

The Effect of Age and Font Size on Reading Text on Handheld Computers

Iain Darroch, Joy Goodman, Stephen Brewster, Phil Gray

Glasgow Interactive Systems Group, Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ, UK
{darrochi, joy, stephen, pdg}@dcs.gla.ac.uk
www.dcs.gla.ac.uk/utopia

Abstract. Though there have been many studies of computer based text reading, only a few have considered the small screens of handheld computers. This paper presents an investigation into the effect of varying font size between 2 and 16 point on reading text on a handheld computer. By using both older and younger participants the possible effects of age were examined. Reading speed and accuracy were measured and subjective views of participants recorded. Objective results showed that there was little difference in reading performance above 6 point, but subjective comments from participants showed a preference for sizes in the middle range. We therefore suggest, for reading tasks, that designers of interfaces for mobile computers provide fonts in the range of 8-12 point to maximize readability for the widest range of users.

1 Introduction

Small screen user interfaces, exemplified by personal digital assistants (PDAs), are becoming more popular and more affordable. Uses include web surfing, reading e-books, reading email and listening to music.

The small screen provides challenges for interface designers but we are lacking design guidelines for creating such interfaces [14]. Some information on interface design for handhelds is given by Weiss [22] but few data are available on how this varies with age. According to estimates from the US Census Bureau's International Database (2004), the proportion of those in the UK who are over 60 is expected to increase from 20% in the year 2000 to 27% by 2025. Increasing age leads to declines in various abilities such as losses in visual contrast sensitivity [1]. In a prior study [11] on designing navigation aids for older people, no information on a suitable font size for handhelds for older people could be found. Desktop guidelines were used but were not entirely satisfactory because there was some indication that older people might be able to read smaller text sizes on handheld computers than recommended by the guidelines. Therefore, we felt it was important to investigate this further and clarify whether there are different requirements for handheld computers.

The problem of how best to display textual information on small screens has been studied. For example, presenting text dynamically (e.g. vertical scrolling) [15] and analyzing web design guidelines for applicability to small screen interfaces [14]. In

this paper we study the effect of the size of the text on readability on small screens, specifically a PDA.

Although few studies exist of text display on small screens there have been a number of studies examining reading text on large screens, such as CRT monitors, (e.g. [2],[19]) and on-line (e.g. [4]). These studies were based on prior reading studies of text presentation on paper (e.g. [20]). Therefore, it is a logical next step to carry out similar text presentation tests on small screen computers. Comparative studies between paper and on-line reading performance have found no significant performance differences [12], but have found differences in users' subjective preferences. Image quality is an important factor in this. It has been found that an increase in image quality results in an increase in subjective performance rating for both paper and on-line reading [13]. Features of CRT monitors, such as screen flicker and luminance, can affect reading performance [10]. Therefore, it may be expected that there will be differences between reading performance on a small screen display compared to a CRT monitor. Recommendations for text sizes from previous studies have indicated font size 14 for children [3], font size 14 for older adults [4], and font size 12 for young to middle-aged adults [2]. This indicates an age-related change in font size on desktop computers but there are no corresponding findings for handheld computers.

In this paper we aim to elicit an indication of a suitable font size to use with text presentation on handheld computers and determine whether different font sizes are required when designing for older people. We also investigate whether the need to scroll when reading text has an effect on which font size should be chosen. We do not consider changing font type in this case to simplify the experiment; it will be investigated in a future study. The next section outlines the experiment used in this study. The results from the experiment are then presented and discussed. Some areas for further investigation are suggested. Finally the conclusions drawn from our experiences are given.

2 The Experiment

The experiment was a 2 x 8 factor within-subjects repeated-measures design. The first factor was age (two levels: younger adults and older adults) the second factor was font size (8 levels: 2,4,6,8,10,12,14 and 16 point).

2.1 Participants

Twenty-four participants took part in the experiment and were divided into two groups of 12 with 6 males and 6 females per group. The Younger Adults group was aged 18-29 and the Older Adults group was aged 61-78. All participants were fluent in English as their first language and educated to at least secondary/high school level. A Snellen near visual acuity test for average reading vision at a distance of 40.6cm was used to test participant's near vision before the experiment. All participants had 20/40 vision or better. Participants had no or very minimal experience of handheld

computers before the experiment. A £5 book token was given to participants as payment for taking part.

