Please find the publisher's version at:

http://hfs.sagepub.com/content/early/2013/04/09/0018720813485089

READINGMATE: THE EFFECT OF THE CONTENT STABILIZING TECHNIQUE, FONT SIZE, AND INTERLINE SPACING ON THE LETTER-COUNTING TASK PERFORMANCE OF TREADMILL RUNNERS

Authors:

Bum Chul Kwon (kwonb@purdue.edu): Purdue University, West Lafayette, IN 47906, USA

Ji Soo Yi (yij@purdue.edu): Purdue University, West Lafayette, IN 47906, USA

Yu Zhu (yuzhu@purdue.edu): Purdue University, West Lafayette, IN 47906, USA

Contact information for the author to whom requests for reprints should be sent: Ji Soo Yi, 315 N. Grant St., West Lafayette, IN 47906, USA

Keywords: reading performance, reading while running, head tracking, font size, interline spacing, ReadingMate, treadmill, letter counting

Précis

The effects of font size, interline spacing, and a technology called, "ReadingMate," which stabilizes contents on a display using a vision-based head tracking technology, were investigated in the reading-while-running context.

Abstract

Objective: Investigate the effects of font size, interline spacing, and a technology called ReadingMate on the letter-counting task performance of users running on a treadmill.

Background: Few studies have investigated how runners read text while running on a treadmill. Our previous studies showed that ReadingMate had positive effects on the reading while running experience (Kwon & Yi, 2009, 2010); however, the effect of other text conditions (i.e., font size and interline spacing) and the interplay between ReadingMate and such text conditions on the letter-counting task performance are not clearly understood.

Methods: Fifteen participants were recruited for the experiment. There were three main factors: display types (Normal and ReadingMate), font sizes (8-, 12-, 16-, and 20-point), and interline spacing (1.0×, 1.5×, 2.0×, and 2.5×). The researchers employed a letter-counting task. The performance was measured regarding task performance time, success rate of counting the target letter *f*, and number of give-ups.

Results: Overall, the letter-counting task performance while running on a treadmill improved as font size and interline spacing increased, as expected. ReadingMate was more effective than normal display particularly when text was displayed in a small font size and with dense interline spacing.

Conclusion: When text must be displayed in a small font size and with dense interline spacing, ReadingMate can be used to improve the users' task performance.

Application: Practical applications of ReadingMate include improving the text reading experience in shaky environments, such as in aviation, construction, and transportation.

INTRODUCTION

It is obviously challenging to read text while running on a treadmill (Kwon & Yi, 2009, 2010). Due to the user's constant head movement, the subsequent eye compensation makes the user become tired easily. To help overcome this challenge, the authors proposed a potential solution called ReadingMate (Kwon & Yi, 2009). ReadingMate (1) tracks users' head movements and (2) adjusts the position of content on a display accordingly; therefore, the content looks still from the runner's point of view. In two prior studies (Kwon & Yi, 2009, 2010), we found positive outcomes in subjective measures regarding ReadingMate.

Despite such encouraging results, these studies did not provide a complete picture. Our observation and interview results alluded that there were other factors influencing the reading-while-running performance, such as font size and interline spacing (Kwon & Yi, 2010). These factors are known to affect reading performance on a static display (Bernard, Chaparro, Mills, & Halcomb, 2003; Grahame, Laberge, & Scialfa, 2004; Ling & van Schaik, 2007; Sheedy, Subbaram, Zimmerman, & Hayes, 2005); however, the effects of the factors on the reading-while-running situation are unknown. Furthermore, the interplay between these conditions and ReadingMate has not been studied.

In this study, we investigated the effects of font size, interline spacing, and ReadingMate on the performance of the letter-counting task, where participants were asked to count the number of a target letter *f* while running on a treadmill. In particular, this study reveals the specific text conditions under which ReadingMate can work more efficiently.

BACKGROUND

ReadingMate is a technology that helps a runner read text while running on a treadmill by adjusting the position of text on a display according to the runner's head movement

(Kwon & Yi, 2009). As shown in Figure 1, ReadingMate is comprised of (a) goggles with a pair of attached light-emitting diodes (LEDs), (b) a Wii Remote, (c) the ReadingMate software, and (d) a display. The Wii Remote is a remote controller for the Nintendo™ Wii™ game console that contains an infrared camera at its tip as well as a Bluetooth wireless communication module. These components enable a Wii Remote to capture the head movements of a treadmill runner by detecting infrared lights emitted from the LEDs and sending the head position data to a laptop via a Bluetooth network. Then, the ReadingMate software computes the position of content, so that the position of contents at the same location relative to the head position of the user and presents the text on the display. The iterative process takes place in real time. More details about the ReadingMate can be found in (Kwon, 2010), but ReadingMate is not commercially available.

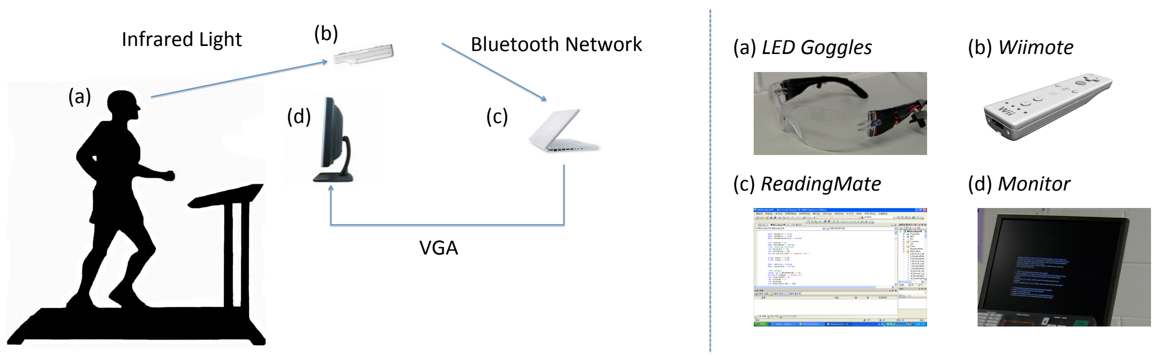


Figure 1. Diagram showing how ReadingMate works: (a) LEDs attached to goggles emit infrared light, (b) a Wii Remote captures head movement, (c) ReadingMate computes the new position of the content, and (d) the position of the content is updated on the display. Adapted from "ReadingMate: The Impact of a Content Stabilization Technique on Reading-While-Running Performance" by B. C. Kwon and J. S. Yi, 2010, Human Factors and Ergonomics Society Annual Meeting Proceedings, 54, p. 647. Copyright 2010 by Human Factors and Ergonomics Society.

