Subjectively homogeneous noise over written text as a tool to investigate the perceptual mechanisms involved in reading

Frédéric J. A. M. Poirier Département de Psychologie, Université de Montréal, Montréal, Québec, Canada

Frédéric Gosselin Département de Psychologie, Université de Montréal, Montréal, Québec, Canada

Martin Arguin Département de Psychologie, Université de Montréal, Montréal, Québec, Canada

In an effort to understand the factors influencing text legibility in natural reading, we adapted the visual spread method (Poirier, Gosselin, & Arguin, 2008) to natural text. Stimuli were sentences conforming to MNREAD standards (Legge, Ross, Luebker, & LaMay 1989) mixed with dynamic probabilistic noise—i.e., each pixel in the image is associated with a probability that its polarity is inverted on a given refresh cycle of the display screen. Noise level varied continuously over the image as initially determined by Gaussian-filtered noise. Participants adjusted noise levels in the text using the mouse until the text appeared homogenously noisy. We assume that participants increased (or decreased) noise at locations where stimulus features were easy (or difficult) to encode and thus that local noise settings correlate with legibility. Data from 11 participants and 30 sentences revealed interesting effects, demonstrating the validity of the method for assessing the impact of various factors on noise resistance in natural text. For example, participants increased noise over (a) spaces and adjacent letters, (b) the second half of words, (c) words with more orthographic neighbors but fewer phonological neighbors, (d) less useful word types, (e) less complex letters, and (f) diagnostic letters (a novel metric). Our observations also offer significant insights on constraints acting upon letter identification as well as on higher-level processes that are involved in reading.

Introduction

Reading performance is dependent on many factors, including word frequency (for review, see Ferrand, 2001; Monsell, 1991), letter similarity or confusability (Appleman & Mayzner, 1982; Bouma, 1971; Fiset, Dupuis-Roy, Arguin, & Gosselin, unpublished observations; Rumelhart & Siple, 1974), orthographic and phonemic neighbors (Andrews, 1997; Coltheart, Davelaar, Jonasson, & Besner, 1977; Ferrand, 2001; Grainger & Jacobs, 1996), and letter position in the word (Nazir, Jacobs, & O'Regan, 1998; O'Regan, 1990, 1992; O'Regan & Jacobs, 1992; O'Regan & Levy-Schoen, 1987; O'Regan, Levy-Schoen, Pynte, & Brugaillère, 1984; Vitu, O'Regan, & Mittau, 1990).

Reading has also been studied via eye tracking (e.g., O'Regan, 1990; Rayner, 1998; Rayner & Pollatsek, 1989; Rayner & Sereno, 1994). Such studies found more fixations on names, verbs, and adjectives (about 85% of them), less on prepositions, conjunctions, articles, and pronouns (about 35%), as well as more fixations on longer words than shorter words.

The purpose of the present study is to assess the merits of a novel approach, the visual spread method (Poirier, Gosselin, & Arguin, 2008; see below), for the investigation of reading. For brevity, we will review the literature where appropriate in the Discussion section.

The visual spread method

We developed the visual spread method to investigate saliency (Poirier et al., 2008; see also Nothdurft, 1993, 2000) where participants adjusted luminance to match the saliency produced by other attributes (e.g., color, orientation) across all locations in that image. This increased data collection efficiency compared to two alternatives forced-choice (2AFC) tasks: (a) participants produced more responses per hour of

Citation: Poirier, F. J. A. M., Gosselin, F., & Arguin, M. (2013). Subjectively homogenous noise over written text as a tool to investigate the perceptual mechanisms involved in reading. *Journal of Vision*, *13*(11):14, 1–19, http://www.journalofvision.org/content/13/11/14, doi:10.1167/13.11.14.

doi: 10.1167/13.11.14 Received October 18, 2012; published September 24, 2013 ISSN

testing, and (b) each response revealed both the location and intensity of perceived saliency.

In the present study, we adapted the visual spread task to evaluate which parts of the written sentence were more noise resistant. Our measure of noise resistance was subjective, in that participants adjusted noise levels until they perceived noise to be homogeneous over the sentence. Actually, the adjustments made by participants systematically contained more noise on some parts of the sentences than others. We assumed that participants spread the noise to equalize signal-to-noise ratios across the stimulus. If that assumption was correct, then the analysis performed on noise distribution would recover the factors contributing to the signal.

Advantages of the visual spread method applied to reading include: (a) the data recovered covers scales ranging from single letters to entire words, (b) any factor that varies within the stimulus sentences can be correlated with the data to estimate its contribution, and (c) regression analyses can determine if factors are redundant or unique (see also Appendix D).

We used regression analysis to determine which factors were correlated with noise resistance, including a vast array of factors related to letters, words, and spaces. Our results are congruent with the literature regarding the factors affecting reading performance despite methodological differences. The ecological validity of our method is also discussed (see Appendix D).

Methods

Ethics statement

The research protocol was reviewed and approved by the University of Montréal Health Research Ethics Committee. Written informed consent was obtained from participants prior to their participation.

Participants

Eleven participants volunteered (four males and seven females), including the first author, as well as university undergraduate and graduate students. Their vision was normal or corrected-to-normal. Participants were French native speakers and fluent French readers.

Apparatus

Testing and data collection were conducted on a PC computer (P4 3 GHz; 800 × 600 pixels; 75 Hz refresh rate). Responses were recorded via mouse button presses. Viewing distance was 68.5 cm, where 16 pixels equaled 1° visual angle. The entire stimulus subtended 8° × 16°. Stimulus generation and data collection were controlled by MATLAB with Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

Procedure

Visual spread task

In the current visual spread task, participants equated noise levels over sentences. Nothdurft (1993, 2000) introduced a "saliency match" task where participants had to indicate which of two stimulus patches was more salient (i.e., which patch grabbed attention). This allowed him to equate saliency across attributes (e.g., luminance, color, orientation). In the study of Poirier et al. (2008), participants distributed luminance over an image made of randomly colored and oriented lines. Adding (or removing) luminance at an image location increased (or reduced) the visibility of features at that location. Participants' luminance matches were significantly correlated with properties known to modulate saliency.