2.2 Equipment

An HP iPAQ hx4700 (www.hp.com) which has a 65,000 colour TFT screen with a resolution of 640x480 pixels was used to present the text (see Figure 1). This has the best quality screen available at the time of writing (January 2005). The screen was backlit and participants sat in a usability lab which was illuminated by overhead fluorescent lights. The iPAQ used the Microsoft Windows Mobile™ 2003 Second Edition operating system and had ClearType enabled to anti-alias the edges of fonts to improve quality. Custom software was used to present the experimental texts. An example of the software running on the iPAQ is shown in Figure 2.



Fig. 1. An HP iPAQ hx4700 as used in the experiment

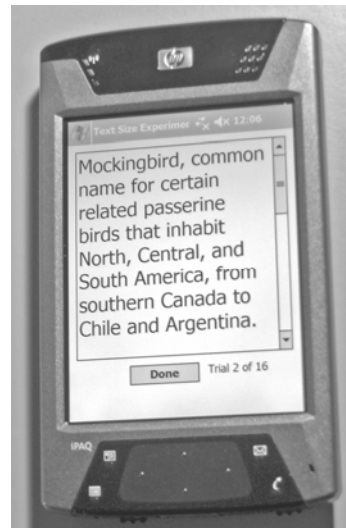


Fig. 2. A screenshot of the application used in the experiment

2.3 Task

Many possible measures could be used to determine the effect of font on reading performance. Setting a task in which participants identify spelling or typographical errors is difficult due to inconsistency in the misspellings used and difficulty in measuring the degree of change in word shape. These types of test also promote skimming behaviour [7]. It has also been found that readers can differ in their ability to detect typographical errors [13]. Post-reading comprehension tests are another option but it is likely that participants will scan passages looking for the main points rather than reading the text. Asking participants to proof read a passage and read the

words out loud would ensure that the passage was read but would not be very realistic since the flow of reading would be broken by having to speak continuously. Dillon [8] points out that many studies into reading performance bear little resemblance to normal reading and argues that tasks should be more realistic.

Jorna and Snyder [13] suggest the introduction of word substitution errors in proof reading tasks, making sentences incomprehensible and which force the subject to read and comprehend the sentence. For example, the word “toe” could be substituted for the word “cake” in the sentence “I baked a cake”, thus making the sentence incomprehensible to someone reading it. However, Gunjar *et al.* [12] found that subjects sometimes re-read sentences to make sense of them and so constrained the words used for substitution in two ways: the substituted word rhymed with the original word; and the substituted word varied grammatically from the original word. For example, the word “fake” could replace “cake” in the sentence “I baked a cake”. This modified proof reading task was successfully used by [2, 3] and was thus chosen for our study (see examples below).

The task ensures a realistic approach because subjects must read the entire passage in order to recognize substituted words. The words chosen for substitution were common English words that were clearly out of context to ensure that fluent English readers would have no trouble in identifying the errors.

2.4 Fonts and Passages

The standard Microsoft Sans Serif font was chosen for displaying text since it has been found that Sans Serif fonts are preferred by subjects in reading computer displayed text [2]. Text was presented to participants at font sizes 2, 4, 6, 8, 10, 12, 14, and 16. Examples of each size are shown in Figure 3. On screen the fonts ranged in size from less than 1mm high for size 2, to 5mm high for size 16.

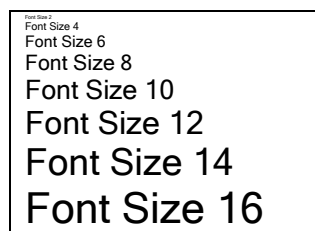


Fig. 3. Examples of the font sizes used in the experiment

Two different lengths of passages were used. The ‘short’ passages were of a length that fitted on a single screen up to font size 12 but required scrolling at font size 14 and 16. The ‘long’ passages fitted on a single screen up to font size 8 and required scrolling at font size 10 and above (see Figure 4). The different lengths and sizes required for scrolling meant that we could investigate the effects of scrolling on reading performance. Within the two groups of passages, the length of passages was adjusted to have approximately the same number of characters (Short: $M = 230.7$

chars per passage, S.D. = 2.9 chars; Long: M = 460.7 chars per passage, S.D. = 4.1 chars).