In previous studies, we found that ReadingMate provided psychological and physiological benefits (e.g., less perceived fatigue and dizziness) to users (Kwon & Yi, 2009, 2010). Despite such perceived effects, we failed to observe statistically significant improvement in the quantitative reading performances (i.e., reading time and reading comprehension test scores). We conjectured three reasons for such results. First, there might have been other factors. In particular, some participants reported that font size and interline spacing might have been larger factors influencing their reading performances. Second, the reading comprehension tests used in the previous studies (i.e., the GRE® example questions) might not have been suitable for revealing differences in reading performances. These tests could be too burdensome, which could result in degraded reading performances at the ends of experiment sessions. Third, the reading activity may be cognitively burdensome while readers are running on a treadmill. It is known that cognitive load increases and reading performance decreases when a user reads text from a mobile phone while walking (Schildbach & Rukzio, 2010). Running could increase users' cognitive load even more.

Font size and interline spacing are among the major factors that influence reading performance. Many researchers attempt to find a font size where reading speed stops increasing. Critical print size (CPS) refers to the smallest font size that generates maximum reading speed (Chung, Mansfield, & Legge, 1998). The reading speed is believed to decline when font size is far below or above such CPS (Legge & Bigelow, 2011). It was reported that the fastest reading speed is achieved at 11-point on paper (Tinker, 1963). Many other studies also show that font size around 11-point is superior to other sizes. For mobile computers, the range of 8-12 point font sizes was suggested to maximize readability (Darroch, Goodman, Brewster, & Gray, 2005). Interline spacing also has similar trends on reading speed and accuracy. In a series of experiments on letter-detection task, reading speed and accuracy increased as font size and interline spacing increased (Lee, Ko, Shen, & Chao, 2011; Lee, Shieh, Jeng, & Shen, 2008; Van Overschelde & Healy, 2005). However, we could not confirm that such trends were consistent when readers were running on a treadmill, especially with ReadingMate.

In addition, the previous reading comprehension task might not be suitable for our present experiment. Quantitative metrics (i.e., reading time and test scores) might be subject to many other factors, including individual differences in reading comprehension skills and memory skills. Furthermore, even if a standardized test (e.g., the SAT® or GRE®) were used, the level of difficulty across reading passages would be difficult to control. More importantly, reading a passage and answering a subsequent multiple-choice question would take 2 to 3 minutes; therefore, the number of data points collected per participant would be very limited, especially when they were running on a treadmill. These concerns could add unnecessary variation to reading time and test scores.

To minimize such unwanted variation as well as to reduce the pressure on participants, we adopted a variation of the letter-detection task, which has been widely adopted for studying how people process words while reading text (Foucambert & Zuniga, 2011; Roy-Charland, Saint-Aubin, Lawrence, & Klein, 2009; Saint-Aubin, Klein, & Roy-Charland, 2003). In the letter-detection task, users are presented words for a period of time (e.g., 250 milliseconds), and they press a button when they see the letter *f*. This procedure is called rapid serial visual presentation; however, such frequent changes in display might confuse runners. Thus, we modified the letter-detection task for our experiment; participants were asked to count the number of the target letter in a given sentence. This letter-counting task could be less burdensome than reading comprehension tests and less subject to individual differences.

METHODS

Participants

Fifteen participants (18 – 42 years old with an average of 22.3; 6 females) were recruited for this study (Approved IRB#: 0906008227). All 15 participants' first language was English. Thirteen of them were undergraduate and graduate students from diverse departments at Purdue University. The other two participants held bachelor degrees from other 4-year universities. Nine participants reported their vision as 20/30 or 20/20 while the others did not specify their vision because they did not know their vision exactly. No participants reported difficulty reading text in a warm-up session. Although we did not test the visual acuity using a standard test, we pretested participants using the following method. Each subject was asked to stand in the middle of treadmill. Then, we showed a page of sample text with 14-point Arial and asked the subject to read. No participant had problems reading the text while standing still. No one attempted to lean over. Twelve participants ran more than once a week at a speed of 6.0 miles per hour for more than 30 minutes. Three participants ran outside more than once a month. In the warm-up session, there was no noticeable difference in the running performance of outside runners versus treadmill runners. All participants wore proper running attire. Professional and semi-professional sport players were excluded from this study as potential outliers. Only two people had had experience in reading text while running on a treadmill. Ultimately, participants reported they were at higher than 80% optimum mental and physical condition for running (averages: eyes = 95.33%, ankles = 95.93%, breath = 95.87%, heart = 97.33%, mental = 96.53%, and brain = 99.27%). These conditions were measured by participants' response to a survey question asking how much they are ready to run considering 100% as their best condition to run for each of the six categories.

Test Conditions

We conducted a within-subject, split-plot experiment with three main factors: display type, font size, and interline spacing. Display type was a whole-plot factor; whereas font size and interline spacing were subplot factors. There were two levels of the display type factor: normal display (ND) and display with ReadingMate (RMD); ND presented text at a fixed position while RMD dynamically adjusted the text position according to the user's head movements.

We used four levels for font size: 8-, 12-, 16-, and 20-point due to the following reasons: First, we wanted to observe a reading speed trend as font size and interline spacing change. Previous studies have shown that 12-point Arial reads faster than 10-point (Bernard, 2002) but not faster than 14-point (Bernard, Liao, & Mills, 2001). To observe the reading speed trend, we added extreme font sizes, such as 8-, 16-, and 20-point, along with 12-point. There were also four levels of interline spacing: 1×, 1.5×, 2×, and 2.5×. Interline spacing has been defined as the center-to-center distance between two adjacent lines (Bernard, Anne-Catherine, & Eric, 2007). We defined 1× interline spacing as one multiple of font size that had zero vertical space between letters. We expected that these four levels would be disparate enough to observe reading speed trends. Since we had disparate levels of font size and interline spacing, we could not find the exact CPS or interline spacing. However, our aim was to figure out whether CPS under the treadmill running condition existed near 12-point as in the normal reading condition. Figure 2 presents text layouts in the different conditions of font size and interline spacing.