In the current study, on each trial, participants were shown a display containing a sentence written over three lines in black and white (Figure 1B, C). Each pixel of the stimulus was white or black and was associated with a probability of polarity inversion, which constitutes the noise level at that location. Participants were required to equate apparent noise levels over the image by left-clicking (or right-clicking) on points where noise levels were perceived as higher (or lower) than elsewhere, which decreased (or increased) the noise level within a small window around that location. The position and duration of button presses was used to dynamically adjust noise levels.

Stimuli

Stimuli consisted of two parts: a sentence and probabilistic noise. Sentences were taken from a bank of French sentences (Senécal, 2001; Senécal, Gresset, & Overbury, 2002) constructed to conform to MNREAD standards (a standard reading speed test; Legge, Ross, Luebker, & LaMay, 1989): (a) 60 characters long, including spaces between words, (b) about 10 words long, (c) divided in three lines, each about equal in physical length and centered on the screen, and (d) using simple sentence structures and high-frequency words easily understood by eight-year-old children. The height of the letter "x" was 7 pixels (0.44°), and

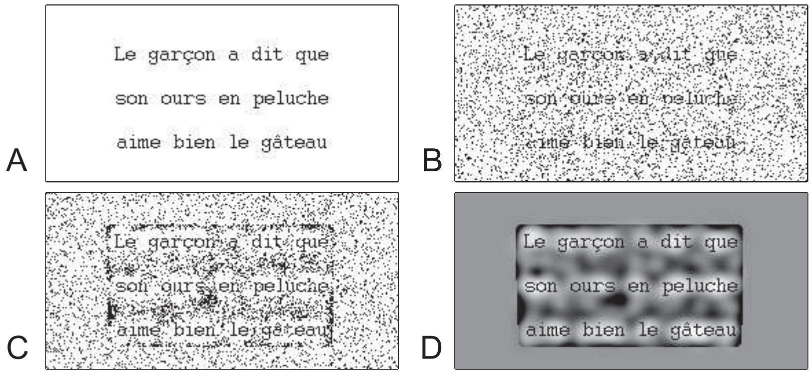

Figure 1. A sample stimulus is shown without noise (A) and after adjustment by participants (B; average). Panel B is also shown with noise differences amplified (C), or with noise levels represented as gray levels (D; surround gray = 12.5% noise; lighter = less noise). The noise levels in Panel B appear roughly homogeneous over the image, despite systematic variation in noise levels as highlighted in panels C, D.

letters could extend 3 pixels (0.19°) above and/or below that. The font used was Courier. Only lowercase nonaccented letters were used in the analyses, with capital letters, accented letters, and apostrophes classified as "miscellaneous characters."

The noise was dynamic and probabilistic. Every pixel was associated with a probability of being displayed at opposite polarity on any frame (13.3 Hz). This probabilistic noise was defined using (detailed below): (a) initial noise levels set using filtered random noise, (b) filtered noise that was normalized, (c) and then transformed into probabilistic noise, and (d) as participants changed noise levels, values were renormalized.

The filtered binary noise (N) was created using binary noise filtered using a Gaussian filter, defined as:

$$G_{i,j} = e^{\frac{-(i^2+j^2)}{\sigma^2}} \tag{1}$$

where $\sigma = 0.5°$, and (i, j) are coordinates. N was linearly scaled to a range of 0 to 1, normalized to a mean of 0.5, and values outside the 0 to 1 range were clipped. Each time N was updated by user responses, it was renormalized and reclipped (if necessary). The resulting scaled filtered noise was transformed into probability noise (P) using the following transformation:

$$P = 50\% * N^2 \tag{2}$$

where pixels varied from: (a) showing the sentence at all times, i.e., never inverting, $N = 0$, $P = 0\%$, (b) showing random noise at all times, i.e., $N = 1$, $P = 50\%$, and (c) showing sentence and noise in various amounts, average $N = 0.5$, average $P = 12.5\%$. The squaring exponent made noise levels vary smoothly over time when adjusting noise levels because small changes are noticeable at low-noise levels. Dynamic noise was used rather than static noise because (a) noise density is easier to evaluate in dynamic noise due to averaging over time, and (b) participants saw real-time effects of adjusting noise.

Participants adjusted noise levels by pressing either mouse button at a location (N varied by 0.5/s; the right/left button increased/decreased noise respectively in a Gaussian window with $\sigma = 0.5°$). Each adjustment was followed by renormalization and reclipping as described above. Normalization was done to prevent participants from setting noise at ceiling or floor values, as well as to keep average noise levels constant across participants and sentences. At 12.5% noise level, words near and far from fixation remained visible. Higher levels of noise would mask the text context, thus potentially bias results towards local properties.

Noise could only be adjusted by participants within an "area of interest" (AOI) on and near text. Outside the AOI, pixels were kept at 12.5% noise. The AOI was defined as the areas containing letters extended two character spaces on each side and two character heights above and below, Gaussian-blurred ($\sigma = 0.5°$) to smooth the transition between inside and outside of the AOI. This encouraged participants to concentrate responses on or near text. "Surrounding space" is space contained within the AOI but located either between or outside sentences.

Following the logic that some sentence parts are perceived more clearly than others, the point of subjective noise equality would be negatively correlated with text clarity. We henceforth use the term "noise resistance" to denote local noise density, in units of percent noise.