There was one substituted word in the short passages, and two in the long passages. Text for the passages was taken from Microsoft's Encarta encyclopaedia [16]; specifically from Life Science > Mammals, Birds, Reptiles & Amphibians, Invertebrate Animals. This ensured consistency between passages since all were written at approximately the same level of difficulty and discussed similar topics. Thirty-two passages were created, 16 for each passage length. The order in which the 16 passages were presented was the same for all participants. There was always two of each font size in the sixteen passages with a different font size ordering for long and short passages. No two participants were given the same font size order. Figure 4 shows an example of each passage length. The substitutions are: the word *shore* at the end of the first passage; *plains*, the fifth word on the second line of the second passage; *sneeze*, the eleventh word on the fourth line of the second passage.

Elephant, huge mammal characterized by a long muscular snout and two long, curved tusks. Highly intelligent and strong, elephants are among the longest-lived, with life spans of 60 years or shore.

Monkey, any of about 160 species of primates that have grasping hands, forward-facing eyes and highly developed plains. Most monkeys also have tails, a characteristic that distinguishes them from their larger primate cousins, the apes. Monkeys are highly skilled climbers, and most spend much of their lives in sneeze. Some have prehensile tails - that is, tails capable of grasping - that they can use as a fifth limb whilst foraging for food or climbing.

Fig. 4. One short and one long example passage from the experiment

2.5 Measurements

Both reading speed and reading accuracy were recorded. A timer within the software recorded the time taken to read a passage. Accuracy was measured by the experimenter noting down the words identified as contextual errors by the participant.

It is important, as Dillon [8] argues, that analysis of readability should consider more than reading performance. The use of subjective measures in addition to visual performance should be included in legibility testing [18]. In other reading preference studies it has been found that no difference exists in reading performance, but the subjective view of reading performance did differ between texts [2]. Therefore, participants were asked what they thought of each font size used, and to pick a preferred font size.

2.6 Procedure

After a briefing on the experiment and some background information questions, participants did a number (minimum 6) of training passages to familiarize themselves with the iPAQ and what was required in the task. Participants were asked hold the

iPAQ and to read the passages from a comfortable position and were told that they could bring the iPAQ closer to the face if necessary.

The software used to present passages to participants had a “Start” button that was pressed to begin reading the passage and a “Done” button, pressed upon completing the passage (see Figure 2). Users were presented with a series of passages and for each pressed the start button, read the passage (saying out loud any word substitutions), then pressing done. Instructions were given to only say the erroneous word(s) and nothing else while reading a passage and keep questions/comments for the breaks between passages. To avoid effects of eye-strain or fatigue participants were told to rest for as long as they wanted between passages. Furthermore, if a font was unreadable or would cause too much discomfort to read, participants were instructed to tell the experimenter and skip the passage. Participants were asked to read passages once through only and identify any errors. The number of substituted words in passages was not told to the participants.

After being presented with an initial set of 16 passages to read users answered questions on what they thought of the different text sizes and were asked to pick a preferred text size by browsing through the passages. This questioning served a secondary purpose in giving users a rest between sets of passages. Having answered the questions, participants were then given a further 16 passages to read after which they were asked about their views on the text sizes used and to choose a preferred size. The first set of passages was of one length (‘short’ or ‘long’) and the second of the other length. The order of presentation of the two lengths was counterbalanced.

3 Results

This section presents the results of the study. Initially the results of the objective measures of reading performance are given before the participants preferred text size and views on the font sizes are presented.

3.1 Reading Performance

Twelve participants from each group read two passages at a given text size for each text length. This gave 24 records of reading time and accuracy per group for the short and long text passages. At the smallest font sizes (sizes 2 & 4) some participants, particularly in the older group, found the text uncomfortable or just impossible to read, so chose not to read it. Font size 2 in particular caused older participants problems, yielding only 2 results for short passages and 3 results for long passages. However, from font size 6 upwards no problems were had with text legibility. Font size 4 posed no problem for younger participants, yielding 24 results while for older participants 19 results were obtained for short passages and 16 for long passages.

