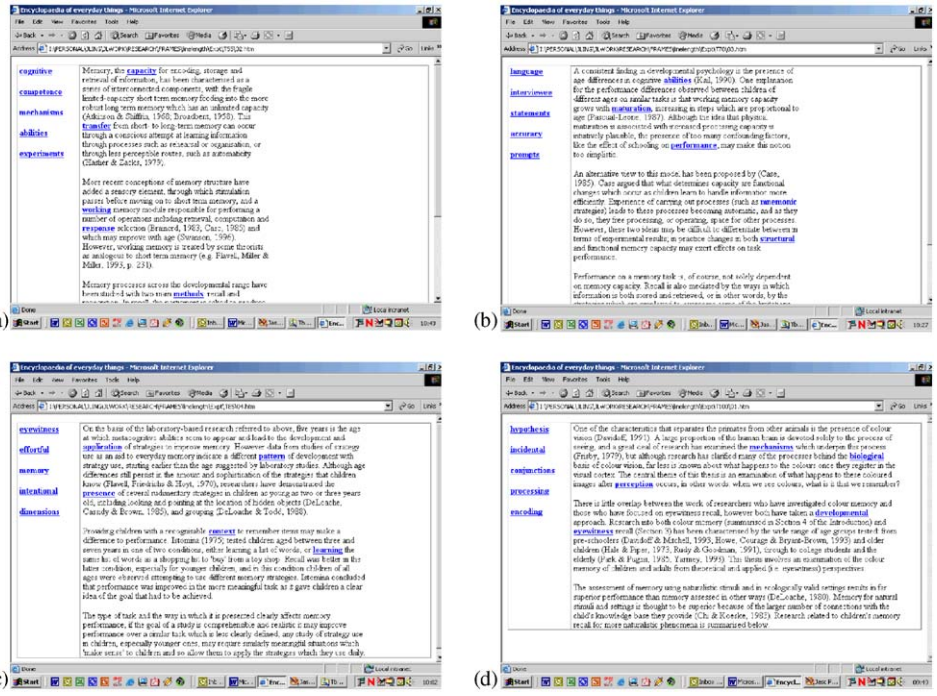

Fig. 1. Typical web pages used in Experiment 1. (a) 55 cpl, Times Roman, (b) 70 cpl, Times Roman, (c) 85 cpl, Times Roman, (d) 100 cpl, Times Roman.

evaluating the quality of web pages) were used to rate the eight types of web page used in the visual search task; these were combinations of 4 line lengths and 2 font types. The items used 7-point Likert scales, for example, 'I judge this web page to be:' with scale end points 'very ordered' and 'very disordered'.

2.1.4. Procedure

The experiment was carried out in classrooms with 15 to 20 participants taking part concurrently. They first completed a practice series consisting of eight trials, followed by a series of 72 experimental trials presented in random order.

Participants were first presented with a set of on-screen instructions. They were told that they were going to perform in a visual search task involving mock web pages and that they had to try to find a hyperlink in a screen of text. If the link was present on the page, they had to press the 'P' key, and if it was absent they had to press 'A'.[1]

After reading the instructions, participants pressed the 'S' key on the keyboard to proceed to a series of experimental trials. In each trial a blank white screen (shown randomly for 1000, 1500 or 2000 ms with a mean of 1500 ms) was presented, then a black target word on a white background (always 1000 ms, in 48-point Arial or Times, depending on the version of the experiment, in the centre of the screen), followed by a blank white screen (shown randomly for 1000, 1500 or 2000 ms with a mean of 1500 ms) and then a web page. The target word was either present in or absent from the content area.

Participants responded by pressing the appropriate keys to indicate whether the target was present or absent from the web page. A participant's response triggered the start of the next trial. If a participant had not responded after 5000 ms then the next trial started automatically. In the practice trials, all four line-lengths appeared twice and in half the trials the target was absent. When all participants had completed the practice task, any questions were answered by the experimenter before they went on to the main experiment. Participants were instructed to perform the task as quickly and as accurately as possible. As the experiment attempted to replicate the conditions under which participants normally use computers, no restrictions were placed on the distance they were seated away from the screen and they were free to move toward and away from the computer screens.

After completion of the final experimental trial, participants completed a series of questions presented by computer. The questions covered demographic details (age, sex, use of the web), aesthetic value of pages and preference for line length and font type. Participants completed the aesthetics scale for each type of page (32 ratings−8 screens × 4 items). For preference, using the procedure employed by Ling and van Schaik (2002), participants were presented with all 15 possible paired combinations of line lengths, and asked to choose which they preferred. This procedure was repeated for font type. Questions were presented in the same order for all participants. Participants took approximately 40 Mins to complete the experiment.