Temporal sequence

The 30 sentences were divided in three test sessions. Each sentence was shown about equally often (details

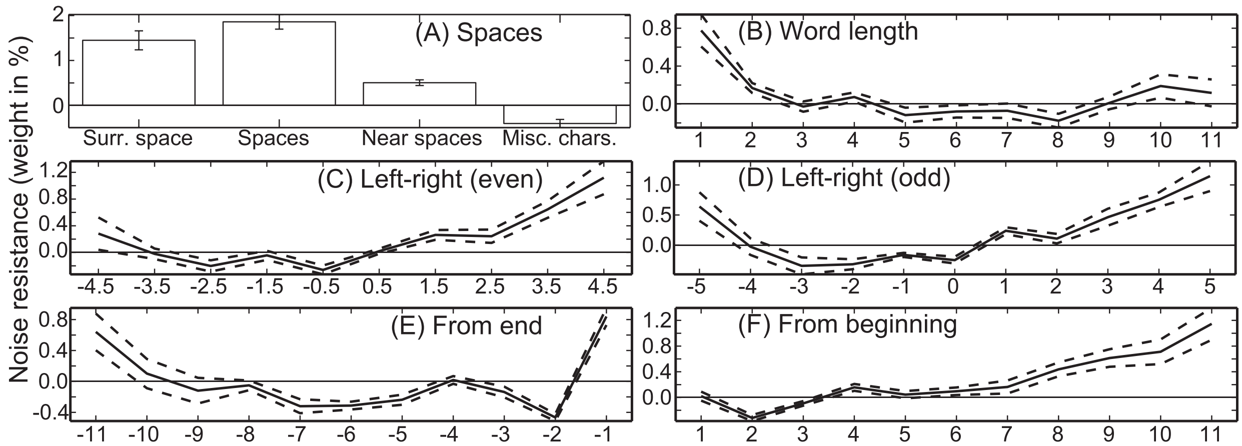

Figure 2. Weights (-*y*-axis) associated with space-based and positional-based factors. A positive (or negative) weight indicates that noise was higher (or lower) in the areas indicated by the factor. Participants increased noise on spaces surrounding and between words (A), on external letters (A), one-letter words (B), on letters further towards the right (C, D, F), and on the last letter. They decreased noise on miscellaneous characters (A). Error bars and dotted lines represent *SEM*.

below). Each session was 30 min long, for a total of 90 min of testing per participant.

Each trial lasted 20 s, during which participants could make as many adjustments as they wanted. Noise levels were adjusted in real-time. Trials were separated by a blank screen for 500 ms. On 50% of trials, the sentence was selected randomly. On other trials, the least-presented sentence was selected, to ensure that every sentence was presented about equally often. Each sentence was presented nine times on average, with repeated presentations using the noise as set in the prior trial to encourage participants to search carefully for uneven noise.

Analysis

Analyses were performed in two steps as recommended by Lorch and Myers (1990). (1) Regression analyses were performed to extract factor weights per participant. Regression weights represent noise change per unit of change in a given factor. The standard error of the means (*SEM*) of those weights represents the accuracy of that mean, despite potential individual differences in strategy. (2) Those weights were analyzed using repeated measures ANOVAs (to compare weights together) and *t* tests (to assess whether individual weights were significantly different from zero). Due to individual variability, *t* tests on individual weights were less powerful than ANOVAs. For orthogonal factors, this is equivalent to measuring mean noise relative to baseline noise and performing standard analyses. However, the method generalizes to nonorthogonal factors. Note that since individual differences in strategies contribute to the error terms of these analyses, effects that are statistically significant are effects that show consistency across participants.

For example, the analysis of space-based factors shown in Figure 2A included two steps: (1) A regression analysis extracted mean noise for surrounding space, spaces, letters near spaces, and miscellaneous characters for each participant, and (2) *t* tests determined independently which of these four factors influenced noise resistance.

Some factors were decomposed into separate predictor variables to clarify trends in two steps: (1) A regression analysis extracted mean noise at each level of the variable, for each participant, and (2) an ANOVA determined if noise varied as a function of that variable. For example, the analysis of the word length effect shown in Figure 2B included two steps: (1) A regression analysis extracted mean noise for each word length and participant, and (2) an ANOVA determined if mean noise varied with word length. Note that these factors were also analyzed not decomposed (e.g., Table 2).

Regression analysis combined factors in three ways: (a) "independent," meaning that factors are entered alone, (b) "blocked," meaning that all factors from that block were included, and (c) "full," meaning that all factors were included. Analyses including more factors were performed to evaluate the relative importance of factors by allowing them to compete for explained variance, as well as controlling for lower-level factors. In all cases, all variables were entered simultaneously; therefore, no preference was given in ability to compete for accounted variance. All regressions include controls for spaces, surrounding spaces, and miscellaneous characters, including "independent" regressions.

	l	e	s		c	h	i	e	n	s
Word length	3	3	3	0	6	6	6	6	6	6
Distance from beginning	1	2	3	0	1	2	3	4	5	6
Distance from end	−3	−2	−1	0	−6	−5	−4	−3	−2	−1
Distance left-right	−1	0	1	0	−2.5	−1.5	−0.5	0.5	1.5	2.5
Noun					1	1	1	1	1	1

Table 1. Some sample factors and their values for each character position in a sequence of two words. *Notes*: "les chiens" in English: "the dogs".

Factors

Most factors are described in the Results sections where they are analyzed. Factors requiring a lengthy explanation are described in Appendices A and B, which also include some factor statistics.