Reading Time. Times for completion of reading a passage were recorded to a tenth of a second and then normalized on the fastest completion time. Normalisation was used to remove any effects of base reading speed and reading abilities among participants. For example, those who read regularly would be expected to read faster than those reading infrequently. The graph in Figure 5 illustrates the normalized reading time for both groups for both sets of passages.

The average reading time for older people at font size 2 has not been plotted since there were an insufficient number of data values to get a reasonable representation of the average reading time. The graph shows that there was little difference in reading time between groups for sizes 6-16. However, each group had a lower bound at which reading becomes difficult and times slow, for the younger group this is at size 2 while the older group it is at size 4. An analysis of variance between different passage lengths and age groups found that there was no significant difference in reading times ($p > 0.05$ in all cases). Though there was a slight increase in reading time at font size 16 it was not statistically significant.

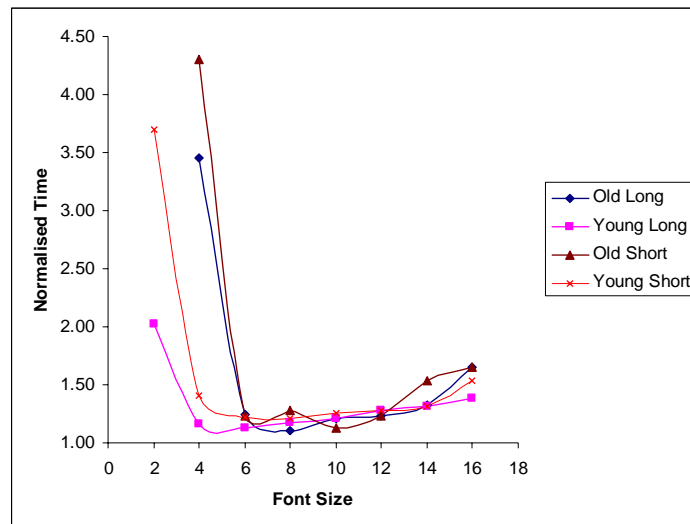


Fig. 5. Time taken to read each passage at each font size by our two user groups.

Accuracy. In the tasks there was 1 error per short passage and two errors per long passage. The percentage of correctly identified errors was over 85% in all but 2 cases. It did drop to 66.67% for young people and short passages at font size 2 and 78.13% for older people and long passages at font size 4.

Overall accuracy is very high, over 90% in many cases. There is a small degradation in accuracy for both the older and younger group at the font size where reading speed slowed. This indicates that those sizes are the lower bounds of readability.

Originally it had been planned to follow Bernard's [2] example and calculate an adjusted accuracy measure. That is the ratio of time take to read the passage to the

percentage of errors found. However, the ceiling effect seen by the accuracy measure meant that nothing meaningful could be drawn from such results.

3.2 Subjective Results

Font Size Preference. Using a Wilcoxon Signed Rank test the preferences for each age group and passage length was examined. A confidence interval of font sizes from 9-11 (12 for older people) are preferred for reading text on the PDA screen. The younger group has a slightly smaller median (10) to the older group (11). These results are the same for both long and short passages.

It is important to consider whether there are any significant differences between young and old participant's preferences or whether preferences change between long and short passages. A Mann-Whitney test was used to analyse whether there was a difference between the groups. The p-values were all greater than 0.05, with no significant difference found between the sets of data. This suggests that, in this case, neither age group nor passage length had a strong effect on size preference.

With no effect of age or passage length found, a 1-sample Wilcoxon Signed Rank test was run on the combined preferences of all participants. This gave a Median of font size 10 and a confidence level of 95% with a confidence interval ranging from 10 to 11. These data indicate that a font size between 10 and 11 is preferred for reading text on a PDA. (Non-parametric tests were used as the data were based on rankings).

