Paragraph Text Reading Using a Pixelized Prosthetic Vision Simulator: Parameter Dependence and Task Learning in Free-Viewing Conditions

Gislin Dagnelie,1 David Barnett,1 Mark S. Humayun,2 and Robert W. Thompson, Jr1

PURPOSE. To investigate the feasibility of adequate reading by recipients of future prosthetic visual implants through simulation in sighted observers.

METHODS. Four normally sighted subjects used a video headset to view short-story segments at a sixth grade reading level, presented in 6- to 11-word paragraphs through a pixelizing grid defined by five parameters (dot size, grid size, dot spacing, random dropout percentage, and gray-scale resolution). Grid parameters were varied individually, and four character sizes and two contrast levels were used.

RESULTS. Reading speeds of 30 to 60 words per minute without errors were recorded for some parameter combinations. In general, reading accuracy and speed were influenced by all parameters. Reading accuracy exceeded 90% if the following conditions were met: At least 3 dots/charwidth were presented, and dropout did not exceed 50%. Reading speed deteriorated below 20 words per minute if accuracy fell below 90% and at low contrast if the grid spanned less than two characters.

CONCLUSIONS. It is uncertain whether and to what extent retinal reorganization may limit the perception of multiple phosphenes by blind prosthesis recipients. If distinct phosphenes can be perceived, these results suggest that a 3 × 3-mm² prosthesis with 16 × 16 electrodes should allow paragraph reading. The effects of stabilizing the dot grid on the retina must be investigated further. (Invest Ophtalmol Vis Sci. 2006; 47:1241-1250) DOI:10.1167/iovs.05-0157

Over the past decade, a growing number of research groups has been reporting slow but steady progress toward the realization of implantable visual prostheses for the blind.1-3 Recently, both subretinal microphotodiode arrays4 and externally driven epiretinal electrode arrays5 have been implanted in volunteers with retinitis pigmentosa who are legally or functionally blind, in U.S. Food and Drug Administration (FDA)-approved phase I clinical trial protocols. Whether recipients of future prosthetic visual implants through simulation (FDA)-approved phase I clinical trial protocols. Whether

...
letter and word recognition at the same eccentricities, in a rectangular (20° × 7° or 10° × 3.5°) area along the lower vertical meridian and reported a decrease in accuracy at and below 6 dots/charwidth, regardless of eccentricity, and a significant practice effect at 15° eccentricity. A further study by Sommerhalder et al.16 demonstrated similar practice effects for full-page text reading at peripheral locations. The direction of gaze was used to scan a 10° × 7° reading window across the text, which consisted of a proportionally spaced font with 1.8° lowercase height, pixelized at approximately 5 pixels/char. In the latest study from the same laboratory,17 the authors varied the dot profile, using circular Gaussian and square profiles, reasoning that the use of square pixels introduced edges with potentially high contrast, which would not occur between neighboring phosphenes. They also used a real-time filter, allowing the dots to move with the viewing window, but only did so for the square profile. They did not find a difference according to dot profile, but because they did not use the most natural condition (a Gaussian filter moving with the viewing window), they may have underestimated the benefit of smooth versus square profile dots. Most important, however, they presented text as black letters on a light background (i.e., as missing dots). Prosthesis designers (and presumably users) would prefer light text on a dark background, because this avoids potential problems of glare and stimulation overload, raises the effective contrast, and reduces implant power consumption.

There are additional ways in which these simulations were not representative of the way a retinal prosthetic wearer would read. In their initial approach Bagboud and Sommerhalder stabilized the pixelized letters and words on the retina, preventing the subjects from scanning the text (as a blind prosthetic wearer could do with a video camera). Moreover, the use of square pixels may have impaired the detection and integration of large scale (letter and word) features across pixels.18,19 Finally, reading in eccentric locations beyond 10° adds an unwarranted degree of difficulty20. The only chronic retinal implants currently in use are located within the macula.5

In preparing pixelized reading stimuli, the choice of character size plays an important role. If characters are small, too few dots are available to define the character shape. If they are too large, only a few characters may fit inside the grid of dots. Studies of reading with spatially sampled text are relatively scarce, but there is a larger body of literature regarding reading performance with spatially filtered text. According to the theory of Shannon,21 a sampled image can convey all information contained in the spatially filtered image, if the number of dots per degree is at least twice the filter cutoff frequency. The validity of the sampling theorem for human visual perception has been demonstrated near the resolution limit of the photoreceptor mosaic.22 Solomon and Pelli23 demonstrated that reading is primarily mediated by spatial filters centered around 3 cycles per character width (c/charwidth), and Legge et al.24,25 used ground-glass filters with approximately second-order filter characteristics26 to demonstrate, in both normally sighted and low-vision observers, that reading performance remains close to optimal if such a low-pass filter extends to at least 2 c/charwidth, regardless of character size. For pixelized text, however, Legge et al.25 found that minimum dot density increases with the logarithm of character size. The value of four samples/charwidth predicted by the sampling theorem only holds near the visual acuity limit. Legge et al.25 also found that reading speed is optimal if the text window contains at least four to five characters, whereas a later study showed that up to seven characters may be necessary for low-vision readers.27