[1]Although touch typists may find 'A' slightly easier to press than 'P' on a keyboard, in practice participants rested their index fingers on each of these keys.

2.2. Results

For the purpose of analysis, the task performance data were divided into hits and correct rejections, which were each examined in terms of accuracy and speed. In addition, participants' preference for and judgement of aesthetic value of web pages was assessed. A series of $2 \times (4)$ analyses of variance (ANOVA) were used to assess the effects of font type (Arial or Times) and line length (55, 70, 85 or 100 cpl) on outcome measures; where appropriate, post hoc analyses were conducted to test for specific differences between line lengths.

2.2.1. Analysis of hits

2.2.1.1. Accuracy. The percentage of hits as a proportion of all responses made was calculated as a measure of accuracy. A hit was defined as a correct response to a target word that was present in the content frame of a subsequently presented web page. None of the main effects of line length or font or the interaction was statistically significant.

2.2.1.2. Speed. Reaction times for hits were examined. The effect of line length was significant, $F(3, 210) = 13.10$, $\eta^2 = 0.038$, $p < 0.001$ (see Table 1, row 1); however, the effect of font type and the interaction were not. Post hoc pairwise t-tests with Bonferroni correction showed significant differences of 55 and 70 with 85 cpl (both $p < 0.001$) and of 55 and 70 with 100 cpl (both $p < 0.05$). Task performance was fastest with 85 and 100 cpl.

2.2.2. Analysis of correct rejections

2.2.2.1. Accuracy. The percentage of correct rejections was calculated. A correct rejection was defined as a correct response to a target word that was absent in the navigation or content frame of a subsequently presented web page. The effect of line length was significant, $F(3, 210) = 2.77$, $\eta^2 = 0.004$, $p < 0.05$ (see Table 1, row 2). The effects of font type and the interaction were not significant. Post hoc pairwise t-tests with Bonferroni correction showed a significant difference of 70 with 85 cpl ($p < 0.05$) and that task performance was more accurate with 70 cpl.

2.2.2.2. Speed. Reaction times for correct rejections were examined. Neither of the main effects of line length and font type or the interaction were significant.

Table 1
Mean scores by line length

		55 cpl		70 cpl		85 cpl		100 cpl	
		M	(S.D.)	M	(S.D.)	M	(S.D.)	M	(S.D.)
Expt. 1									
1. Reaction times for hits (ms)	Arial	1786	(339)	1805	(392)	1639	(363)	1699	(412)
	Times	1752	(534)	1718	(387)	1511	(311)	1609	(451)
2. % of targets correctly rejected	Arial	86.04	(22.47)	88.89	(22.37)	84.05	(25.84)	85.75	(24.31)
	Times	81.14	(25.98)	83.84	(28.07)	80.81	(29.69)	79.80	(29.72)
3. Pref. for line length (% of times preferred)[a]	Arial	25.64	(21.58)	23.93	(13.68)	19.23	(14.07)	14.10	(16.46)
	Times	22.73	(19.90)	29.29	(12.52)	23.74	(11.81)	20.20	(17.06)
4. Aesthetic value (rated on 7-point Likert scale)	Arial	4.22	(1.81)	4.81	(1.26)	4.62	(1.42)	4.71	(1.71)
	Times	4.07	(1.48)	4.35	(1.45)	4.63	(1.63)	4.34	(1.72)
Expt. 2									
5. Pref. for line length (% of times preferred)	Arial[b]	30.45	(19.73)	30.77	(13.36)	22.12	(13.09)	16.67	(18.38)
	Times[c]	40.07	(17.26)	29.08	(11.24)	20.21	(11.49)	10.64	(14.51)

[a]Percentages add up close to 100 because of missing data.
[b]Arial presented as font type for rating of aesthetic value.
[c]Times presented as font type for rating of aesthetic value.

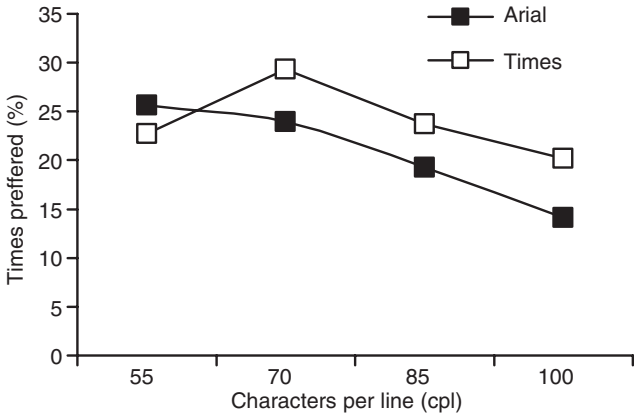

Fig. 2. Preference for line length by font.