We used a battery of factors that could potentially influence noise resistance using multiple regression

Category	Factor	Independent	Blocked	Full	Figure reference	R^2
Spaces & misc.	Surrounding space	1.448 ± 0.211	0.659 ± 0.175	1.615 ± 0.396	2A	21.7% ± 4.1%
	Spaces	1.861 ± 0.163	1.939 ± 0.169	0.754 ± 0.259		
	Near spaces	0.506 ± 0.063	0.506 ± 0.063	0.182 ± 0.119		
	Miscellanous characters	−0.402 ± 0.090	−0.454 ± 0.089	−0.092 ± 0.111		
Positional	Word length	−0.018 ± 0.011		—	2B–F	20.8% ± 3.9%
	Left-right	0.107 ± 0.022		—		
	From end	0.084 ± 0.016		—		
	From beginning	0.060 ± 0.016		—		
Letters	Width	−0.397 ± 0.268	1.181 ± 0.323	—	3	27.7% ± 4.3%
	Confusability in noise	0.795 ± 0.120	1.339 ± 0.328	—		
	Confusability low contrast	0.581 ± 0.102	−0.880 ± 0.257	—		
	Complexity	−1.092 ± 0.150	−1.446 ± 0.186	—		
	Log letter frequency	0.272 ± 0.046	0.074 ± 0.062	—		
	Diagnosticity	0.363 ± 0.132	0.099 ± 0.137	—		
Sublexical	Grapheme length	0.000 ± 0.052	−0.003 ± 0.053	0.019 ± 0.052	4A, B	20.3% ± 3.7%
	Syllable length	−0.017 ± 0.019	0.006 ± 0.018	−0.021 ± 0.011		
	Token syllable frequency	0.019 ± 0.025	−0.128 ± 0.040	−0.101 ± 0.032		
	Type syllable frequency	0.145 ± 0.034	0.258 ± 0.046	0.154 ± 0.050		
Frequencies	Log frequency	0.014 ± 0.008	−0.017 ± 0.013	0.022 ± 0.028	4C, D	20.6% ± 3.7%
	Log cumulative Frequency	0.015 ± 0.009	0.024 ± 0.014	−0.021 ± 0.029		
	N homographs	0.021 ± 0.015	−0.011 ± 0.013	0.013 ± 0.019		
	N homophones	0.001 ± 0.006	0.001 ± 0.007	−0.007 ± 0.007		
	Log orthographic neighbors	0.052 ± 0.023	0.097 ± 0.022	0.056 ± 0.027		
	Log phonetic neighbors	0.000 ± 0.019	−0.056 ± 0.020	−0.037 ± 0.020		
Word type	Auxiliary verb	0.445 ± 0.090	0.302 ± 0.129	0.484 ± 0.253	4E, F	20.6% ± 3.8%
	Conjunction	0.331 ± 0.188	0.191 ± 0.219	0.528 ± 0.255		
	Article	0.159 ± 0.077	0.014 ± 0.117	0.369 ± 0.222		
	Pronoun	0.142 ± 0.114	−0.001 ± 0.146	0.393 ± 0.187		
	Verb	0.017 ± 0.029	−0.121 ± 0.098	0.425 ± 0.213		
	Preposition	−0.004 ± 0.066	−0.138 ± 0.109	0.286 ± 0.161		
	Noun	−0.038 ± 0.073	−0.159 ± 0.080	0.339 ± 0.168		
	Determiner	−0.093 ± 0.113	−0.223 ± 0.158	0.320 ± 0.259		
	Adverb	−0.108 ± 0.064	−0.233 ± 0.105	0.199 ± 0.185		
	Adjective	−0.146 ± 0.064	−0.264 ± 0.119	0.264 ± 0.210		

Table 2. Summary of analyses. *Notes*: All analyses include the factors "spaces," "miscellaneous characters," "surrounding space," and a constant. Weights are shown for independent, blocked, and full regressions. Intervals are *SEM* across participants. Factors marked by "—" were included in the analysis as controls. Empty lines represent factors not included in the given regressions. R^2 for the full analysis was 29.5% ± 4.2%.

analysis. Table 1 shows some sample factors for two sample words, along with the numerical value assigned to each letter position on those factors. Note that all letter positions received the same value if a factor characterizes the entire word. Note also that categorical variables were assigned a numerical code of nominal value.

Factors were selected if they met the following criteria: (a) they had to be easily quantifiable based on physical characteristics or previously collected data, and (b) they had to vary naturally within the selected sentences, rather than requiring specific manipulations. Note that the analysis could have included further factors, and data could be re-analyzed as further factors are added. However, we believe this set of factors constitutes a fairly comprehensive list of factors present in normal reading situations. The main factors are listed in Table 2, and Table 1 shows some examples of factors used. Factors are described in Appendices A and B. Diagnosticity was a novel factor to estimate how important single letters were to word recognition. Confusability measured how likely a letter was to be confused as another letter.

Results

Participants systematically set noise levels higher in certain locations than others, resulting in systematic noise modulations over the image, which we quantified as "noise resistance" (in units of percent noise).

Factors underlying noise resistance were analyzed in four blocks: (a) space-based, (b) positional, (c) letter-based, and (d) text-based factors. The main results are summarized in Table 2. As new factors are added, some weights on previous factors changed (e.g., reverse direction, drop to zero). Such weight changes occur when other factors compete to explain the same variance. Such changes are thus informative and discussed below.

Space-based factors

This block included: (a) "surrounding space" indicating the empty space directly surrounding the text, but not including spaces between words, (b) "spaces," (c) "near spaces" indicating letters adjacent to one or two spaces, and (d) "misc. letters" including apostrophes, accented letters, and uppercase letters (Figure 2A). For these factors, a weight of zero indicated that noise was as high as noise over regular letters.

Participants had a higher resistance to noise for surrounding space, spaces, and the external letters of words than for internal letters, $ts(10) \geq 6.9$, $ps < 0.0001$. Participants put less noise on miscellaneous characters than on letters, $t(10) = -4.5$, $p = 0.001$.