Some properties and operational aspects of retinal implants were not available in these studies. A visual prosthetic wearer may pan a head-mounted or handheld camera across the text, and optimize text size by changing the working distance, but the prosthesis wearer has no control over the contrast and gray-scale resolution of the phosphenes “image.” Intraoperative tests and results with long-term implants suggest that subjects can distinguish six to eight different stimulation levels on the basis of phosphenes brightness and/or size. Subjects also report seeing these phosphenes on a gray rather than a black background.3,25 Finally, it is not certain that all electrodes will provide phosphenes of the same clarity, due to either electrode location relative to viable target cells in the retina or to safety limits for charge density at the electrode interface.28 In addition, some electrodes or target cells may lose functionality over time.

In contrast with previously published studies, we examined reading with larger-sized yet noncontiguous dots in the center of the visual field and, in addition to dot spacing, grid size, and optimal character size, addressed properties of a prosthetic implant over which only limited control is possible: perceived gap width, dropouts, gray-scale resolution, and effective contrast. We did not address essential properties of electrical stimulation of the degenerated retina, most of which can be investigated only through detailed tests with actual implant recipients: irregular phosphenes shapes reported in some acute tests but not in others, safe charge levels during long-term stimulation, effects of altered neural connectivity in the retina, and phosphenes stabilization due to the anchoring of electrodes to the retina. These simulations are only a simplified representation of prosthetic vision, but an analysis of sighted subjects’ ability to read under these adverse conditions may still be helpful in setting parametric limits for a future retinal prosthesis.

METHODS

Subjects

The volunteer subjects were four college students, two female and two male, 20 to 32 years of age, with best corrected visual acuity in the right eye of at least 20/20. Subjects received a meal coupon and a small financial remuneration for their participation. Before participation in the study, written informed consent was obtained from all subjects. The research protocol adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of the Johns Hopkins University School of Medicine.

Apparatus

Subjects used only the right eye to view dot displays in a field subtending 36° × 48° at optical infinity inside a video headset (Infrared Eyetracking Low Vision Enhancement System, IVES; Visionics Corp., formerly of Golden Valley, MN), with an equivalent screen luminance of 225 cd/m². The screen has 480 × 640 (VGA) resolution, so each screen pixel subtends 4.5 × 4.5 arcmin. Within this field, large-print text in Arial font was displayed as short paragraphs of up to 20 characters per line, two to three lines per paragraph. Subjects viewed this text through a grid of dots, with 11 grid parameter combinations, as specified in Table 1. The dots acted as “apertures” onto the text image: Text information was shown only within the dots, with each dot containing a single intensity, equal to the mean intensity of the text image across its aperture, quantized to the nearest available gray level (Fig. 1, right). As the dot grid was moved by the subject, the aperture of each dot would shift across the text image; hence, the mean gray level across each dot was recalculated, and the dot intensity adjusted, before the next 17-ms frame. This allowed subjects to inspect the text by scanning the dot grid across the image.

Figure 1 shows the subject wearing the IVES headset (left) and two views of a text fragment image: The filtered text fragment is shown in
TABLE 1. Grid Parameter Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Dot Diameter (Arcmin)</th>
<th>Gap Width (Arcmin)</th>
<th>Grid Count (Dots)</th>
<th>Dropout Percentage (% of Dots)</th>
<th>Gray Scale (Level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard condition</td>
<td>22.5</td>
<td>4.5</td>
<td>16 × 16</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Alt. Dot 1</td>
<td>13.5</td>
<td>4.5</td>
<td>16 × 16</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Alt. Dot 2</td>
<td>40.5</td>
<td>4.5</td>
<td>16 × 16</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Alt. Gap 1</td>
<td>22.5</td>
<td>18</td>
<td>16 × 16</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Alt. Gap 2</td>
<td>22.5</td>
<td>31.5</td>
<td>10 × 10</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Alt. Grid 1</td>
<td>22.5</td>
<td>4.5</td>
<td>25 × 25</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Alt. Grid 2</td>
<td>22.5</td>
<td>4.5</td>
<td>16 × 16</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>Alt. Dropout 1</td>
<td>22.5</td>
<td>4.5</td>
<td>16 × 16</td>
<td>70</td>
<td>2</td>
</tr>
<tr>
<td>Alt. Level 1</td>
<td>22.5</td>
<td>4.5</td>
<td>16 × 16</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>Alt. Level 2</td>
<td>22.5</td>
<td>4.5</td>
<td>16 × 16</td>
<td>30</td>
<td>4</td>
</tr>
</tbody>
</table>

The 10 alternate (Alt.) conditions were made by changing a single parameter (bold) relative to the standard condition.

the middle panel, and the image seen by the subject through the dot grid is shown on the right.