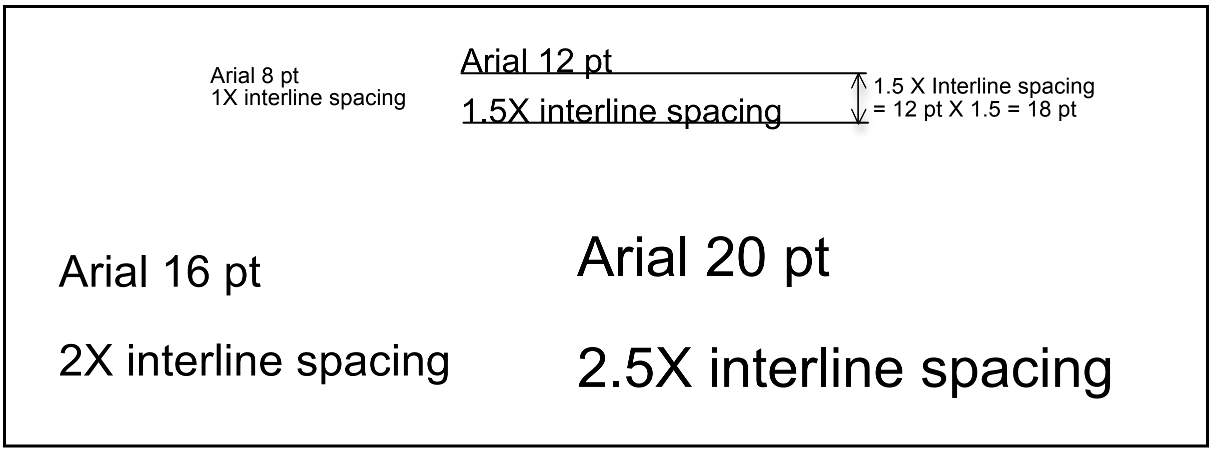


Figure 2. The size of font and interline spacing in the experiment.

In summary, the experiment had three variables in Table 1 and a total of 32 (i.e., 2×4×4) treatments. Each treatment was replicated six to seven times depending on the length of the given reading passage; therefore, each participant performed 205 trials on average. For each trial, a participant performed the letter-counting task while running on a treadmill.

Table 1

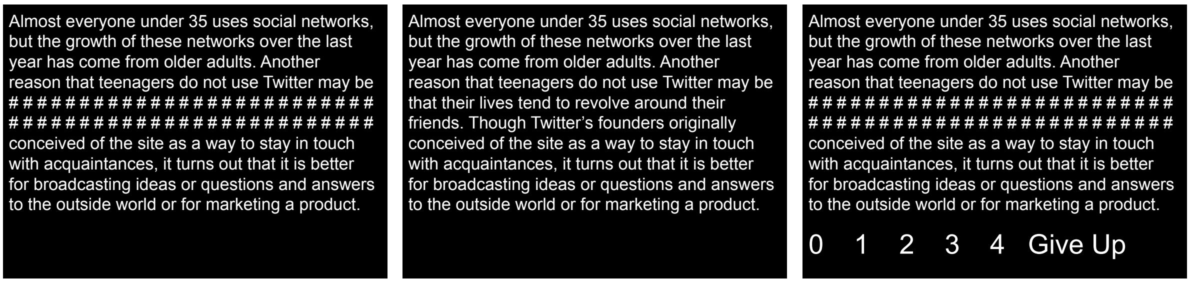
Summary of Independent Variables

Independent Variables	**Categories**
Display Type	Normal Display (ND) and ReadingMate (RM)
Font Size	8-point, 12-point, 16-point, and 20-point
Interline Spacing	1×, 1.5×, 2×, and 2.5×

We did not switch display types between trials because we noticed that participants took time to get accustomed to different display types. Instead, participants were randomly given ND or RMD on one day and the opposite the next day. Each day we randomized the 16 combinations of font size and interline spacing.

Letter-counting Task

The letter-counting task was to count the number of appearances of a target letter in a given sentence. The letter *f* was chosen as the target letter in our experiment. Previous studies reported that the letter *f* was more difficult to detect because it was often included in function words (e.g., "of"; Goldman & Healy, 1985). We assumed that more omissions would occur in less legible conditions. In each trial, 10 lines of text were presented. The task was to count the number of *f*s in the fifth and sixth lines (the target lines). Other lines were used as distractions to test the effect of interline spacing more clearly. In our pilot studies, sentences in the first or last lines were easier to read probably because they were placed at the boundary of a paragraph, working as visual cues for tracing. The target lines of text were pulled from 32 passages of the GRE® reading comprehension test used in the authors' previous studies (Kwon & Yi, 2009, 2010). Other text lines were pulled from random text articles on various news websites (e.g., CNN and BBC).



(a) Before Trial. (b) In Trial. (c) After Trial.

Figure 3. Screen modes (a) before, (b) in, and (c) after a trial.

The two target lines, the fifth and sixth lines, were initially masked with an array of pound signs (i.e., #s), as shown in Figure 3(a). Each participant revealed the words beneath the pound signs by pressing a button on a separate remote controller (not the Wii Remote), which started the trial, as shown in Figure 3(b). After counting the number of letter *f*s in the target lines, the participant finished the trial with another button-press of the remote. Task performance time was measured by computing the duration between the two button-presses (one to begin a trial and another to end the trial). Then, the participant was asked to report the number of *f*s in the trial by choosing one out of five integer options (e.g., 0 through 4) or a sixth “Give-up” option, as shown in Figure 3(c); the five options in each trial were generated to include a right answer plus four random neighboring numbers. To avoid the reporting of random guesses, Participants were instructed to choose the “Give-up” option when it was too difficult to complete the task.