Qualitative Analysis of Comments. Users' comments on specific font sizes were examined and ranked on a five point scale: -2 (very negative), -1 (negative), 0 (neutral), 1 (positive), 2 (very positive). Examples of comments received were: "Just impossible for me to read. Well maybe if I screwed up my eyes but I would not be comfortable reading that size of text", "Rubbish. Too big. A waste of space", "perfectly clear. Nice and bold. I like that one". Two researchers independently rated the comments. Pearson product moment correlation statistical analysis was used and it was found that the ratings from the two researchers correlated ($r > 0.5$; $df = 10$ in all cases). However, this does not show any indication of the difference in magnitude of the two researchers' ratings. Further analysis revealed disagreement in less than 22% of cases and disagreement was never greater than one point on the rating scale. Therefore, there was close agreement between researchers' ratings. The ratings were combined by averaging them. The graph in Figure 6 illustrates the average comment ratings about each font size from the old and young group with respect to the long and short passages.

All groups of participants agreed that sizes 2 and 4 were undesirable. Size 6 had a slight positive comment from all users while size 8 was considered positive. The younger group rated size 10 as positively as size 8 for both long and short passages but were more negative toward size 12 and were negative about sizes 14 and 16. The older group commented most favourably on size 12 for short passages but this was only 0.04 more positive than size 10 (0.08 more than size 8). Sizes 14 and 16, received positive comments but distinctly less positive than sizes 8-12. The comments

about the long passages from the older group gave size 10 as the font size most positively commented upon. Sizes 8 and 12 were also given positive comments. Once again sizes 14 and 16 received less favourable comments, both fairly neutral.

A Mann-Whitney test was used to analyse whether there was a difference between the groups. The p-values were all greater than 0.05, with no significant difference found between the sets of data. This suggests that, in this case, neither age group nor passage length has a strong effect on subjective views of font size.

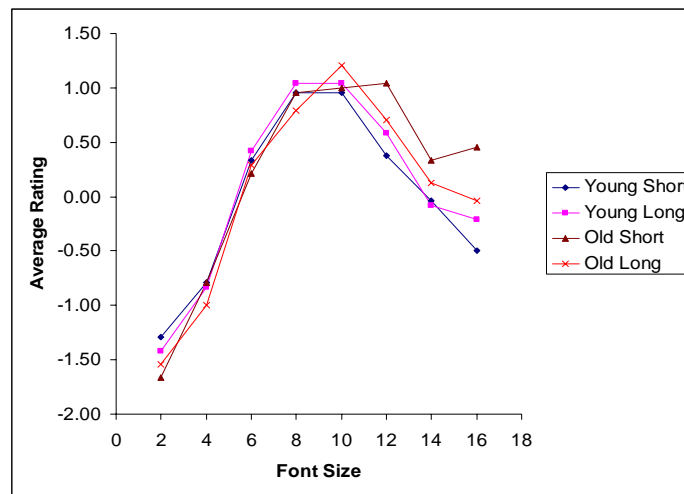


Fig. 6. Average ratings of the comments received about each font size.

Summary. There was no difference in preference identified due to passage length or between age group preferences. Overall a font range of 10-11 was preferred. This was reflected by users' comments about font sizes where size 10 received high positive comments, as did size 8. In addition, older people also commented positively about size 12. The smaller font sizes (2, 4) were disliked as were the larger font sizes (14, 16) by younger group.

4 Discussion

The results show that objective measures (time and accuracy) of reading performance are not affected significantly by changes in font size (within limits) or passage length for both the old and young group. This corresponds to the findings of Tinker [21] that varying the typeface and size of text within common parameters only has limited impact on readability. This owes much to the adaptability of the human visual system to be able to process diverse presentation of data. The lower bound for text size appears to be size 4 for the younger group and size 6 for the older group since below that size reading time degrades. However, no upper bound was apparent from the

results. However, subjective comments show a negative feeling toward the largest font sizes indicating that there is may be an upper bound, but perhaps we did not go high enough in our font sizes to find the upper bound. Mills & Weldon [17] found that 80 characters per line were easier to read than 40 characters per line. As font size increases the number of characters per line decreases. Therefore, it would be expected that there will be an upper bound to maintain reading performance as the font size increased. At font size 4 we see a large disparity between young and old participants' times for reading passages. These deficits are potentially due to age-related losses in visual contrast sensitivity [1]. This is backed up by some comments from older participants stating that they preferred higher contrast text (e.g. "a decided black is better than grey for text colour.....the smaller text sizes were lighter").