Text images were prepared as follows: We used 24 short text segments of sixth-grade reading material, 150 to 200 words in length, covering different topics. Segments were split into phrases averaging 12 words (50 characters), displayed across two to three lines. Thus, we obtained 384 phrases, presented to preserve the continuity of the original text segments. Subjects never saw the same text segment twice.

To allow real-time scanning of the text, bitmap images of the 384 phrases, centered vertically and horizontally, were low-pass prefiltered (as in Fig. 1, middle) to match the two-dimensional comb filter formed by the dot grid to be used in the trial. The cutoff spatial frequency of this filter is the inverse of the spatial periodicity (two dot-gap pairs) of the grid. This reduced the real-time operation needed to prepare the dot grid for the next video frame, in response to the updated dot positions, to a look-up task in which each dot is filled with the gray-scale value of the pixel at its center in the filtered image, rounded to the nearest gray level allowed (note the example in Fig. 1, right, which has four gray levels). Filter and presentation software were developed (Visual Basic, Microsoft, Redmond, WA) and run on a 400-MHz computer, with a video card (Diamond Stealth VGA; Diamond Multimedia, Chatsworth, CA) with 2 MB of memory for stimulus presentation.

Procedure

Before the initial test session, subjects were allowed to familiarize themselves with the equipment by scanning and reading three short practice sentences, presented with different parameter and character size combinations that have been found to be easy for inexperienced users. In subsequent sessions, subjects practiced by repeating the last few trials of the previous session. At the start of the session, if necessary, the refractive correction in the headset was set to the subject’s average, and gray scale resolution) and character size were varied independently within each block. Character size changed on every trial, cycling through 29, 38, 48, and 58 screen pixels cap height (i.e., sizes from 2.2° to 4.4°); average character widths corresponding to these cap heights were 1.5° to 3°. Grid parameters also changed on every trial, with 11 combinations of five parameters presented in pseudorandom counterbalanced order. We defined a standard condition of 16 × 16 dots, 22.5-arcm screen size, 4.5-arcm gap width, 30% random dropout, and four gray levels and created 10 additional conditions by modifying one parameter relative to this standard. Alternate values for the grid parameters were 13.5 or 40.5 arcmin dot size, 18 or 31.5 arcmin gap width, 10 × 10 or 25 × 25 grid size, 50% or 70% dropout; and two or eight gray levels. The 11-grid parameter combinations and four character sizes result in 44 distinct stimulus conditions at each contrast level. Not all conditions could have an equal number of repetitions across 64 trials, but there was an equal number of trials for each character size, and close to an equal number (five or six) for each of the 11 dot/grid parameter combinations.

Figure 2 shows the same text fragment shown in Figure 1, as seen through the 11 grid parameter combinations, with examples of the three remaining character sizes and a low-contrast image.

Stimulus Parameters

Stimuli were presented in six blocks of 64 trials each, at either high (99%; blocks 1, 2, and 4) or low (12.5%; blocks 3, 5, and 6) contrast. Five grid parameters (dot size, gap width, grid size, dropout percentage, and gray scale resolution) and character size were varied independently within each block. Character size changed on every trial, cycling through 29, 38, 48, and 58 screen pixels cap height (i.e., sizes from 2.2° to 4.4°); average character widths corresponding to these cap heights were 1.5° to 3°. Grid parameters also changed on every trial, with 11 combinations of five parameters presented in pseudorandom counterbalanced order. We defined a standard condition of 16 × 16 dots, 22.5-arcm screen size, 4.5-arcm gap width, 30% random dropout, and four gray levels and created 10 additional conditions by modifying one parameter relative to this standard. Alternate values for the grid parameters were 13.5 or 40.5 arcmin dot size, 18 or 31.5 arcmin gap width, 10 × 10 or 25 × 25 grid size, 50% or 70% dropout; and two or eight gray levels. The 11-grid parameter combinations and four character sizes result in 44 distinct stimulus conditions at each contrast level. Not all conditions could have an equal number of repetitions across 64 trials, but there was an equal number of trials for each character size, and close to an equal number (five or six) for each of the 11 dot/grid parameter combinations.