2.2.3. Analysis of subjective measures

2.2.3.1. Preference for line length. The percentage of times participants chose a particular line length over the others was calculated as a measure of preference. The effect of line length was significant, $F(3, 210) = 3.95$, $\eta^2 = 0.044$, $p < 0.05$ (see Table 1, row 3 and Fig. 2) as was the effect of font type $F(1, 70) = 4.60$, $\eta^2 = 0.010$, $p < 0.05$. The interaction was not significant. Fig. 2 shows a negative linear relationship of line length with preference for Arial with, shorter line lengths preferred. Post hoc pairwise t-tests with Bonferroni correction showed significant a mean difference of 70 cpl with 100 cpl ($p < 0.05$).

2.2.3.2. Preference for font type. The number of times participants who chose Arial and Times when presented with these two font types was recorded. Of those using Arial in the visual search task, 74% chose Arial. Participants using Times in the visual search task chose Arial in 72% of cases. A chi square test of independence of font type used during visual search and preference for font demonstrated independence, $\chi^2\ (1) = 0.029$, $p > 0.05$. Over both font types during visual search Arial was preferred over Times, $\chi^2\ (1) = 14.63$, $w = 0.45$, $p < 0.001$.

2.2.3.3. Aesthetic value. The aesthetics scale was checked for reliability. The instrument was found to be reliable for only half of the eight line length-font type combinations (Cronbach's alpha > 0.70). Therefore, in subsequent analyses a single item (the 'overall' item: 'I judge the web page to be' with anchors 'very bad' and 'very good') was used as a measure of aesthetic value, as this item had the highest correlations with other items.

The effect of presented font type was significant, $F(1, 67) = 5.63$, $\eta^2 = 0.007$, $p < 0.05$, with Arial rated more highly than Times (see Table 1, row 4). There were no other significant effects.

2.2.4. Summary of results

Experiment 1 showed that there was no effect of font type on speed or accuracy for either hits or correct rejections in a visual search task. Longer line lengths (85–100 cpl) led to quicker searches for hits but reduced accuracy (85 cpl) for correct rejections. In terms of subjective measures, participants preferred shorter line lengths over longer ones and Arial over Times. In addition, Arial was also rated by participants as having a higher level of aesthetic value.

3. Experiment 2: Introduction

Visual search is only one type of user interaction with a web page; other modes of usage may exert different

demands on the user and may be influenced in different ways by variables like line length and font. Experiment 2 sought to examine the impact of these variables on users who were performing an information retrieval task. Information retrieval is a common use of the Internet (Brooks, 2003), and is a goal-oriented task in which the user aims to locate a specific piece of information from within a web site, such as product prices or contact details.

3.1. Method

3.1.1. Experimental design

The same experimental design was employed as in Experiment 1, although with this time using an information retrieval task rather than a visual search task and with speed and efficiency of task/performance as well as subjective measures as dependent variables.

3.1.2. Participants

There were 99 participants, consisting of 82 females and 17 males, with a mean age of 24 (S.D. = 7.1). Of these, 52 performed the information retrieval task with Arial, and 47 with Times. Unequal sample sizes were due to differences in the number of participants attending lab classes in which the data was collected. All participants used the web and 93% of them had done so for more than one year. Frequency of using the web varied from more than once a day to less than once a month, with a majority (77%) using the web at least once a day. None of the participants had taken part in Experiment 1.

3.1.3. Materials and apparatus

A set of six web sites was produced. All sites consisted of a hierarchy of three levels. Each site was produced in eight versions, corresponding to four line lengths (55, 70, 85, 100 cpl) combined with two font types (Arial, Times). The first site consisted of 16 pages and was used for the practice experiment and represented the domain of gardening. The other five sites, each consisting of 30 pages, were used for the main experiment and represented five domains: sport, shopping, music, software and computer equipment. The total amount of text in the content area on any given page was kept constant and pages did not scroll (i.e. all text in the content area was visible all the time), regardless of the line length used.

The experiment ran on personal computers (Intel Pentium, 333 MHz, 64 Mb RAM, Microsoft NT4 operating system, 14 in monitors). The screen dimensions were 800×600 pixels; contrast and brightness were set to optimum levels. Experimental software, written in Microsoft Visual Basic version 6, used the web browser control (based on version 6 of the Microsoft Internet Explorer web browser) to display pages, which enabled participants to navigate through the sites. The program recorded all participants' computer interaction, including the sequence of web pages visited for each task (including time spent on each page) and all answers they gave, as well as the time taken to answer each question. Participants were asked a series of 40 questions based on information available on the web sites which they had to answer by browsing the site for the correct information. Answers to the questions were one (for 20 questions) or two (for the other 20 questions) links away from the home page. Each line length was presented 10 times.