The ceiling effect seen in the accuracy results was unfortunate since it did not allow analysis of reading time versus accuracy. The effect could be addressed in future experiments by increasing the number of contextual errors per passage. However, too many errors would make the task unrealistic compared to normal reading. Further research is needed to determine an optimum number of errors.

The results of the objective measurements showed no significant difference due to font size, age, or passage length. The subjective results also show no significant effects due to age or passage length. However, there is a clear effect due to change in font size.

Overall, participants preferred a font in the range 10-11 but more interesting were the comments about the text sizes. The comments, to some extent, reflected the preference findings. Young people were most positive about sizes 8 and 10 and preferences showed a range of 9-11. The preferred font size of older people fell in the range 9-12 and they commented positively about sizes 8, 10, and 12. Unsurprisingly neither group commented positively about the smallest (2, 4) or largest font sizes (14, 16). The discomfort experienced in reading small text explains the dislike for the small sizes. The larger sizes were rejected because the "words are spread out more" which "breaks up the flow of reading". Interestingly, the objective measures do not show any significant effect due to the broken reading. However, the subjective comments show this was disliked by participants, maybe because more effort was required to derive the meaning of sentences.

It can be seen from both font size preference and comments about font sizes that a slightly larger range is associated with the older group than with the younger group. This indicates that older users vary more widely in their subjective preferences. Therefore, when considering older users in design, a slightly wider range of fonts should be allowed, including larger ones. This applies particularly to the shorter passages. A possible explanation is that size 12 is the largest font that requires no scrolling with the short passages. However, from the objective measures it was seen that passage length had no significant effect on performance. Therefore, it may be the case that users would prefer not to have to scroll even though it has little effect on their reading performance. Comments from some users reflected this with a preference for "seeing text on one page". Allowing font sizes in the range 8-12 would provide reasonable user satisfaction and ensure good reading performance

The sizes in this range may seem smaller than one would expect based on previous desktop computer based text reading studies. For example, Bernard *et. al.* [2] found size 12 produced greater subjective readability and lower levels of perceived

difficulty, therefore, one would expect size 12 to be in the middle of the range. However, font size 10 at a resolution of 640x480 is approximately the same height as font size 12 at a resolution of 1024x768 [2] for the same screen. Therefore, the lower resolution of our screen compared to that of previous desktop computer reading studies could account for the smaller font sizes we found.

4.1 Comparing Our Results to Previous Research

It may have been expected that one font size would come out as the “best” or most favourably commented upon for reading text on handheld computers. Instead, we ended up with a range of sizes. This can be explained by the fact that the reading distance during the experiment was not fixed. Subjects could move the iPAQ closer or further from their eyes as necessary; experimental observation confirms participants varying the distance of the iPAQ from their face. This allowed the angular character size to be changed. Akutsu *et. al.* [1] found that reading speed was maximal for both young and old people within a given angular character size range (0.3° to 1.0°). This would explain the similar performance of all groups from font size 6-16 and possibly the range rather than a specific preference. An analogy to consider is that of reading a book. Publishers use different font sizes (and book sizes), each requiring the book to be held a different distance from the eyes for the most comfortable reading.

The handheld computer used in this study had a screen resolution of 640x480 pixels (currently the best available, and a significant improvement over the previous generation of the iPAQ device where characters below 6 point were not rendered clearly) which is becoming the common screen resolution on handheld computers. Therefore, our findings will continue to apply to handheld technology for the near future. In 2002, Karkkainen suggested 14 point font for reading text on handheld computers. The resolution of the device he used was 320x240. This is lower than that used in our experiment and is the likely explanation for the different findings. The LCD screen technology for handheld devices is changing rapidly for the better. The anti-aliasing used in the current version of the Windows Mobile operating system makes the characters much easier to see at small sizes. This indicates that our findings may not be applicable to future displays with improved resolutions. However, our findings that a range of sizes is preferred are likely to be the same for future screens, with the bounds of the range changing with resolution changes.

The results in this paper should be taken with the caveat that they only apply to the particular device and screen used in this study, although are likely to apply to other small screen devices with similar displays. The quality of presentational format can have a major influence on both reading speed for learning and comprehension [10]. As the quality of the screens on handheld computers improve, better performance could be expected just as improvements in computer monitors lead to improvements in screen reading such that they are now comparable to print reading speeds.