Analysis

Reading speed (RS) in (correct) words/minute and accuracy in percentage correct (%C) were the direct outcome measures. To obtain

FIGURE 1. The Infrared Eyetracking Low-Vision Enhancement System (IVES: Visionics Corp.) worn by a subject (left) and a text fragment used in the experiment in two processing stages: the prefiltered image used as the input in the real-time presentation (middle) and the pixelized image, with the standard grid parameters and high contrast as seen inside the headset by the subject (right).
RESULTS

Reading Accuracy and Speed as a Function of Grid Parameters

Figure 3 shows parameter dependence of reading speed (number of words read correctly per minute; filled symbols) and accuracy (percentage of words read correctly; open symbols), as a function of the five grid parameters: dot size, gap width, grid size, dropout percentage, and number of gray levels. The left column shows performance averaged across all four character sizes, and the right column shows performance for the optimal character size. Each data point gives mean and SD across the four subjects. Further details are provided in the figure caption.

Data in the left column show that most of the conditions tested allowed reading with greater than 90% accuracy, averaged across character sizes and subjects. The only conditions for which accuracy fell below 90% were the largest dot size (40.5 arcmin), the larger gap widths (18 and 31.5 arcmin), and the highest dropout (70%). At low contrast, accuracy also fell below 90% for 50% dropout and the 10 × 10 grid size. In other words, 7 of 11 conditions tested allowed 90% accurate reading at high contrast, and 5 of 11 at low contrast. If the same 90% accuracy criterion is applied to the data in the right column (i.e., reading at optimum character size) only the 70% dropout condition (high and low contrast) and the 31.5-arcmin gap at high contrast yielded accuracies below 90%; most other conditions yielded accuracies at or near 100% correct.

Reading speeds averaged across character sizes and subjects generally ranged from 20 to 30 words per minute at high contrast for all parameter conditions allowing better than 90% accuracy; speeds were slightly lower at low contrast. Trends for reading speed as a function of grid parameters mirror those observed for accuracy, in general, and in fact accentuate these, falling below 10 words/min as accuracy falls below approximately 80%. Reading was slower at low than at high contrast, especially for those conditions allowing high reading speeds. If the character size is limited to the optimum for each condition (right column), reading tended to be 20% to 50% faster than when averaged across character sizes. Trends were similar otherwise.

Especially if character size was chosen optimally, the single most important parameter affecting reading speed was grid size, with 25 × 25 dots allowing a 67% gain in reading speed over 16 × 16 dots, and reading speeds on the order of 50 words per minute. At high contrast, an increased number of gray levels also allowed a noticeable increase in reading speed, but an opposite effect occurred at low contrast. This difference is statistically significant at optimal character size. Reading speeds at high and low contrast for two gray levels are equal, whereas the difference at eight gray levels is significant (39 and 22 wd/min; \( t = 2.61, P = 0.04 \)). Another effect of contrast on reading speed was noted for the smallest dot size only (top panels): Reading was more than twice as fast at high contrast, and the difference is significant at optimal (\( t = 3.84, P < 0.02 \)) size, as well as across all (\( t = 6.14; P < 0.001 \)) character sizes. We will return to these effects in the Discussion section.

Effect of Character Size

From the comparison of the left and right columns in Figure 3 it is clear that the choice of character size has a noticeable effect on pixelized text reading. To illustrate this effect better, we examined the dependence of reading speed and accuracy on grid parameters that show strong interactions with character size, either because they affect the number of dots spanning the width of an average lowercase character (dots/charwidth) or because they affect the number of characters contained in the grid window. Letter recognition and reading performance are known to depend critically on the number of spatial frequency cycles, and hence dots, per character.\(^{23-25}\) Only dot size and gap width affect the dots/charwidth measure; the number of characters in the viewing window is affected by the same two parameters and by the grid size in dots. Therefore, Figure 4 plots along the abscissa either character size in dots/charwidth, to show the effect of dot and gap changes (top), or the number of characters, to show the effect of changes in grid size (bottom). High- and low-contrast data are shown in the left and right columns, respectively.

As can be seen in Figure 4, top, accuracy for text sizes at or above 4 dots/charwidth was at or near 100%; maximum read-
Inking speeds were reached at 5 to 6 dots/charwidth. With the same 90% correct criterion as before, accurate text reading remained possible down to 3 dots/charwidth, even at low contrast, although reading speeds decreased to approximately half of the maximum speed. Looking at the data in the top left panel, which were all collected with high-contrast text, 16×16 grids, 30% dropout, and four gray levels, for both accuracy and reading speed, the five characteristics lined up to form a smooth envelope, with performance limited by spatial sampling (dots/charwidth) toward the left and by grid size (characters) toward the right. Beyond 5 dots/charwidth, three characters, at most, fit within the window at any given time. The bottom left panel shows how this windowing restriction scaled the reading speed. For the two larger text sizes (5.5 and 6.7 dots/charwidth) reading speed was close to proportional with the number of letters in the dot window, as indicated by the dashed diagonal regression line forced through the origin. The decline in performance with a larger number of characters, first in reading speed and then in accuracy is caused by the decrease in text size (in dots/charwidth).