Procedure

On an experiment day, each participant completed the following procedure: Each participant was scheduled for two types of displays on two separate days, respectively. Once the participant arrived at the experiment site, the participant was asked to fill out a back-ground questionnaire. The experimenter examined the participant for appropriate attire, and the participant was asked to warm up. To avoid a learning effect, each participant was asked to run for 2 to 3 minutes at a speed of his or her choosing and to read sample documents of the day's chosen display type during his or her warm-up session. Once the participant was ready, the participant began running and adjusted his or her speed to 6 miles per hour. The task began when the participant pressed a remote controller. While running on the treadmill, the participant read two lines of text and reported the number of letter *f*s appearing in those two lines. After the experiment, the participant was asked to discuss his or her experiences. If the participant completed the first day, then he or she was asked to come back on a subsequently scheduled day to repeat the procedure for a different display type. After the second day, each participant was given $16.00 as compensation.

Equipment

A treadmill, the Smooth 5.45 manufactured by Smooth Inc., was used in this experiment. The dimensions of the treadmill were 1.93 meters long by 0.91 meters wide and 1.52 meters high. The tread belt had a 0.50-meter by 1.40-meter walking surface and bar rails as well as a safety clip. The main computer running the ReadingMate software was an Apple MacBook (MB466LL/A). A Wii Remote and an LCD monitor (a Dell 1908FP with a 1280 × 1024 resolution) were also used. A separate remote control (the Kensington Wireless Presenter with Laser Pointer) was given to participants for interactions such as proceeding to the next trial and choosing an answer.

The experimental site was located in a basement without a window in order to minimize any environmental distraction such as noise or lighting. The luminance of lighting in the room was maintained at approximately 45 lux. The computer monitor was located on bookshelves behind the treadmill. According to each individual's height, the tilt angle of the monitor was adjusted to remove any glare on the display. The LED goggles had no lenses to provide better sight and less irritation from sweat. The average distance between the monitor and runner's head was 0.7075 m (standard deviation = 0.06028). Figure 4 presents the general layout of the equipment.

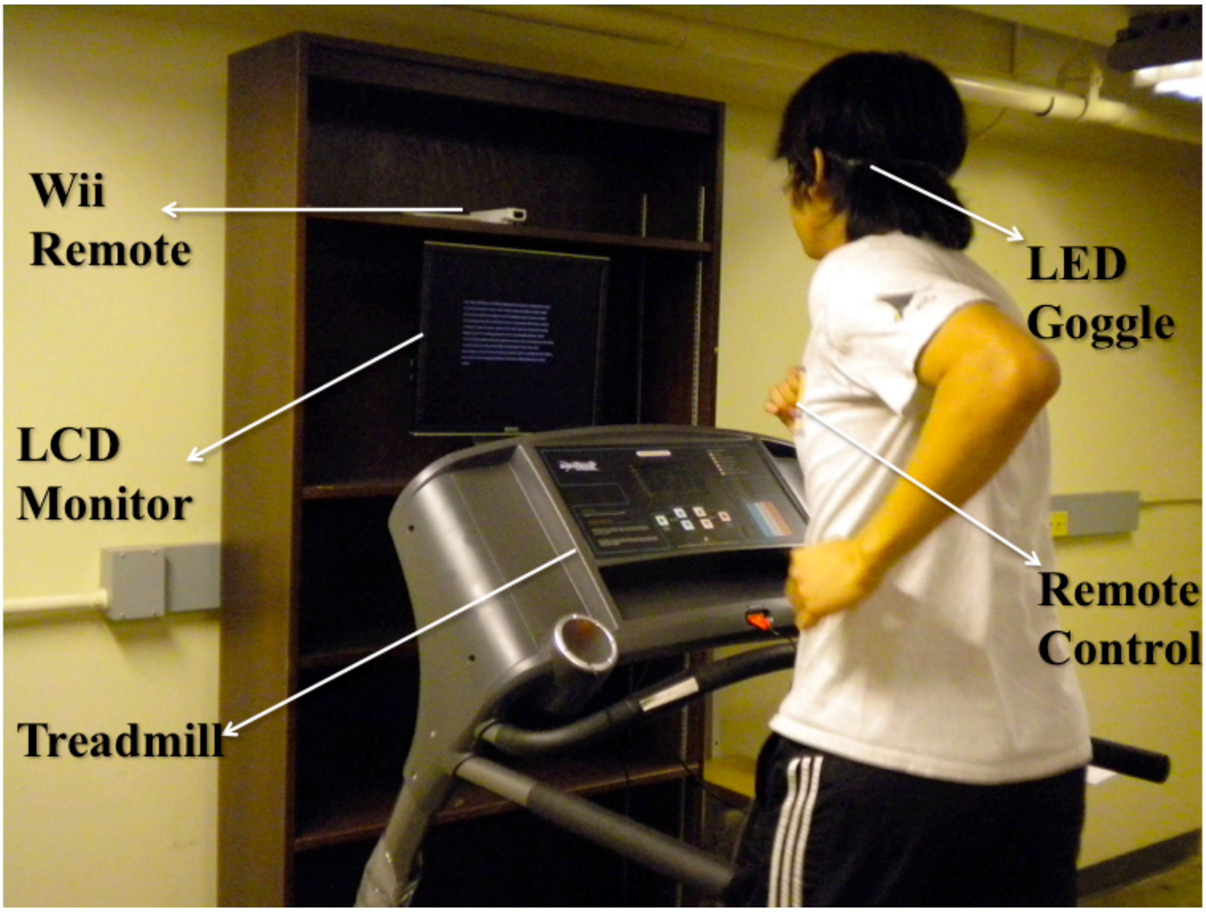


Figure 4. Equipment settings at the experimental site. The runner is holding the remote control with his right hand. Adapted from "A Head Tracking-Based Content Stabilization Technique to Help Runners Read Text While Running on a Treadmill" by B. C. Kwon, 2010, Purdue University, p. 647. Copyright 2010 by B. C. Kwon.