Positional factors

Positional factors refer to the position of a letter in a word (in number of letters; see Table 1): (a) word length ($M = 5.54$, $SD = 2.26$), (b) "left-right" distance from the word's center keeping directional information (positive was towards the right), and (c–d) distance from the beginning/end of words. The factors were analyzed separately for odd- and even-length words as well as combined into a single analysis, with similar results.

Positional factors were decomposed into independent levels to analyze trends (see "Procedure: Analysis" section). For example, the factor "distance from beginning" is decomposed into the sub-categories "first letter," "second letter," etc., each of which indicates if a letter belongs to the subcategory (1 = yes, 0 = no).

Participants increased noise on (Figure 2B through F): (a) one-letter words, $F(10, 100) = 7.55$, $p < 0.001$; LSD $ps < 0.02$; Figure 2B, (b) letters further towards the right of a word, linear $Fs(1, 10) > 14.0$, $ps \leq 0.004$; Figure 2C, D, and F, (c) letters further from the center of a word, quadratic $Fs(1, 10) > 13.6$, $ps < 0.004$; Figure 2C, D, and (d) the last letter (last letter different from any previous letter; $F[10, 100] = 12.06$, $p < 0.001$; LSD $ps < 0.015$; Figure 2E). Taken together, these factors indicate that participants increased noise levels on letters away from the center and towards the right of words.

Letter-based factors

This block included six factors (Appendices A and B, "Factors" section in the Methods, and Table 3): (a) letter complexity, (b) letter width, (c–d) confusability in noise or in low contrast, (e) log letter frequency, and (f) diagnosticity. The variables complexity, letter width, and diagnosticity were normalized to a maximum value of 1 for the regression analysis.

Noise-resistant letters were frequent, easily confusable, diagnostic, and less complex (Table 3; Figure 3). Letter width had no effect on its own, $t(10) = -1.5$, $p = 0.17$, but if entered with other factors, wider letters received more noise, $ts(10) \geq 3.7$, $p \leq 0.0045$. That is, the letter width effect was too small to be observed on its own.

Both confusability factors were positively correlated with noise resistance (independent: $ts[10] \geq 5.7$, $ps <$ .0003). Contrary to intuition, letters of high confus-

Factor	Noise resistance	Diagnosticity	Log letter frequency	Complexity	Confusability in low contrast
Confusability in noise	0.646	0.410	0.566	−0.756	0.953
Confusability in low contrast	0.676	0.487	0.644	−0.713	
Complexity	−0.561	−0.239	−0.457		
Log frequency	0.428	0.810			
Diagnosticity	0.442				

Table 3. Correlations between several measures based on the same font. *Notes*: If the factor complexity is multiplied by −1, all correlations become positive here.

ability received more noise. When entered with other factors, "confusability in noise" remained, $ts(10) \geq 3.8$, $ps < 0.0035$, whereas "confusability in contrast" became negatively associated with noise resistance, $ts(10) \leq -3.2$, $ps < 0.01$. In other words, participants *increased* noise on letters with low confusability in *low contrast*, and greater confusability in *noise*.

Complex letters were less resistant to noise in all analyses, $ts(10) \leq -6.9$, $ps < 0.0001$.

Frequent letters received more noise, $t(10) = 5.9$, $p = 0.0002$, but this effect disappeared when other factors were included in the analysis, $ts(10) \leq 1.2$, $ps \geq 0.26$, suggesting that letter frequency did not contribute uniquely to explained variance.

Higher diagnosticity was associated with an increase in noise resistance, $t(10) = 2.8$, $p = 0.020$. The effect disappeared when other factors were included, $ts(10) \leq 0.54$, $ps \geq 0.60$, suggesting that diagnostic letters are noise resistant due to other features.

Taken together, letter-based factors offer a solid account of noise resistance for individual letters of the alphabet (Figure 3C). The correlation between predicted noise using letter-based factors (y-axis), and data averaged over multiple instances of a letter (-x-axis)

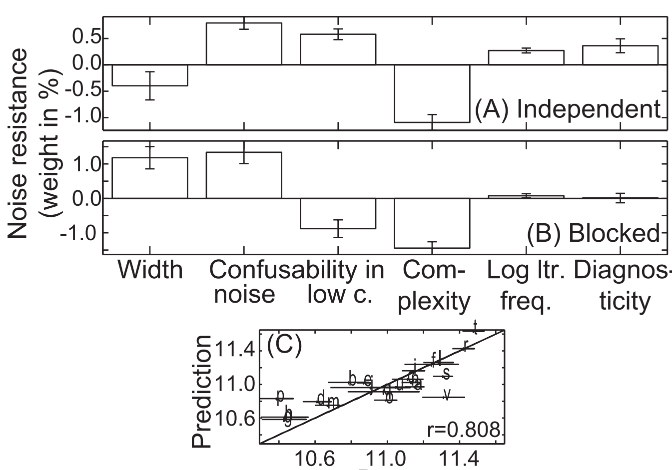

Figure 3. Weights (y-axis) associated with letter-based factors, when entered independently (A) or "blocked" together (B). A weight of "0" indicates that the factor is not associated with noise increase or decrease. (C) Correlation between averaged noise and predicted noise per letter. "Low c." is an abbreviation for "low contrast."

was high ($r = 0.808$). Error bars indicate *SEM* based on individual variability, with error bars for the model based on individual variability of regression fits.

Text-based factors

The last block of factors includes 24 factors, classified into: (a) sub-lexical statistics, (b) frequencies, and (c) word type.