The sampling and windowing limitations are more obvious for low-contrast stimuli (Fig. 4, right). For the coarsest grid (filled squares), reading speeds were similar at low and high contrast, but as grid pitch decreased, reading speeds were reduced, so the smooth envelopes shown in Figure 4, top left, are absent in the top right panel. For the finest grid pitch (i.e.,
13.5 ± 4.5 arcmin; filled diamonds), reading speed was depressed at all character sizes, presumably because only 1.6 to 3.2 characters fit in the 16-dot window.

In summary, the top panels of Figure 4 confirm an important expectation: Both accuracy and reading speed declined at character sizes with sparser sampling, as expected on the basis of the sampling theorem. Both top and bottom panels show that the limited extent of the grid restricted reading speed when the text window was confined to less than five characters. The bottom panels show that for the two largest character sizes accuracy was near perfect, and window size constituted the only major limitation of reading speed. Contrast reduction to 12.5% in these conditions reduced reading speed, but only by approximately 20%.

**Practice Effects**

During the initial test session, all subjects indicated frustration at the difficulty of the high-contrast reading task. This complaint gradually subsided as they read the initial two blocks of sentences, resurfaced when the third low-contrast block was presented, and subsided again over time. To examine the response patterns underlying this apparent habituation, we calculated accuracy and reading speed for groups of 22 trials (i.e., two trials at each of 11 parameter settings), taken from each of the six blocks of sentences. Because of the four different character sizes used, character sizes are represented unequally in each group of 22, but this is balanced in the next group. The reason to consider groups of 22 rather than 44 trials is the desire to see practice effects that may occur over a time scale shorter than 44 trials. Practice trends were tested statistically by using the reading efficiency (geometric mean of accuracy and reading speed) as a compound dependent variable, as previously stated.

Figure 5, top, shows the progression of 22-trial group averages across time. Error bars represent the average of between-subjects standard deviations in individual trials. Because of the small number of words per trial, these standard deviations can be substantial. Over the first two blocks of high-contrast trials (groups 1a through 2c) both accuracy and reading speed increased substantially. Improvements early within block 1 (t = 6.06) and between blocks 1 and 2 (t = 8.14) are highly significant (P < 0.001). The last block of high-contrast trials (groups 4a, 4b, and 4c) did not show further improvement over block 2 (t = −0.09; NS). The first low-contrast block (groups 3a, 3b, 3c) shows a significant (t = −4.06; P < 0.001) decrease in accuracy relative to high-contrast block 2. Low-contrast performance remained stationary through two blocks (t = 0.82; NS, between blocks 3 and 5, no significant improvement within either block), but then improved substantially.
between blocks 5 and 6 ($t = 7.29; P < 0.001$), whereas there was no further improvement during the last block (groups 6a, 6b, and 6c; $t = 0.54; -0.92$; NS). Both the figure and the statistical analysis demonstrate that the effects of practice in low-contrast conditions are delayed relative to those in high-contrast conditions.

As was shown in Figure 3, performance levels substantially depended on the parameter combination used. It is plausible that practice effects occurred at different rates with different parameter settings. To examine this possibility, separate trend analyses were performed for the three best and the three worst performing parameter combinations, with results shown in Figure 5, bottom. Each data point in these two panels represents 12 trials (3 parameter sets $\times$ 4 character sizes), averaged across subjects. Error bars again represent the average of between-subject standard deviations in individual trials. For the "easy" sets of the high-contrast trials (blocks 1, 2, and 4), accuracy was near perfect from the start, and reading speed reached its asymptote in the second block of trials, whereas the low-contrast trials (blocks 3, 5, and 6) showed improvement in both accuracy and reading speed between the last two blocks. These are the same trends recorded in the top half of the figure, and the statistical significance levels are very similar. For the difficult trials (bottom right), error rates were at least 10 times those of the easy trials (bottom left), and reading speed was roughly one quarter that for the easy trials (note different scales, y-axis). Overall, however, the practice trends were similar to those in the top and bottom left panels. High-contrast performance reached an asymptote by the second block, as confirmed by ANOVA (SAS Proc GLM; SAS Institute, Cary, NC), whereas low-contrast performance improved significantly from block 3 to block 5 ($P = 0.028$) and again in the final block ($P = 0.037$).

Note that, as shown in all three panels, similar accuracy levels were reached after extensive training for high- as for low-contrast stimuli, whereas reading speeds at low contrast stayed approximately 25% below those at high contrast. It appears, then, that practice effects were qualitatively similar for all stimuli and occurred over a relatively quick time frame (i.e., within approximately 100 trials), albeit lower than at high contrast (i.e., over the second rather than the first 100 trials).