Measurements

We collected three quantitative measurements during each trial: (1) task performance time (time per word elapsed during a trial), (2) counts of successful letter *f* search, and (3) counts of "Give-up." In addition, survey responses and interview results were collected after each day's experiments.

Data Analysis

For accuracy (i.e., the probability of successfully finding *f*s) and give-up counts (i.e., the probability of giving up), we conducted logistic regression analysis. The main effects of display type, font size, and interline spacing and their interaction effects were considered fixed whereas the effects of the participants and the interaction between the participants and the display types were considered random. In particular, Type III tests were used. Once the fitted logistic regression model was obtained, we calculated and plotted the probabilities of successfully counting *f*s for the 32 treatments and further compared the treatments in pairs using odds ratios. Analysis of Variance (ANOVA) and the Tukey HSD test (for pairwise comparison) were employed for analyzing task performance time.

RESULTS

When we analyzed the data, we realized that the structure of the data was unique due to give-ups. Since give-up was an option a participant could use to opt-out of a given trial, it affected the outcome, such as accuracy in counting letter *f*s and task performance time, significantly; therefore, we decided to exclude give-up cases (N = 406) from all cases (N = 3292) in the analysis of task performance time and accuracy and to analyze give-up cases separately. In addition, because over-counting errors (N = 88) were relatively rare and different from under-counting errors, we excluded over-counting cases when analyzing accuracy.

Give-up. Give-up cases tended to decrease as font size and interline spacing increased, but this tendency did not occur in all cases. The probability of having give-up cases was affected by font size, $F(3,3239) = 57.66$, $p < .0001$, and by interline spacing, $F(3,3239) = 32.03$, $p < .0001$. Pairwise comparison using odds ratios (Table 2) revealed that three pairs were significant, as denoted with asterisks. The probability for 8-point font size ($\pi = 0.4783$)

was higher than that for 20-point font size (π = 0.1338). The probability for 1× interline spacing (π = 0.3811) was higher than that for 1.5× interline spacing (π = 0.2327) and for 2.5× interline spacing (π = 0.1673).

Table 2

Odds Ratios of the Main Effects for Font Size and Interline Spacing in Give-up Cases

Treatments in Comparison	**Odds Ratios**	**95% Confidence Intervals**
12 point vs. 8 point	0.0010	0.0001 – 9.9999
16 point vs. 8 point	0.0001	0.0001 – 9.9999
20 point vs. 8 point*	0.0074	0.0036 – 0.0154
16 point vs. 12 point	0.2627	0.0001 – 9.9999
20 point vs. 12 point	7.2993	0.0001 – 9.9999
20 point vs. 16 point	27.778	0.0001 – 9.9999
1.5× vs. 1.0×*	0.1073	0.0537 – 0.2146
2.0× vs. 1.0×	0.0001	0.0001 – 9.9999
2.5× vs. 1.0×*	0.0545	0.0286 – 0.1035
2.0× vs. 1.5×	0.0001	0.0001 – 9.9999
2.5× vs. 1.5×	0.5074	0.2251 – 1.1442
2.5× vs. 2.0×	351.83	0.0001 – 9.9999

* Significant difference between two levels at the error rate of 0.05.

The probability was also affected by the interaction effect of font size and interline spacing, $F(9,3239) = 4.97$, $p < .0001$. Table 3 shows that the probability of 8-point font with 2.0× interline spacing (π = 0.3159) and of 8-point font with 2.5× interline spacing (π = 0.3164) were significantly lower than that of 12-point font with 1× interline spacing (π = 0.4009).

Table 3

Odd Ratios of Interaction Effects for Font Size and Interline Spacing

Treatments in Comparison	**Odds Ratios**	**95% Confidence Intervals**
8 point, 2.0× vs. 12 point, 1.0×*	0.4880	0.2732 – 0.8717
8 point, 2.5× vs. 12 point, 1.0×*	0.2746	0.1482 – 0.5088

* Significant difference between two levels at the error rate of 0.05.

Figure 5 shows the probabilities of having give-up cases with 95% confidence intervals. The trend shows that give-up cases decreased as font size and interline spacing increased; however, the gap between different font sizes decreased as interline spacing increased. The graph shows three conditions that had probabilities of give-up higher than 0.2 (i.e., 8-point font and 1× interline spacing, 12-point font and 1× interline spacing, and 8-point font and 1.5× interline spacing).