In a rating task, participants completed the same four items from Tractinsky et al.'s (2000) aesthetics scale as in Experiment 1 for each of the four line lengths (producing 16 ratings in total), using the same font they had been presented with during the information retrieval task. These questions asked participants to rate aesthetic value of the pages in terms of order, meaningfulness, comprehension and overall impression. The scale was found to be reliable, Cronbach's alpha for 55 cpl was 0.82; for 70, 0.76; for 85, 0.81 and for 100 cpl, 0.78. Scores on each of the items were averaged to give an overall rating.

3.1.4. Procedure

Participants were presented with instructions before completing a practice task, consisting of a series of five information retrieval questions that were presented in random order using the gardening site. They were told that a series of questions would appear at the top of the screen. After reading each question (see Fig. 3 for examples of questions) they had to click on a button labelled 'Show Web site'. The home page of the first site was then displayed in the browser window (see Fig. 3); the question remained displayed along the top of the screen. Participants were instructed to find the answer to each question using the site. Once they found the answer they had to click a button labelled 'Your answer'. A dialog box then appeared into which participants entered their answer. Once they had typed their answer, the next question appeared. Participants were instructed to take the most direct route to the answer possible.

When all participants had completed the practice task, any questions were answered before they went on to the main experiment. The five other sites were used and participants completed a series of 40 further randomised questions (5 sites $\times$ 8 questions). Questions and answers were similar in length and complexity.

After the experimental trials, participants completed the aesthetics items for each line length and gave their preferences for line length and font type, using the same procedure as in Experiment 1. Participants took approximately 40 minutes to complete the experiment.

3.2. Results

For the purpose of analysis, task performance on trials in which a correct answer was given was examined in terms of speed and efficiency.[2] Subjective measures included

[2]Percentages of correct answers were 96–98% for the different combinations of line lengths and font type.

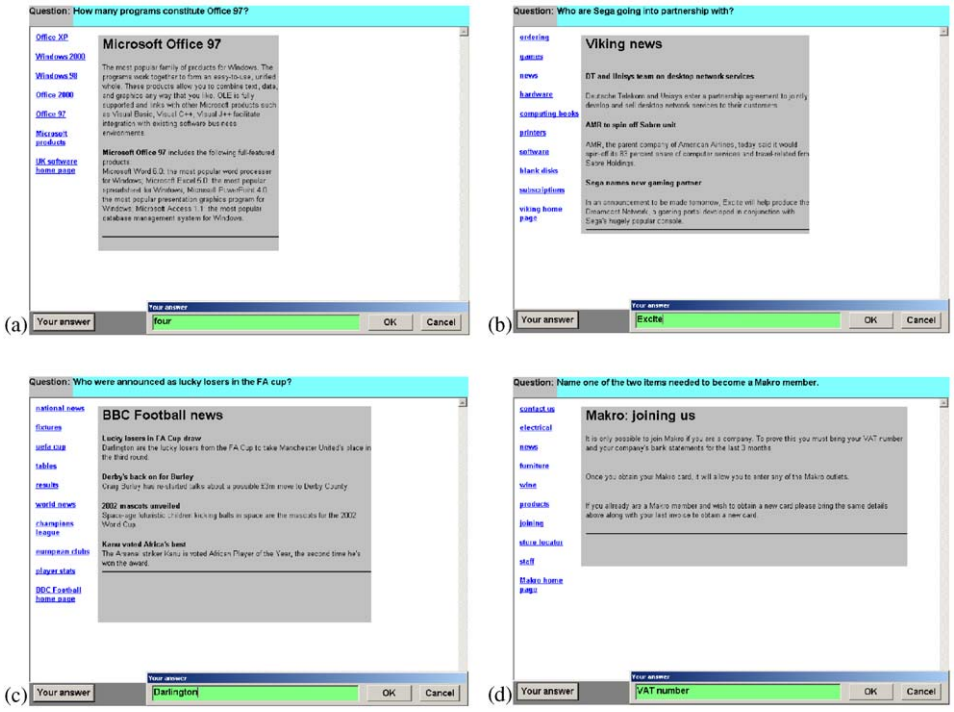

Fig. 3. Typical web pages used in Experiment 2. (a) 55 cpl, Arial, (b) 70 cpl, Arial, (c) 85 cpl, Arial, (d) 100 cpl, Arial.

preference for and aesthetic value of web pages. A series of $2 \times (4)$ ANOVA were conducted to assess the effects of line length and font type on outcome measures; where appropriate, post hoc tests were conducted to identify specific differences between line lengths.