**Interindividual Differences**

Interindividual differences in reading performance were substantial, not only in reading speed (where the between-subjects SD ranged from 20% to 60% even for conditions with near perfect accuracy), but also in accuracy for most conditions with appreciable error percentages. As Figure 5, bottom, shows, the intersubject coefficient of variability (CoV; the ratio of error and mean score) is no greater for the hard than for the easy parameter settings, especially after training. In contrast, all three panels of Figure 5 suggest that the CoV decreases with training and/or at low contrast. Although performance differences, even among normally sighted subjects, are to be expected, it is important to know whether they vary across conditions or, instead, operate primarily as a scaling factor between individual subjects.

To study the intersubject difference in performance in a single ANOVA, we reduced the two dependent variables, reading speed and accuracy, to a single compound measure, similar to the identification index used in the analysis of face identification data under pixelized vision conditions. In choosing this measure, we wanted to minimize the effect of any tradeoff of reading speed and accuracy that might affect the results of practiced subjects as they became less careful, in a desire to complete the experiment. Such speed–accuracy tradeoffs are well known in spatial tasks such as pointing and are adequately described by Fitts’s law. Unfortunately, the temporal speed–accuracy tradeoff has received much less attention, and there is no generally accepted model similar to Fitts’s law. Using reading efficiency (RE), defined as the geometric mean $\sqrt[3]{(C \times RS)}$, as the dependent measure, we not only minimized the effects of such trade-offs but also normalized the distribution of the dependent measure (skewness $= 0.13$; reduced kurtosis $= -0.86$).

Subject, grid parameter condition, contrast, and block number were independent variables in the ANOVA. The analysis was performed twice (i.e., for all text sizes combined and for the optimal text size only). All interactions of these variables were initially built into the model, and nonsignificant effects were sequentially eliminated until only the effects remained for which $P < 0.01$. In the two analyses, these effects accounted for 55% ($n = 1536$, $F = 31.2$) and 72% ($n = 384$; $F = 15.3$) of the variance, respectively. The relative importance of the significant effects was very similar in both analyses. For this reason, only the results of the analysis for the optimal text size are shown in Table 2. The third column of this table shows the significant effects. Most of the explained variance is associated with single variables (62%) and two-way interactions (20%). The two- and three-way interactions indicate that besides the overall performance difference between subjects there is only one parameter that contributed significantly to intersubject variability (i.e., contrast level). In particular, this means that practice effects (block) and grid parameter effects were substantially the same in all four subjects.

**DISCUSSION**

The results presented herein indicate that reading with pixelizing masks is not just possible but accurate under a wide range of grid conditions and text sizes. In a limited set of conditions,

### Table 2. ANOVA Results for Reading Efficiency

<table>
<thead>
<tr>
<th>Parameter</th>
<th>df</th>
<th>Type 3 SumSq</th>
<th>Type 3 MeanSq</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid parameters</td>
<td>10</td>
<td>125.69</td>
<td>12.57</td>
<td>13.60</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Subject</td>
<td>3</td>
<td>41.21</td>
<td>13.74</td>
<td>14.86</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Block</td>
<td>1</td>
<td>18.42</td>
<td>18.42</td>
<td>19.93</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Contrast</td>
<td>1</td>
<td>21.16</td>
<td>21.16</td>
<td>22.89</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Contrast parameters</td>
<td>10</td>
<td>28.45</td>
<td>2.84</td>
<td>3.08</td>
<td>0.0009</td>
</tr>
<tr>
<td>Contrast subject</td>
<td>3</td>
<td>23.10</td>
<td>7.70</td>
<td>8.33</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Block subject</td>
<td>3</td>
<td>16.25</td>
<td>5.41</td>
<td>5.86</td>
<td>0.0007</td>
</tr>
<tr>
<td>Block contrast parameters</td>
<td>21</td>
<td>62.10</td>
<td>2.96</td>
<td>3.20</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Reading efficiency, $\sqrt[3]{(\text{accuracy} \times \text{reading speed})}$, was the dependent variable of subject, grid parameter combination, contrast and block number, and their significant ($P < 0.01$) interactions. Data are for 384 trials from each of the four subjects, with optimal text size only ($F = 15.3$, $R^2 = 0.716$).
reading speeds similar to those of patients with low vision can be achieved (30–60 words/min). With sufficient practice these performance levels are also achieved at low contrast (12.5%). Reading performance in this study was generally consistent with that reported by other investigators using pixelized and spatially filtered text, yet there are a few discrepancies.