Of the sublexical factors (Figure 4C, D), only syllable frequency by type was independently correlated with noise resistance, $t(10) = 4.3$, $p = 0.0016$; all others: $|ts|(10) < 0.75$, $ps > 0.39$. In the blocked analysis, noise resistance increased with type frequency, $t(10) = 5.7$, $p = 0.0002$, and decreased with token frequency, $t(10) = -3.23$, $p = 0.009$. The syllable frequency factors (type and token) remained significant (or at least near-significant) in the full analysis, $|ts(10)| > 3.09$, $ps < 0.012$; except for token syllable frequency in the full analysis: $t(10) = -2.0$, $p = 0.076$. Grapheme and syllable length were not significant in any analysis, $|ts|(10) < 1.82$, $ps > 0.099$. This suggests that the two syllable frequency factors are unique contributors to explained variance, where words are more noise resistant if they have high type frequencies yet low token frequencies.

Of the frequency factors (Figure 4E, F), log orthographic neighbor frequency was associated with an increase in noise, $ts(10) \geq 2.2$, $ps \leq 0.049$; near-significant in the full analysis: $t(10) = 2.07$, $p = 0.065$, and log phonological neighbor frequency was only significant in the blocked analysis, $t(10) = -2.8$, $p = 0.019$; independently: $t(10) = 0.0$, $p = 0.99$; full: $t(10) = -1.84$, $p = 0.095$. None of the other frequency-based factors was significant in any of the analyses (log lexical frequency, log cumulative frequency, number of homographs, number of homophones; $|t|s \leq 1.8$, $ps \geq 0.10$. Thus participants increased noise on words that were graphically similar to other words, and had a more unique pronunciation. The phonological effect was more dependent on which factors were included in the analysis.

There was an effect of word type, $F(9, 90) = 4.1$, $p = 0.0$ (Figure 4E, F), with words associated with sentence

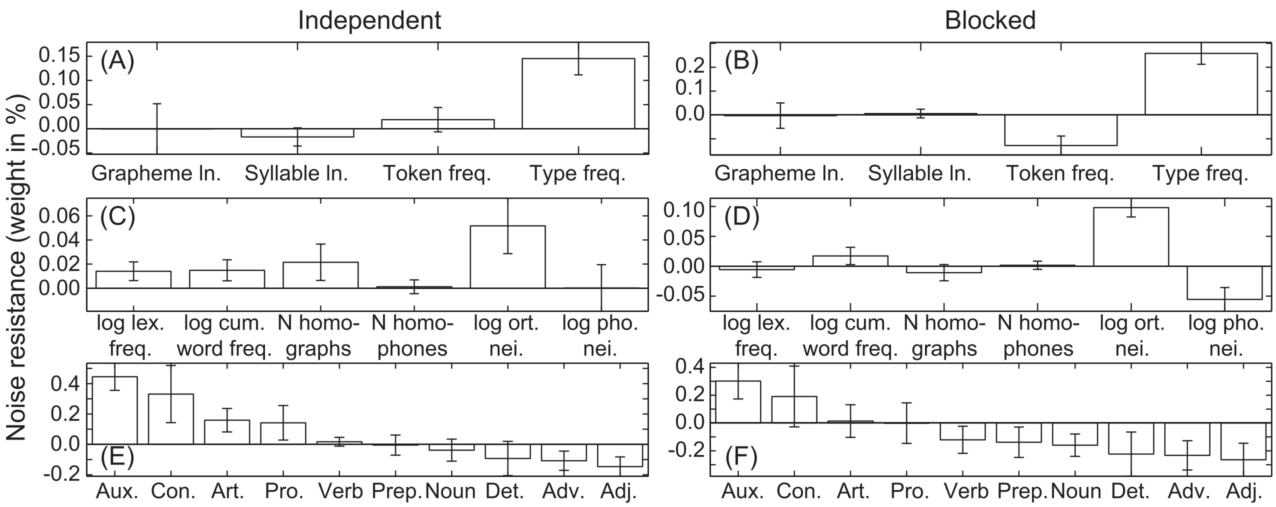

Figure 4. Weights (*y*-axis) associated with text-based factors, when entered together (blocked; right) or independently (left). Factors are arranged in panels according to the blocks they were entered in. A weight of "0" indicates that the factor is not associated with noise increase or decrease. (A, B) Of the four subword statistics, only the syllabic frequencies were significantly correlated with noise levels, where participants increased noise on high syllable type frequency and low syllable token frequency. (C, D) Participants increased noise on words with more orthographic neighbors and fewer phonological neighbours. (E, F) Word type also influenced noise levels.

understanding generally receiving less noise. We observed the following ranking of word type from lowest to highest noise resistance (all ps < 0.05): (a) adjectives, (b) verbs, (c) articles, and (d) auxiliary verbs. Other word categories were too variable across participants to establish a statistically reliable ordering. The word-type effect was not significant in the full analysis, $F(9, 90) = 1.7$, $p = 0.098$, suggesting that the word-type effect could be due to other factors.

Discussion

Summary of results

Within 90 min of data collection per participant, the novel and intuitive method of visual spread produced reliable data on perceived text clarity or noise resistance throughout a set of sentences.

Noise resistance, as measured using the visual spread task, was significantly correlated with many factors consistent with the reading literature, including (a) position of spaces, (b) letter position within words, (c) letter complexity, (d) letter confusability, and (e) lexical characteristics. We also document an effect of the novel factor of letter diagnosticity. Thus, although participants were not explicitly asked to read the sentences, their responses nevertheless correlated with some factors known to influence readability.

Although many factors were consistent across analyses, some factors were not. This suggests either non-additive effects or smaller effects that can be masked by larger effects. Examples of nonadditivity include: (a) the log number of phonetic neighbors was significant depending on which other factors were included, (b) the "confusability in low contrast" factor reversed sign when the "confusability in noise" factor was included, (c) token syllable frequency was only significant when type syllable frequency was included in the analysis, and (d) many effect sizes were reduced with controls for letters and position. These interaction effects are discussed below where appropriate, along with what they tell us about reading.