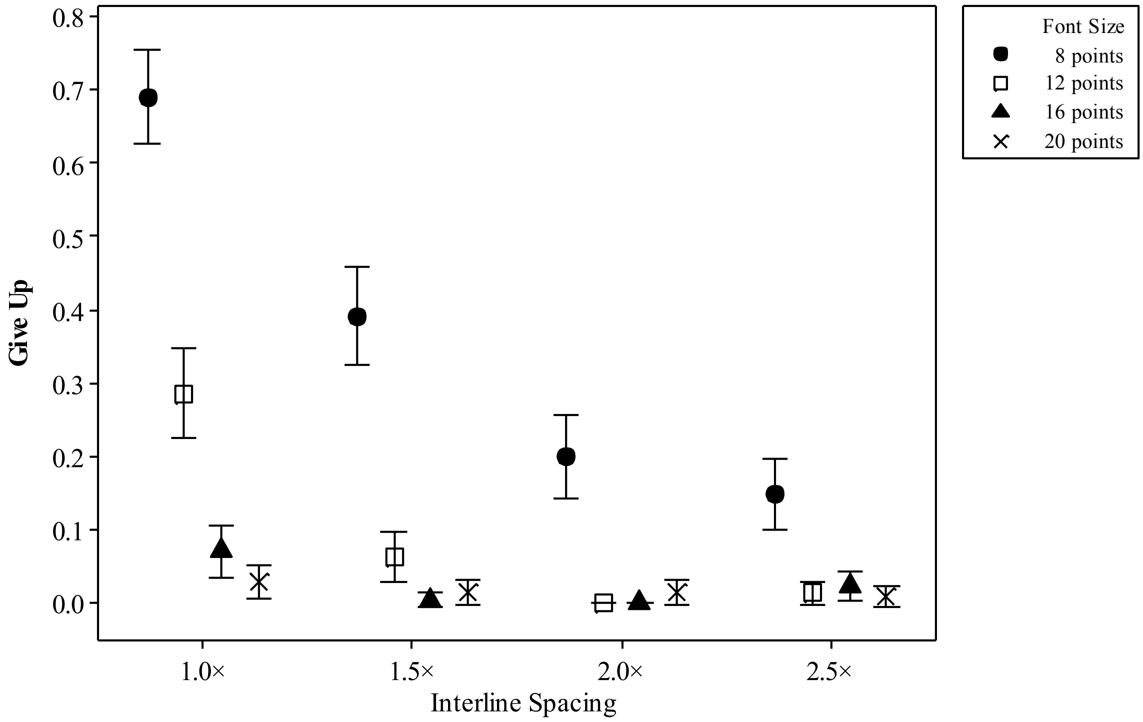


Figure 5. Probabilities of having give-up with 95% confidence intervals for 16 combinations of font size and interline spacing.

Accuracy. The type III tests revealed that the accuracy of the letter-counting task was significantly affected by the main effects of display type, $F(1, 2745) = 5.67$, $p < .0003$ and font size $F(3, 2745) = 4.66$, $p < .0001$. The main effects of interline spacing were not found to be significant $F(3, 2745) = 0.76$, $p = 0.5178$, and none of the interaction effects were found to be significant.

Table 4 shows the 95% confidence intervals of odds ratio for the main effects of display type and font size. The probability of successfully counting *f*s tended to increase as font size increased (8-point ≈ 12-point < 16-point); however, the probability tended to stabilize or slightly decrease as font size reached 20-point (8-, 12-, and 16-point ≈ 20-point). The probability was higher for RMD than for ND (RMD > ND).

Table 4

Odds Ratios of Main Effects

Treatments in Comparison	Odds Ratios	95% Confidence Intervals
RMD vs. ND*	1.3966	1.0351 – 1.8832
12 point vs. 8 point	1.1751	0.8857 – 1.5600
16 point vs. 8 point*	1.6367	1.2315 – 2.1786
20 point vs. 8 point	1.3245	1.0000 – 1.7544
16 point vs. 12 point*	1.3928	1.1062 – 1.7544
20 point vs. 12 point	1.1274	0.8985 – 1.4124
20 point vs. 16 point	0.8084	0.6447 – 1.0132

* Significant difference between two levels at the error rate of 0.05.

To test the hypotheses that RMD could be more beneficial for certain font sizes and interline spacings than ND, we computed the odds ratios between RMD and ND at multiple font size and interline spacing levels. RMD demonstrated greater benefits than ND at 8- and 16-point font sizes and 1.0× and 1.5× interline spacings (see Table 5). In particular, we found the positive effects of RMD on letter-counting performance in dense text conditions such as 8-point and 1.0×, 8-point and 1.5×, and 8-point and 2.5×. The positive effects of RMD were present at 16-point and 1.0× as well. There was no difference between RMD and ND in other text conditions.

Table 5

Odds Ratios of Multiple Combinations of Font Size and Interline Spacing between RMD and ND

Treatments in Comparison	Odds Ratios	95% Confidence Intervals
RMD, 8 point vs. ND, 8 point*	2.0407	1.2914 – 3.2247
RMD, 12 point vs. ND, 12 point	1.1358	0.7696 – 1.6762
RMD, 16 point vs. ND, 16 point*	1.4857	1.0052 – 2.1960
RMD, 20 point vs. ND, 20 point	1.1058	0.7558 – 1.6181
RMD, 1.0× vs. ND, 1.0×*	1.7422	1.1296 – 2.6870
RMD, 1.5× vs. ND, 1.5×*	1.5268	1.0203 – 2.2849
RMD, 2.0× vs. ND, 2.0×	1.2298	0.8367 – 1.8077
RMD, 2.5× vs. ND, 2.5×	1.1640	0.7907 – 1.7137

* Significant difference between two levels at the error rate of 0.05.

Figure 6 shows the 95% confidence intervals of the probabilities of successfully finding *f*s (i.e., accuracy) in all 32 settings of display type, font size and interline spacing. The accuracy of ND demonstrated large variance, especially for small font size (i.e., 8-point). On the other hand, the variance of the accuracy of RMD was consistent across 16 combinations of font size and interline spacing. In particular, the probability of successfully counting *f*s tended to be higher for RMD than for ND.