Effects of Contrast and Gray-Scale Resolution
To limit the number of different parameter sets to be studied, we used only two contrast levels in this study. The choice of contrast levels, 100% and 12.5%, was not accidental. The 100% contrast represents an idealized situation, in which phosphophones stand out brightly against a dark background, whereas 12.5% is well below the contrast qualitatively described by subjects during both intraoperative and chronic implant testing at suprathreshold stimulus levels. Thus, testing normally sighted subjects with contrast levels closer to threshold would not have served a useful purpose. The reason for using a lower contrast level than that reported for single phosphophones is the expectation that contrast may decrease or phosphophones may fuse if multiple electrodes are activated simultaneously at charge levels well above threshold (Humayun MS, personal communication, September 2004; also Rizzo et al.28). In general, data in Figures 5, 4, and 3 show strong similarities between high- and low-contrast reading performance. As long as phosphophones are readily discernible against the background, the results presented herein should remain valid.

Figure 3 provides two notable examples of differences between high- and low-contrast reading performance. Figure 3, top, shows that reading speed at the smallest dot size was strikingly slower at low than at high contrast. This can be understood by realizing that the 16 × 16 grid in this condition spans just 4.8° (i.e., very similar to the 10 × 10 grid in the center two panels), for which reading speeds very similar to the low-contrast, small-dot condition are shown. In other words, it is the high reading speed achieved with small dots at high contrast that is the exception. A second discrepancy between high- and low-contrast results can be seen as a function of gray-scale resolution in the bottom panels of Figure 3: As more simultaneous gray levels were added, reading speed at high-contrast increased, whereas reading speed at low contrast decreased slightly. This is understandable in terms of the contrast represented by each gray level in the two situations. Low-contrast edges have only a fraction of the 12.5% contrast, whereas at high contrast, even with eight gray levels, the lowest gray level is equal to the full low-contrast value. Subjects described the high contrast, eight-gray-level stimulus as “smooth,” and a similar effect of gray-scale resolution was noted in the face recognition test data reported by Thompson et al.11

Interindividual Differences
Interindividual differences in this study were analyzed in a subject group of similar age and educational background. It has been suggested that older normally and partially sighted individuals may have been more representative of retinal prosthesis wearers, whose ages and central visual processing may differ from those of our subjects. We do not expect this to be of importance, however. The reading material was simple enough to be easily understood by adults of any age in good mental health, and the single most important limitation on performance was the adverse visual environment of the simulations. It is possible that older normally sighted subjects would have more difficulty learning to perform pixelized vision tasks than our younger subjects, but it is unlikely that they would have been more representative of subjects skilled in working with severely restricted visual information through years of advanced RP. Recent studies involving subjects with severe visual impairment in our laboratory have demonstrated that these subjects are, if anything, less hindered by the use of degraded pixelized imagery than those with normal sight (Dagnelie G, et al. IOVS 2005;46:ARVO E Abstract 1490).

Comparison with Other Reports on Filtered and Pixelized Text Reading
Findings in pixelated text reading reported by Legge et al.,24,25 Cha et al.,15 Sommerhalder et al.,15,16 and Forinos et al.17 Bagnoud et al.14 all seem to agree that performance is impaired if fewer than six samples/character width are available, particularly with larger characters. Two recent studies,16,17 arrive at a slightly lower limit, showing that practice at 5 pixels/ character width in the periphery is possible. In this study, however, accuracy at high contrast remains at 100% down to 4 dots/character width and remains better than 80% for most conditions, with as few as 2.5 dots/character width. With the exception of the narrowest dot–gap combination, no significant decrease in reading speed occurs above 4 dots/character width (Fig. 4, top left panel). Similar trends are seen at low contrast, with accuracies for the largest dot–gap combination even exceeding those at high contrast. Reading was painfully slow (3–8 words/min) in these conditions, but it was possible.

Our subjects’ ability to read at lower sampling densities than subjects in similar studies elsewhere must be due to one of two important differences in scanning. First, our subjects could scan the dot raster across the underlying text in small and rapid increments, thus gaining spatial information about the stimulus pattern through temporal sampling; most previous studies of filtered and sampled text reading primarily dealt with static images. Cha et al.15 used scrolling as well as stationary text, but this text movement was imposed on the observer. The subjects in Forinos et al.17 were given real-time image feedback to eye movement scanning in some tests, but they only used peripheral vision, and letter edges were represented by the absence rather than the presence of Gaussian dots or square pixels, making the detection task quite different from that of subjects in the present study. A second difference was the freedom of our subjects to scan their eyes across the displayed image, using any benefits that might be derived from foveal vision for all locations in the display. Such scanning would not be available for phosphophones at fixed retinal locations, leaving only the first form of scanning (typically by a head-mounted camera) for inspection of the underlying text image. This advantage may be greatest at low contrast, where the dot edges at high spatial frequency may interfere less with the detection of the lower-spatial-frequency text and may explain why for some conditions in this study, accuracy was better at low than at high contrast. It is important to test the contribution of this free-viewing condition in a separate series of tests, but this does not take away the advantage of our subjects in resampling the text by moving the mouse. The same advantage is available to a prosthesis wearer scanning a head-mounted camera across a stimulus, which is known to confer a distinct advantage over passive stimulus movement or stationary stimuli.3