Overall, noise resistance seems to increase around spaces and external letters, towards the right of words, on easily confusable letters, on less complex letters, and on word with more orthographic neighbors. Other effects were found but were less consistent across analyses (see below).

Implications for optimal viewing position

The optimal viewing position for word recognition is at the center of the word, or slightly left of center, and performance is better for ocular fixations on the first half than on the second half of the word (Nazir et al., 1998; O'Regan, 1990, 1992; O'Regan & Jacobs, 1992; O'Regan & Levy-Schoen, 1987; Vitu et al., 1990). Fixations away from the optimal position bring performance costs of about 10–20 ms per letter (Nazir et al., 1998; O'Regan, 1992; O'Regan & Levy-Schoen, 1987; O'Regan et al., 1984; Vitu et al., 1990). One reason suggested for an advantage for external letters is that they are less exposed to crowding (Bouma, 1970; Pelli, Palomares, & Majaj, 2004), i.e., letters inside words are masked by flanking letters.

In our study, participants increased noise on and around spaces more than on the letters themselves. Noise resistance was lowest at about two letters left of word center, i.e., around the optimal viewing position. Participants increased noise on word endings (especially on the last letter), which, in French, are usually less informative (Blais et al., 2009; Clark & O'Regan, 1999). This suggests a noise placement strategy to move noise away from meaningful and crowded areas of words.

The size of the Gaussian spread of noise adjustment used in this study could explain in part why participants increased noise on external letters. Adding noise on spaces also adds some noise over external letters. This possible artefact does not explain many effects however, e.g., lower noise on the first half of words or the letter-based effects. Nevertheless, a replication of this study could be performed with noise density set per letter instead of spread using a Gaussian.

Implications for letter processing

Noise resistance was best predicted by the block of variables linked to letter characteristics. This suggests that modifying alphabet characteristics could substantially influence reading performance. Indeed, font choice does impact reading speed, comprehension, perceived attractiveness, and legibility (Bernard, Lida, Riley, Hackler, & Janzen, 2002; Mansfield, Legge, & Ban, 1996). This is convenient, because it is easier to change fonts (or even an alphabet) than word frequencies or statistics of word types. Our results include measures of variability *within* the alphabet chosen. Thus we can examine the relationships among letters that are often overlooked when reading performance is measured across sentences or paragraphs. The pattern of correlations between letter-based variables (Table 3) suggests that the alphabet is more optimal for writing than for reading (see Appendix C for details). Frequent letters are simple, making them easier to write. However, frequent and diagnostic letters are easily confusable with other letters, making them less-than-optimal for reading.

Implications for letter degradation methods

Our study compares three methods of degrading letter visibility: probabilistic noise (here) and confusability in both noise and low contrast (Fiset et al., unpublished observations). Not surprisingly, they were all correlated. Results suggest that participants increased the visibility of letters that are already easier to identify. However, the "confusability in low contrast" weight changed sign when the "confusability in noise" weight was included in the same analysis. This suggests that these two measures have a shared component (they both degrade letter recognition) as well as a unique component (the specific way letter recognition is degraded).

The three degradation techniques differ in how much they affect letter shape and contrast: (a) "confusability in noise" used additive noise, which influenced both letter shape and contrast, (b) "confusability in low contrast" used contrast reduction, which mainly influenced contrast, and (c) "probabilistic noise" (as used here), which influenced letter shape mainly. Thus, a fair approximation of shape degradation (probabilistic noise) would be the "shape + contrast" degradation (confusability in noise) *minus* the "contrast" degradation (confusability in contrast). Thus a regression approach as used here shows that different methods to degrade letters are not fully equivalent, with letters being differentially sensitive to different degradation methods.

Implications for sublexical effects

A grapheme is defined as the written representation of a phoneme. Studies have shown that words controlled for length take longer to read or identify if they have longer graphemes or syllables (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Plaut, McClelland, Seidenberg, & Patterson, 1996; Rastle & Coltheart, 1998; Rey, Jacobs, Schmidt-Weigand, & Ziegler, 1998; Rey, Ziegler, & Jacobs, 2000), suggesting longer processing time for longer graphemes. Our results show that neither grapheme length nor syllable length played a role in noise resistance. Our experiment did not include time pressure, and it is possible that performance costs associated with longer graphemes mainly occur under time pressure.

Conrad, Carreiras, and Jacobs (2008) studied the effects of syllable frequency on lexical decision: quantifying syllable frequency both by type (number of syllabic neighbors) and by token (summed word frequencies across syllabic neighbors). They found facilitation for type frequency, and inhibition for token frequency. Our study replicated their results: noise resistance increased with type frequency and decreased with token frequency. However, our findings differ from Conrad et al.'s on the sensitivity of the two effects to methodological changes or the inclusion of control variables. Conrad et al. (2008) report that the token frequency effect was reliably found across the literature

Figure 5. Hypothesized relationships between the variables related to letters. All correlations were positive between the variables shown. Diagnostic letters are frequently used (left). Efficiency was measured using noise resistance and simplicity (the opposite of complexity). Letter confusability was measured in low contrast and in noise (right). The results are consistent with an alphabet that promotes faster writing, but not necessarily faster or easier reading. Dotted lines indicate correlations that should be inverted when designing an alphabet to promote reading performance (without necessarily making writing performance worse), by changing the relationship between confusability and all other variables, while retaining all other relationships.

and their own study, whereas the type frequency effect was dependent on methodological demands and/or statistical controls. In contrast, in our study, the type frequency effect was reliably found across analyses whereas the token frequency effect was dependent on which other variables were included in the analysis. One possibility is that the two effects are differently sensitive to methodological demands (e.g., speeded responses). Further studies could investigate the temporal properties of these effects to confirm.