4.2 Other Observations

At the smallest text size (2), few participants from the older group attempted to read the text. This was because they had been given the option to pass on a passage if it was going to be too much of a strain. However, they may have been able to read it if they had tried. In such a situation the experimenter is faced with a dilemma. There are ethical issues involved in forcing a participant to perform a task that may cause discomfort. However, as in this case, there can be a fine line between extracting useful research results and the comfort of the participant. This adds difficulty in designing tasks and procedure for an experiment, especially involving older people. An alternate view is that gaining a measurement for reading a passage that would never be read in practice is not a useful result. What is useful, however, is finding the limits of what would be read in practice.

A point of interest to those considering doing similar research to this study is to choose words for rhyming carefully. For example, one replacement used in this study was the word “clear” for the word “deer”. This meets the requirements of rhyming and sufficiently out of context as to not require rereading of sentences. However, at smaller fonts the letters ‘c’ and ‘l’ become less distinguishable and look very much like the letter ‘d’ (e.g. *ca*). This makes the substituted word very like the original word and means it can easily be missed. The unfortunate choice of word was pointed out by one of the last participants, but it only affected the passage at font size 2. Furthermore, the overall performance of accuracy was such that this did not have a serious impact on the results.

One other import factor is mobility. The research presented in this paper was all done with the participants seated in a quiet usability laboratory. The iPAQ is a handheld computer designed to be used in mobile situations. If the user is mobile then that is likely to have a large impact on the size of font required. As the user moves the device moves, making the screen harder to see. A mobile environment can also have changing lighting conditions which can make the screen hard to see and so change font size requirements. Brewster [6] found a very significant effect on performance when users used a stylus/touch screen interface when on the move. A 32% reduction in tapping performance and a 17% increase in subjective workload were found when users were walking outside as compared to sitting in a usability lab. Therefore the experiments described in this paper should be replicated in a mobile situation to gain more knowledge of appropriate font sizes (something we are planning to do in the near future). However, there are very many cases where users of handheld computers use them when sat down or stood still, so the results described here are significant.

5 Future Work

Our study has given some indications of the text size that should be used and paved the way for further research into suitable text formats to be used for reading on handheld computers. This study used only one font type but previous studies have compared different font types, particularly serif and sans-serif fonts. However, Boyarski *et al.* [5] found that 10 point Georgia (serif) and Verdana (sans-serif) were

equally readable. Since these two font types were specifically designed for screen use it is quite possible the same findings would occur if they were used on handheld computers. It was also found that fonts designed for screen that had relatively large x-heights performed well in on-line reading performance [5]. Future studies should investigate reading performance by varying font type (both serif and sans-serif), x-height, and font size. This study has provided bounds within which font size should be varied.

It has been found that line length is a more important factor in reading than line height [9]. Therefore, a future investigation could examine the differences in reading when text is displayed in portrait or landscape format.

As was discussed previously, text reading performance experiments are not always realistic to actual reading. Handheld computers bring another factor to the realism, that of environment. The portability of small screen devices means they can be used in many locations each varying in the distracters it contains. For example, a commuter could read the latest news or novel on a PDA on the train to work. However, the stop-start nature of travel, background noise and vibrations could all have an effect on reading. A planned future experiment is to repeat the study in this paper in a 'real' mobile environment. The difficulty is being able to ensure a consistent environment for comparable results with so many potential variables to account for.

6 Conclusions

The purpose of this study was to examine the effect of different font sizes on reading text on handheld computers and to consider the differences between young and old people. Although there were no significant differences (for sizes 6-16) in reading performance or accuracy due to either passage length or age, there was variation in subjects' preferences on the text sizes used. The range of preferred or positively commented upon sizes was slightly greater (at the large side) for older participants than for younger participants. The amount of text presented and so the amount of scrolling required does not have an effect on reading performance.

We recommend that designers creating applications for reading text on a small screen with resolution of 640x480 should offer the choice of small (font size 8), medium (font size 10), or large (font size 12) sizes to cater for the needs of most users. The choice should consider the amount of text that will be presented at once. Ideally, designers will allow for a range of text sizes to accommodate most users.

Acknowledgements

This work was funded by SHEFC through the UTOPIA project (HR01002).