Implications for Retinal Prostheses
As stated in the introduction, any extrapolation of reading performance with pixelized stimuli to reading with prosthetic vision hinges on the assumption that multiple phosphophones generated by a retinal implant bear some similarity to the regular patterns of identical dots in our simulations. Bearing in mind this crucial limitation, the results presented herein may have design implications for retinal prostheses, and can help limit the expectations for the benefits of such devices.
According to the right column in Figure 3, appropriately chosen optical magnification will allow text to be read with greater than 90% accuracy under a wide range of conditions, even at reduced contrast. If we define reading with 90% accuracy at 20 to 30 words/min as acceptable performance for a visual prosthesis, then we learn from Figure 4 that a retinal implant with 16 × 16 electrodes and 22.5 ± 18 arcmin (≈200 μm) center-to-center distance may allow its wearer to attain this measure of reading success. Such a device would span a visual field area of approximately 10° × 10° and measure 5 × 3 mm², and its electrodes, while a factor of 2 to 3 smaller than those currently in use, would be similar to some that have been used successfully for acute intraoperative stimulation. Such electrodes are capable of activating ganglion cells without exceeding safe charge-density limits if a high capacitance surface material such as iridium oxide is used. Although 16 × 16 electrode implants may not be available for several years, even a 10 × 10 array with similar electrode properties may allow some reading at rates of 10 to 15 words/min, but it is obvious that such a device would be inadequate for paragraph reading. In fact, even the 16 × 16 electrode implant cannot compete with Braille or speech output reading devices. Bearing in mind that current assistive technologies do not work in every location and on all text materials (e.g., handwritten notes, price tags), there may still be a limited reading benefit for implants as small as 10 × 10 electrodes. However, it is clear from the bottom panels of Figure 4 that even under ideal circumstances, reading speeds acceptable to low-vision readers would necessitate a device with at least 25 × 25 electrodes, in line with the findings of Cha et al.

## Remaining Shortcomings of These Simulations

The simulations presented to our subjects were simplistic representations of what may be experienced by a retinal prosthesis wearer. The most significant shortcomings fall into two areas: phosphene stabilization and phosphene appearance. Phosphene stabilization is an inevitable consequence of the fixed location of the electrode array relative to the retinal substrate. If a stationary stimulus is presented through such an electrode array, at a stimulus repetition rate exceeding the critical flicker fusion frequency of the target cells, then the prosthesis wearer is likely to report a fading percept, just as is experienced when a stationary light pattern is stabilized on the retina. According to reports by Humayun et al., the importance of this stabilization effect may have been overestimated in the situation of a prosthesis wearer who uses a camera as a scanning device or who otherwise has the ability to pan and/ or zoom through the scene. Nonetheless, it is essential to perform reading tests similar to those presented herein under conditions of retinal stabilization of the dot grid.

The appearance of our stimuli as grids of clean, round dots, even with the added feature of random dropouts is a more serious shortcoming of our simulations. Phosphenes elicited electrically in the retina do not resemble sharp-edged round dots as in these tests. Intraoperative retinal stimulation test subjects and early chronic implant recipients describe phosphenes as dots or (sometimes) rings, fuzzy and not necessarily perfectly round. Adjusting the properties of the dots presented in a simulation so that they are more similar to realistic phosphenes is possible and will lead to more realistic simulations. However, this does not address another aspect of our simulations: the regularity of the dot grid. Reports from retinal implant wearers describe the ability to recognize horizontal and vertical lines, and even angles, received from camera input, but these subjects do not describe seeing distinct dots making up such shapes. In an advanced prosthesis design this may be partially addressed through adjustment of stimulus parameters and other technical advances, but it is doubtful that phosphene images will approximate the idealized grids (even with dropouts) presented to our subjects. For example, even if prosthesis wearers perceive a grid of fuzzy dots, these are likely to vary in size and brightness across the grid. Adjusting pulse parameters for individual electrodes may allow the grid to achieve greater homogeneity, but this will probably be at the expense of dynamic range and resolution. It is conceivable that reliable retinal prostheses will have to await the development of penetrating electrodes that more closely contact the remaining retinal cells. Although this will postpone the advent of functional implants, it does not invalidate the implications of these simulations for prosthetic vision in the long run.

### Acknowledgments

The authors thank the volunteer participants.

### References