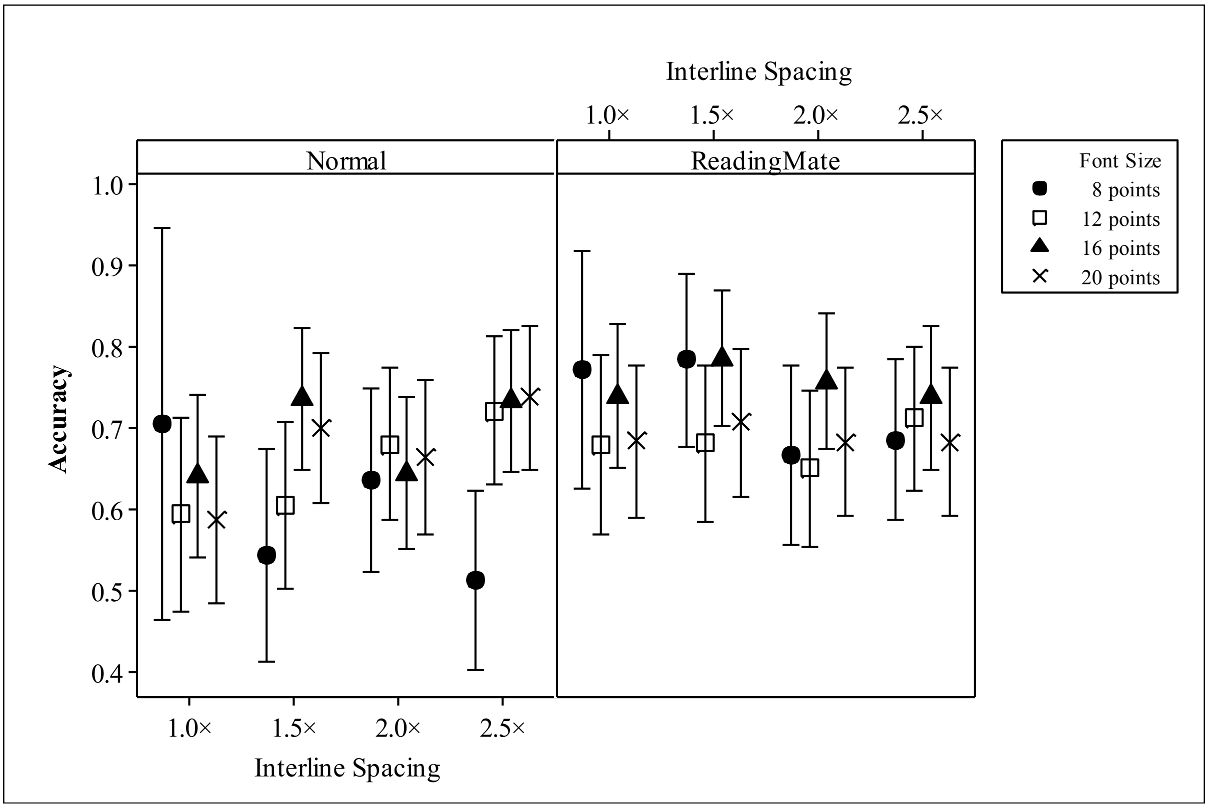
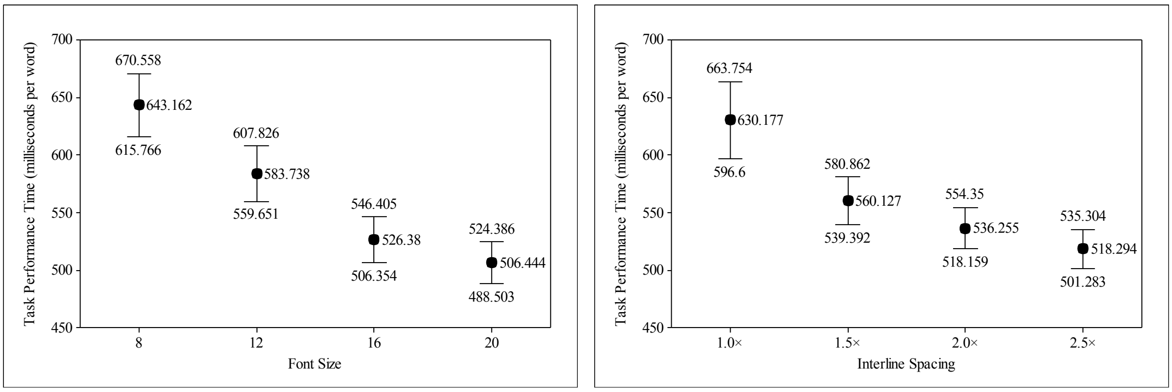


Figure 6 Accuracy with 95% confidence intervals in 32 combinations.

Task Performance Time. There were significant differences in the main effects of font size, $F(3, 2863) = 27.61$, $p < .0001$, and interline spacing, $F(3, 2863) = 17.72$, $p < .0001$. Consistent with our results regarding the probability of finding *f*s, the Tukey HSD test results (Figure 7) showed that task performance time decreased and stabilized as font size increased (8-point > 12-point > 16-point ≈ 20-point) and interline spacing increased (1× > 1.5× ≈ 2× ≈ 2.5×). There was no significant difference in task performance time between RMD and ND.

There did not exist significant interaction effects on task performance time among the three factors.



(a) Task Performance Time versus Font Size. (b) Task Performance Time versus Interline Spacing.

Figure 7. 95% confidence interval of task performance time (i.e., milliseconds per words) by (a) font size and (b) interline spacing.

DISCUSSION

We find the text conditions that are unsuitable for displays in front of runners: 8-point font and 1× interline spacing, 8-point font and 1.5× interline spacing, and 12-point font and 1× interline spacing. We find that the participants are likely to give up more than 20% of the time under these conditions. Even if they do not give up, they tend to take a longer time to read, resulting in highly inconsistent performance in identifying letters. These text conditions should be avoided if possible because they seem to be illegible for many runners. The results also show that give-up is not significantly affected by the main effect of display type. Unless font size and interline spacing are legible enough for runners, ReadingMate cannot be helpful in recognizing letters.

For text conditions under which participants do not give up, ReadingMate can improve letter-counting performance. In particular, we find significant effects of ReadingMate in accuracy in the following text conditions: 8- and 16-point font as well as 1× and 1.5× interline spacings. Though task performance time is not significantly different between RMD and ND, we observe higher probability of successfully counting *f*s when participants are using ReadingMate. This trend stands out in small font size and dense interline spacing, such as 8-point font and 1× interline spacing. According to the interviews with participants, the benefits of ReadingMate included “staying on the line,” “making posture relaxed,” “less blur,” and “reducing the text shake.” On the other hand, participants offered the following comments on their experiences with a normal display: “blurry/blended text lines,” “losing track,” and “accidentally skipping words.” ReadingMate tends to help participants recognize letters in certain text conditions.

We find a trend suggesting that letter-counting performance increases as font size and interline spacing increase. We observe that the probability of give-up decreases as font size and interline spacing increase. Task performance time also decreases and stabilizes as font size and interline spacing increase. This trend implies that CPS might also exist in the reading-while-running context. We also estimate that CPS might exist at font sizes larger than the CPS for sedentary reading (i.e., 12-point) because we notice the stabilization trend between 16- and 20-point in task performance time and letter-counting accuracy.

The letter-counting task proves suitable for this study. The accuracy of finding the letter *f* reveals the effects of ReadingMate. Another interesting measurement of our study is the count of give-up cases. By allowing participants to give up, we observe the perceived difficulty. Furthermore, give-up cases reveal text conditions in which participants cannot read; however, this measurement introduces unexpected variation in other measurements (i.e., in task performance time and accuracy). Such unwanted variation needs to be removed in order to show a legitimate trend in the measurements. With some additional follow-up studies, we believe the letter-counting task could be used as an alternative experimental method to measure reading performances.