3.2.1. Task performance

3.2.1.1. Speed. Average time-on-task for correct answers was calculated. None of the main effects of line length and font type, or the interaction was significant.

3.2.1.2. Efficiency. The number of pages visited before a correct answer was given was calculated. The main effects of line length and font type, and the interaction were not significant.

3.2.2. Subjective measures

3.2.2.1. Preference for line length. The percentage of times participants chose a particular line length over the others was calculated as a measure of preference. The effect of line length was significant, $F(3, 291) = 29.00$, $\eta^2 = 0.224$, $p < 0.001$, as was the interaction between line length with font type, $F(3, 291) = 3.60$, $\eta^2 = 0.028$, $p < 0.005$ (see Table 1, row 5 and Fig. 4). The effect of font type was not significant. Simple effect tests showed that the effect of line length was significant with both Arial, $F(3, 153) = 6.80$, $\eta^2 = 0.118$, $p < 0.001$, and Times, $F(3, 138) = 28.99$, $\eta^2 = 0.387$, $p < 0.001$. Fig. 4 shows a negative linear relationship of line length with preference for both font types, indicating that longer line lengths were preferred less often. Simple comparison pairwise t-tests with Bonferroni correction showed significant mean differences of 100 with 55 cpl ($p = 0.05$) and 70 cpl ($p < 0.005$) for Arial. Mean differences were significant between all pairs for Times (all $p < 0.05$ or less). Further simple effect pairwise t tests with Bonferroni correction showed that the effect of font type was significant for 55 cpl, $t(97) = 2.571$, $p < 0.05$, but not for other line lengths.

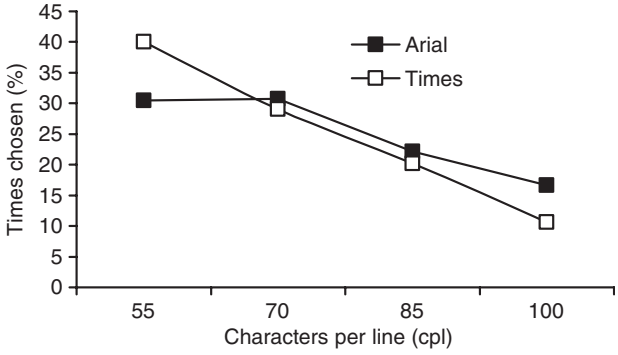

Fig. 4. Interaction between choice of font and characters per line.

3.2.2.2. Preference for font type. The number of times participants who chose Arial and Times when presented with these two font types was recorded. Of those participants using Arial in the information retrieval task 58% chose Arial, while 60% of participants using Times preferred Arial. A χ^2 test of independence of font type used during information retrieval and preference for font demonstrated independence, $\chi^2(1) = 0.036$, $p > 0.05$. This preference for Arial over Times in the information retrieval

Table 2
Preference for font type (percentage of times preferred) (Experiment 2)

Font in presented visual search	Font presented in preference task					
	Arial		Times Roman		Trebuchet	
	M	(S.D.)	M	(S.D.)	M	(S.D.)
Arial	39.10	(26.17)	25.00	(23.69)	35.90	(21.74)
Times	36.88	(21.12)	31.91	(29.45)	31.21	(26.38)

tasks showed a trend towards significance though did not reach it, $\chi^2(1) = 2.92$, $w = 0.17$, $0.05 < p < 0.10$ (see Table 2).

3.2.2.3. Aesthetic value. The main effects of line length and font type, and the interaction were not significant.

4. Discussion

In two experiments, we examined the effect of font type and line length on visual search and information retrieval in web pages.

4.1. Summary of results

The effect of line length on task performance measures that were found with the visual search task (Experiment 1)—with relatively small and very small effect sizes (Cohen, 1988)—disappeared with the information retrieval task (Experiment 2). Over the two experiments, the results relating to subjective measures show a preference for shorter line lengths over longer (with a very large effect size in Experiment 2 and a small to medium one in Experiment 1) and a preference for Arial over Times (with close to a large effect size in Experiment 1 and a small to medium effect size in Experiment 2).