References

1. Akutsu, H., Legge, G.E., Ross, J.A., Schuebel, K.: Psychophysics of Reading: X. Effects of Age Related Changes in Vision. *Journal of Gerontology: Psychological Sciences*, 46, 1991, 325-331.
2. Bernard, M.L., Chaparro, B.S., Mills, M.M., Halcomb, C.G.: Comparing the Effects of Text Size and Format on the Readability of Computer-Displayed Times New Roman and Arial Text. *International Journal of Human-Computer Studies*, 59 (2003) 823-835.
3. Bernard, M.L., Chaparro, B.S., Mills, M.M., Halcomb, C.G.: Examining Children's Reading Performance and Preference for Different Computer-Displayed Text. *Journal of Behaviour and Information Technology*, 21(2), 2002, 87-96.
4. Bernard, M.L., Liao, C.H., Mills, M.M.: The Effects of Font Type and Size on the Legibility and Reading Time of Online Text by Older Adults. In Vol. II Proceedings of ACM CHI 2001, ACM Press, 2001, 175-176.
5. Boyarski, D., Neuwirth, C., Forlizzi, J., Regli, S.H.: A Study of Fonts Designed for Screen Display. In Proceedings of ACM CHI 98, ACM Press, 1998, 87-94.
6. Brewster, S.A.: Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing*, 6(3), 2002, 188-205.
7. Creed, A., Dennis, I., Newstead, S.: Proof-Reading on VDUs. *Journal of Behaviour and Information Technology*, 6, 1987, 3-13.
8. Dillon, A.: Reading from Paper versus Screens: A Critical Review of the Empirical Literature. *Ergonomics*, 35(10), 1992, 1297-1326.
9. Duchnicky, R.L., Kolars, P.A.: Readability of Text Scrolled on Visual Display Terminals as a Function of Window Size. *Human Factors*, 25(6), 1983, 683-692.
10. Garland, K.J., Noyes, J.M.: CRT Monitors: Do They Interfere with Learning?. *Journal of Behaviour and Information Technology*, 23(1), 2004, 43-52.
11. Goodman, J., Gray, P.D., Brewster S.: How Can We Best Use Landmarks to Support Older People in Navigation? *Journal of Behaviour and Information Technology*, 24(1), 2005, 3-20.
12. Gujar, A.U., Harrison, B.L., Fishkin, K.P.: A Comparative Evaluation of Display Technologies for Reading. In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Chicago IL, 1998, 527-531.
13. Jorna, G.C., Snyder, H.L.: Image Quality Determines Differences in Reading Performance and Perceived Image Quality with CRT and Hard-Copy Displays. *Human Factors*, 33(4), 1991, 459-469.
14. Karkkainen, L., Laarni, J.: Designing for Small Display Screens. In Proceedings of the Second Nordic Conference on Human-Computer Interaction (Aarhus, Denmark, October, 2002), ACM Press, 227-230.
15. Laarni, K.: Searching for Optimal Methods of Presenting Dynamic Text on Different Types of Screens. In Proceedings of the Second Nordic Conference on Human-Computer Interaction (Aarhus, Denmark, October, 2002), ACM Press, 219-222.
16. Microsoft Corporation: Encarta Online Encyclopedia. <http://encarta.msn.com>, 2004.
17. Mills, C.B., Weldon, L.J.: Reading Text from Computer Screens. *ACM Computing Surveys*, 19 (4), 1987, 329-358.
18. Mustonen, T., Olkkonen, M., Hakkinen, J.: Examining Mobile Phone Text Legibility while Walking. In Vol. II Proceedings of ACM CHI 2004, ACM Press, 2004, 1243-1246.
19. O'Hara, K., Sellen, A.: A Comparison of Reading Papers and On-Line Documents. In Proceedings of ACM CHI 97, ACM Press, 1997, 335-342.
20. Poulton, E.C.: Letter Differentiation and Rate of Comprehension of Reading. *Journal of Applied Psychology*, 49, 1955, 358-362.
21. Tinker, M.A.: Legibility of Print. Iowa State University Press, Ames, Indiana (1963).
22. Weiss, S.: Handheld Usability. John Wiley & Sons (2002).