The results of this study could guide one to design an interface for reading text when readers are running or under turbulence. First, we suggest avoiding small font and interline spacing, such as 8-point font size and 1× interline spacing. Second, ReadingMate can be used to improve reading performance. In particular, ReadingMate shows improvements in accuracy when font size is small (i.e., 8-point) and interline spacing is dense (i.e., 1×). There are clear tradeoffs between how much text one can put on one screen and how efficiently one can read. When one needs to display a large amount of text on one screen, ReadingMate can be helpful in maintaining reading performance.

There are some limitations. Though we pretested participants' vision, we did not measure their visual acuity, which could affect the performance on the letter-counting task. In addition, we used the letter-counting task instead of asking participants to read text articles. Despite the merits of this task, the results of this study cannot be directly extended to the reading comprehension performance.

CONCLUSIONS AND FUTURE WORK

From this experiment, we find that small font and interline spacing are not suitable for runners because they cannot read while running on a treadmill; however, ReadingMate can improve letter-counting performance. Specifically, task performance time decreases significantly as font size and/or interline spacing increases. We also find that the letter-counting task might be a suitable method for evaluating the reading performance of runners. These findings could be a stepping-stone for further investigation of ReadingMate.

Some future work remains. The effects of other factors—such as paragraph margins and contrast of text—are still unknown. One might also more comprehensively investigate

CPS in the context of reading-while-running. We also hope to investigate reading performance in realistic environments, such as in military (e.g., High-Mobility Multipurpose Wheeled Vehicles) and construction.

ACKNOWLEDGEMENTS

The authors are especially grateful to Dr. Zygmunt Pizlo for his critical advice during preliminary investigations and to Yuming Zhang who assisted post-experiment data analysis.

KEY POINTS

- When a user reads text while running on a treadmill, the reading performance increases as font size and interline spacing increase.
- ReadingMate improves reading-while-running performance by helping runners recognize letters when font size and interline spacing are legible.

REFERENCES

Bernard, J.-B., Anne-Catherine, S., & Eric, C. (2007). Page mode reading with simulated scotomas: A modest effect of interline spacing on reading speed. *Vision Research*, *47*(28), 3447–3459.

Bernard, M. (2002). Examining user expectations for the location of common e-commerce web objects. *Usability News*, *4*(1).

Bernard, M. L., Chaparro, B. S., Mills, M. M., & Halcomb, C. G. (2003). Comparing the effects of text size and format on the readibility of computer-displayed Times New Roman and Arial text. *International Journal of Human-Computer Studies*, *59*(6), 823–835.

Bernard, Michael, Liao, C. H., & Mills, M. (2001). The effects of font type and size on the legibility and reading time of online text by older adults. In *Extended Abstracts on Human Factors in Computing Systems* (pp. 175–176).

Chung, S. T. L., Mansfield, J. S., & Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Research*, *38*(19), 2949–2962.

Darroch, I., Goodman, J., Brewster, S., & Gray, P. (2005). The Effect of Age and Font Size on Reading Text on Handheld Computers. In M. Costabile & F. Paternò (Eds.), *Human-Computer Interaction - INTERACT 2005* (Vol. 3585, pp. 253–266).

Foucambert, D., & Zuniga, M. (2011). Effects of Grammatical Categories on Letter Detection in Continuous Text. *Journal of Psycholinguistic Research*, *41*(1), 33–49.

Goldman, H. B., & Healy, A. F. (1985). Detection errors in a task with articulatory suppression: Phonological receding and reading. *Memory & Cognition*, *13*(5), 463–468.

Grahame, M., Laberge, J., & Scialfa, C. T. (2004). Age Differences in Search of Web Pages: The Effects of Link Size, Link Number, and Clutter. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(3), 385–398.

Kwon, B. C. (2010). *Readingmate: A head tracking-based content stabilization technique to help runners read text while running on a treadmill*. Dissertation Purdue University.

Kwon, B. C., & Yi, J. S. (2009). ReadingMate: an infrared-camera-based content stabilization technique to help joggers read while running on a treadmill. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* (pp. 3449–3454).

Kwon, B. C., & Yi, J. S. (2010). ReadingMate: The Impact of a Content Stabilization Technique on Reading-While-Running Performance. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *54*, 647–651.

Lee, D.-S., Ko, Y.-H., Shen, I.-H., & Chao, C.-Y. (2011). Effect of light source, ambient illumination, character size and interline spacing on visual performance and visual fatigue with electronic paper displays. *Displays*, *32*(1), 1–7.

Lee, D.-S., Shieh, K.-K., Jeng, S.-C., & Shen, I.-H. (2008). Effect of character size and lighting on legibility of electronic papers. *Displays*, *29*(1), 10–17.

Legge, G. E., & Bigelow, C. A. (2011). Does print size matter for reading? A review of findings from vision science and typography. *Journal of Vision*, *11*(5).

Ling, J., & Van Schaik, P. (2007). The influence of line spacing and text alignment on visual search of web pages. *Displays*, *28*(2), 60–67.

Roy-Charland, A., Saint-Aubin, J., Lawrence, M. A., & Klein, R. M. (2009). Solving the chicken-and-egg problem of letter detection and fixation duration in reading. *Attention, Perception & Psychophysics*, *71*(7), 1553.

Saint-Aubin, J., Klein, R. M., & Roy-Charland, A. (2003). Direct Assessments of the Processing Time Hypothesis for the Missing-Letter Effect. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1191–1210.

Schildbach, B., & Rukzio, E. (2010). Investigating selection and reading performance on a mobile phone while walking. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (pp. 93–102).

Sheedy, J. E., Subbaram, M. V., Zimmerman, A. B., & Hayes, J. R. (2005). Text Legibility and the Letter Superiority Effect. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(4), 797–815.

Tinker, M. A. (1963). *Legibility of print*. Iowa State University Press.

Van Overschelde, J. P., & Healy, A. F. (2005). A blank look in reading: the effect of blank space on the identification of letters and words during reading. *Experimental Psychology*, *52*(3), 213–223.

.