These results have implications for the way in which designers display web content. We found a significant difference in performance that was related to the type of task participants performed. When the task was visual search, participants performed faster with longer line lengths. In comparison to shorter line lengths, longer line lengths allow users to scan quickly across the page, while reducing the number of separate lines that need to be scanned for a given amount of information (although it is likely that extremely long lines will cause difficulty in returning accurately to the start of the next line; see Dyson and Haselgrove, 2001). In addition, although scrolling was not necessary in the visual search task (Experiment 1), this would obviously be reduced for longer line lengths, hence increasing speed. With an information retrieval task (Experiment 2), however, the superiority of longer line lengths disappeared. In fact, users performed better with a line length of 70 cpl in Experiment 1, when presented with web pages that did not contain the sought after information (i.e. correct rejections). This indicates that although longer line lengths facilitate faster scanning, shorter lines may be better for accuracy. Indeed, there was a preference for shorter lines in both experiments, with a very large effect size in Experiment 2 and a small to medium effect size in Experiment 1.

Perhaps surprisingly, given the beliefs of some commentators, font type had no effect on either visual search or information retrieval. Previous commentators have argued that particular fonts, usually sans serif, lead to better levels of performance than other, usually serif, fonts (Davidov, 2002). This view has been challenged by Bernard et al. (2003) who found no differences in performance across a range of widely used fonts. Our results support their conclusion. However, the effect of font type may depend on type of task. In tasks that are not dominated by reading and comprehension of text, such as the visual search and information retrieval tasks used in the current study, font type may not significantly affect task performance, but this may be different in other tasks, for example online reading of academic text (Dyson and Haselgrove, 2001).

The finding that longer line lengths facilitate faster scanning supports earlier findings on comprehension by Dyson (Dyson and Haselgrove, 2001; Dyson and Kipping, 1997). This was an important confirmation of Dyson's work because in her studies participants read blocks of text presented in a window on the left of the screen with the right side of the screen light grey. In the current experiments, information was displayed as web pages and therefore was a better representation of what a typical web user encounters when going online. This is an important finding; nevertheless further research needs to establish whether Dyson's findings on comprehension can also be replicated within a web-based environment. Although we found no performance-related differences between fonts, like Bernard et al. (2003) we found participants had a preference for Arial over Times.

4.2. Methodological considerations

Visual search is an essential part of information retrieval in web pages (Kitajima et al., 2000) and therefore a search task was included in Experiment 1. However, search is typically only one part of a larger information retrieval task (van Schaik and Ling, 2003) and so information retrieval was examined in Experiment 2. Our results show that performance on neither task was sensitive to differences in font type, although the performance on the search task was influenced by line length with effect sizes below medium. This type of disparity in results between the two types of task is consistent with the findings of Pearson and van Schaik (2003). Their findings were even more striking because different results were obtained in a repeated measures design, where each participant completed both a visual search task and an information retrieval task (whereas in the current study different participants performed the different types of task). Pearson

and van Schaik found first, no effect of link colour in the visual search task, but blue was found to be significantly quicker than red for the information retrieval task; second, the best performing positions for presenting a navigation bar, in terms of time, for the visual search task were the top and bottom; however, for the information retrieval task, especially for the blue link colour, they were the left and right positions. The authors proposed two possible explanations for the discrepant findings that in principle also apply to the current study. First, the visual search task may lack external validity, or second, the information retrieval task may be systematically biased in some way. After exploring these two explanations they concluded that the information retrieval task appears to mimic real-world tasks more closely and have more real utility than the visual search task, and so perhaps constitutes a more thorough measurement instrument of interaction with web sites. Pearson and van Schaik's research would have been strengthened by a more comprehensive use of subjective measures (only preference for link colour was measured, but this was not separated out per task). However, the current study assessed both preference and aesthetic value for both line length and font type. Preference was sensitive to differences in line length, but much stronger after completing the information retrieval task, and preference was sensitive to differences in font type after completing the search task.

These findings suggest that the greater the importance of the visual search component in a large task, the more important the choice of line length becomes. Furthermore, even when task performance is unaffected by line length, as in the information retrieval task, participants still preferred shorter line lengths, reinforcing the robustness of preference as an outcome measure over both tasks. The sensitivity of preference for differences in font type may reflect an expectation of better performance with a more familiar font (in this case, Arial), perhaps taking into account prolonged use of the font types that were presented (i.e. longer than in the experimental task). Overall, the preference measure was the most sensitive measure used in either study (see also Spenkelink and Besuijen, 2003). In contrast to the preference measure, the aesthetic measure was not sensitive to differences in line length; therefore preference is a more valuable subjective measure. Although both subjective and objective measures were significant, the effect sizes for the subjective measures, were much higher than for the objective methods. Therefore, web designers would appear to be better served by focusing on the findings related to the subjective measures in the absence of effects on objective measures when they develop a site.

In Experiment 2 we found that aesthetics scale was reliable, though in Experiment 1 there were clear reliability problems (perhaps due to the excessive number of line length and font combinations that participants had to evaluate which may have led to a reduction in participants' motivation for this part of the task). Nonetheless, we believe that the scale is itself reliable (see also the results reported in van Schaik and Ling, 2003), and should be used in the future to assess aesthetic qualities of web pages.

4.3. Design recommendations

The research reported herein used multiple methods to explore the effects of varying text presentation on user behaviour. In addition to using two types of task—visual search and information retrieval—we also examined objective and subjective measures. As the summary of the results indicates, there was an effect of line length on both objective and subjective measures, a finding that differed with type of task, although font type only had an effect on subjective measures. These results have practical implications for individuals involved with web design. First, longer line lengths should be used when information is presented that needs to be scanned quickly. Second, shorter line lengths should be used when text is to be read more thoroughly, rather than skimmed. Given a choice between using long (85–100) and short line lengths (55–70), designers should opt for the latter, because although scanning web pages is also a major activity when finding information on web sites (Kitajima et al., 2000), the majority of textual web content is presented to be read rather than skimmed. Third, Arial font should be used. Although, based on the present results, the costs of using Times may not be significant; users expressed a preference for Arial. Such a divergence between performance and preference has been found elsewhere (e.g. Ling and van Schaik, 2002), and has significant implications for the way in which web pages are designed. This preference is likely to become more pronounced, as users will increasingly expect web sites to conform to the already widespread use of sans serif fonts, in accordance with design guidelines (W3C, 2004). Designers may too readily focus on performance data in the construction of web pages; however, our results show that preference data should also be taken into consideration (particularly when there are no differences in performance) as this is likely to have a positive impact on users' acceptance of sites.

Although guidelines exist for choosing fonts for reading On line text, there is a lack of empirical testing of their validity. Moreover, the assumption that design guidance for presentation of large bodies of screen text for reading (such as On line newspapers) is applicable to reading smaller segments of text is highly questionable. This assumption needs to be supported by empirical evidence as it is possible that reading large bodies of text exerts different demands on a user. For example there may be no effect of font type when there is a relatively small amount of text to be read; however, when the length of the text is increased, fatigue or lack of visual clarity of some fonts may lead to poorer performance than with others.

The issues outlined above need to be explored with a wide range of user groups. Although the participants in the present research had a mean age of less than 25, the demographics of web use are shifting and becoming more

comparable to the general population (Cutler et al., 2003). This shift needs to be accompanied by further research. Web designers cannot assume that just because their site is accessible for young users that it will be equally accessible for older users or those with disabilities. Indeed recent research indicates that if web sites are created without attention to accessibility issues, a sizable minority of the adult population will have difficulty using these sites (Microsoft, 2004). Although some researchers have explored the use of the web by older users or those with disabilities (e.g. Bernard et al., 2003; Ingraham and Bradburn, 2003) this is an area that is requires detailed empirical study. What may become clear in the future is that a focus on designing for specific groups may come to have a positive impact on usability for all computer users.

4.4. Conclusion

The current study, using two types of task essential for use of web pages, demonstrated a significant effect of line length. Shorter lengths were overwhelmingly superior in terms of preference and longer line lengths faster for finding information, but only in a visual search task. Although font had no effect on performance measures, overall participants expressed a preference for Arial. Future research should elucidate the generality of these findings for web pages and other online materials, such as academic text.

References

Bernard, M.L., Chaparro, B.S., Mills, M.M., Halcomb, C.G., 2003. Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text. International Journal Human-Computer Studies 59, 823–835.

Bradley, R., 2002. HESDA Briefing Paper 98. HESDA, Sheffield.

Briem, G.S.E., 2002. How to arrange text on web pages. In: Sassoon, R. (Ed.), Computers and Typography, vol. 2. Intellect, Bristol, pp. 10–20.

Brooks, T.A., 2003. Web Search: how the Web has changed information retrieval. Information Research 8(3) paper no. 154 [Available at http://InformationR.net/ir/8-3/paper154.html]

Cedrus Corporation, 2002. SuperLab Pro (Version 2.0) [Computer software]. San Pedro, CA.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Erlbaum, Hillsdale, NJ.

Cutler, S.J., Hendricks, J., Guyer, A., 2003. Age differences in home computer availability and use. Journal of Gerontology 58B, S271–S280.

Davidov, A., 2002. Computer screens are not like paper: typography on the web. In: Sassoon, R. (Ed.), Computers and Typography, vol. 2. Intellect, Bristol, pp. 21–41.

Duchnicky, R.L., Kolers, P.A., 1983. Readability of text scrolled on visual display terminals as a function of window size. Human Factors 25, 683–692.

Dyson, M.C., Haselgrove, M., 2001. The influence of reading speed and line length on the effectiveness of reading from screen. International Journal of Human-Computer Studies 54, 585–612.

Dyson, M.C., Kipping, G.J., 1997. The legibility of screen formats: are three columns better than one? Computers and Graphies 21, 703–712.

Graf, W., Krueger, H., 1989. Ergonomic evaluation of user-interface by means of eye-movement data. In: Smith, M.J., Salvendy, G. (Eds.), Work with Computers: Organizational, Management, Stress, and Health Aspects. Elsevier, Amsterdam, pp. 659–665.

Ingraham, B., Bradburn, E., 2003. Converting OLF Materials for Use Online Multi-PALIO: A Case Study. Open Learning Foundation, London.

Kitajima, M., Blackmon, M.H., Polson, P.G., 2000. A comprehension-based model of web navigation and its application to web usability analysis. In: McDonald, S., Waern, Y., Cockton, G. (Eds.), People and Computers XIV—Usability or Else!. Proceedings of HCI 2000. Springer, London, pp. 357–373.

Lindgaard, G., Dudek, C., 2003. What is this evasive beast called user satisfaction? Interacting with Computers 15, 429–452.

Ling, J., van Schaik, P., 2002. The effect of text and background colour on visual search of Web pages. Displays 23, 223–230.

List, D., 2001. Screenreading. Retrieved 7 May, 2003 from the World Wide Web: http://www.dennislist.net/scread.html.

Microsoft Corporation, 2004. The wide range of abilities and its impact on computer technology. Retrieved 8 July 2004 from the World Wide Web: www.microsoft.com/enable/research/default.aspx

Moneta, G., Csikszentmihalyi, M., 1996. The effect of perceived challenges and skills on the quality of subjective experience. Journal of Personality 64, 275–310.

Morkes, J., Nielsen, J., 1997. Concise, scannable and objective: How to write for the Web. Retrieved 14 September, 2003 from the World Wide Web: http://useit.com/papers/webwriting/writing.html.

Nielsen, J., 2002. Let users control font size. Retrieved 14 July, 2004 from the World Wide Web: http://www.useit.com/alertbox/20020819.html

Nielsen, J., 2005. Lower-literacy users. Retrieved 14 April, 2005 from the World Wide Web: http://www.useit.com/alertbox/20050314.html.

Nygren, E., 1996. From paper to computer screen: human information-processing and user interface design. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology, 188. Almqvist and Wiksell, Stockholm.

Pearson, R., van Schaik, P., 2003. The effect of spatial layout and link colour of web pages on performance in a visual search task and an interactive search task. International Journal of Human-Computer Studies 59, 327–353.

Shneiderman, B., 2000. Universal design. Communication of ACM 43, 84–91.

Schmidt, K.E., Bauerly, M. Liu, Y., Sridharan, S., 2003. Web page aesthetics and performance: a survey and an experimental study. Proceedings of the Eighth Annual International Conference on Industrial Engineering: Theory, Applications and Practice. Las Vegas, Nevada, November 10–12, 2003.

Scott, D., 1993. Visual search in modern human-computer interfaces. Behaviour and Information Technology 12, 174–189.

Spencer, H., 1968. The Visible Word. Royal College of Art, London.

Spenkelink, G., Besuijen, K., 2003. Visual display quality. In: Daily, J.E., Kent, A., Lancour, H. (Eds.), Encyclopedia of Library and Information Science. Marcel Dekker, New York, pp. 3066–3082.

Streveler, D.J., Wasserman, A.I., 1984. Quantitative measures of the spatial properties of screen designs. In: Shackel, B. (Ed.), Proceedings of Interact '84 Conference on Human-Computer Interaction. Elsevier, Amsterdam, pp. 125–133.

Tractinsky, N., Katz, A.S., Ikar, D., 2000. What is beautiful is usable. Interacting with Computers 13, 127–145.

Trewin, S., Zimmerman, G., Vanderheiden, G., 2004. Abstract representations as a basis for usable user interfaces. Interacting with Computers 16, 477–506.

Tullis, T.S., 1997. Screen design. In: Helander, M., Landauer, T.K., Prabhu, P. (Eds.), Handbook of Human-Computer Interaction. Elsevier, Oxford, pp. 503–531.

van Schaik, P., Ling, J., 2003. The effects of screen ratio and order on information retrieval in web pages. Displays 24, 187–195.

W3C, 2004. Web content accessibility guidelines 2.0. Retrieved 19 July, 2004 from the World Wide Web: http://www.w3.org/TR/WCAG20/.

Wolfe, J.M., 1994. Guided search 2.0: a revised model of visual search. Psychonomic Bulletin and Review 1, 202–238.