Implications for lexical frequency

Frequently occurring words are easier to recognize than less frequent words (for review, see Ferrand, 2001; Monsell, 1991) including in French (Ferrand, 2000; Grainger, Spinelli, & Ferrand, 2000; O'Regan & Jacobs, 1992). This effect is seen across word types (Gordon & Caramazza, 1985; Segui, Frauenfelder, Lainé, & Mehler, 1987). Shorter words are frequent (Zipf, 1935). Word frequency and word length seem to have independent effects on performance (O'Regan & Jacobs, 1992; Vitu et al., 1990). Reading speed is also improved for words with a higher cumulative frequency (word frequencies summed across same-family words; Beauvillain, 1996; Bradley, 1979; Burani & Caramazza, 1987; Burani, Salmaso, & Caramazza, 1984; Colé, Beauvillain, & Segui, 1989; Holmes & O'Regan, 1992; Taft, 1979).

In our study, neither lexical frequency nor cumulative lexical frequency played a significant role, although both showed trends in the right direction, consistent with a facilitory effect. Controlling for other factors did not help these two factors reach statistical significance. It is possible that these effects may require time constraints or reading performance measurement.

Implications for orthography

Orthographic neighborhood was measured as (a) frequency of the most frequent neighbor, and (b) the number of orthographic neighbors (Andrews, 1997; Ferrand, 2001; Grainger & Jacobs, 1996). Words with several orthographic neighbors are facilitated (Andrews, 1989, 1992; Forster & Shen, 1996; Sears, Hino, & Lupker, 1995; but see Carreiras, Perea, & Grainger, 1997; Coltheart et al., 1977), but high-frequency orthographic neighbors inhibit performance (Carreiras et al., 1997; Grainger, 1990; Grainger & Jacobs, 1996; Grainger, O'Regan, Jacobs, & Segui, 1989, 1992; Grainger & Segui, 1990; Perea & Pollatsek, 1998). These effects are dependent on various factors (Andrews, 1997; Grainger & Jacobs, 1996), including interactions with task and language.

Here, orthographic neighbors had a facilitation effect, consistent with the literature. The orthographic-neighbor effect resisted the introduction of controls, suggesting this effect was a true higher-level effect.

Implications for phonology

Overall, homophones impair performance: (a) less frequent words in homophonic pairs get inhibited (Davelaar, Coltheart, Besner, & Jonasson, 1978; Rubenstein, Lewis, & Rubenstein, 1971; but see Clark, 1973; Coltheart et al., 1977), (b) pseudo-homophones are harder to reject than other non-words (Coltheart et al., 1977; Rubenstein et al., 1971), and (c) higher misclassification rates for words into the category of its homophone (e.g., "rows" misclassified as a flower; Daneman & Reingold, 1993; Daneman, Reingold, & Davidson, 1995; Jared, Levy, & Rayner, 1999; Peter & Turvey, 1994; Van Orden, 1987; Van Orden, Johnston, & Hale, 1988; Van Orden et al., 1992; Ziegler & Jacobs, 1995; Ziegler, Van Orden, & Jacobs, 1997). Some of these effects are small, observable mainly with pseudo-words, present mostly in low-frequency words, and/or possibly dependent on strategy (Clark, 1973; Coltheart et al., 1977; Coltheart, Patterson, & Leahy, 1994; Davelaar et al., 1978; Jared & Seidenberg, 1991; McCusker, Hillinger, & Bias, 1981; Seidenberg, 1985a, 1985b).

Our results show inhibition on words with more phonological neighbors, but only when the effect of orthographic neighbors is controlled for. The direction and fragility of these effects is consistent with the literature. This suggests that extra processing is required to discriminate between phonologically close words.

Implications for word type and word length

Eye movements recorded during reading (O'Regan, 1990; Rayner, 1998; Rayner & Pollatsek, 1989; Rayner & Sereno, 1994) show that participants make more fixations on longer words (100% for eight or more letters vs. 25% for two to three letters), names, verbs, and adjectives (85%), and fewer on prepositions, conjunctions, articles, and pronouns (35%).

Our results are consistent with a strategy of moving noise away from words that are important for understanding sentences (i.e., noise increased on auxiliary verbs & articles, and decreased on verbs & adjectives). Our analyses suggest that the word-type effect is partly or mainly due to lower-level statistics. For example, word types may be statistically different in complexity or frequency, and those differences may underlie the word type effect.

We also found a word-length effect in the form of greater noise for single-letter words relative to longer words. This suggests that longer words may be fixated more because they extend beyond the visual span (Legge, Cheung, Yu, Cheung, Lee, & Owens, 2007; Pelli et al., 2007) or are more likely to be important, and not because they are somehow more masked or noisy.

Conclusion

Our study is consistent with the literature on (a) optimal viewing position, (b) letter complexity, (c) syllable frequency, and (d) orthographic and phonological neighbors. We did not replicate effects of lexical frequency or word length, and it is unclear if this was due to lack of statistical power, or differences in methodology. There are three novel conclusions from our study: (a) the word type effect may reflect different word statistics across word types (e.g., length, letters), (b) letter degradation techniques have similar but not fully equivalent effects, and (c) the alphabet may have been more optimized for writing rather than for reading, given that frequent and diagnostic letters were also easily confusable.

The efficiency of the method for data collection is noteworthy. This efficiency is gained by combining the advantages of adjustment methods to the ability to make adjustments at any location over a sentence. Essentially, participants were free to concentrate their responses at locations that deviated most from subjectively homogeneous noise, thus quickly reducing apparent noise variations throughout the stimulus. This method quickly converged on responses that were similar across participants that, once analyzed, showed effects similar to those reported in the literature.

Our novel results can be used to generate predictions that can be tested in future work. Specifically, we predict that (a) making diagnostic letters more dissimilar from each other may help improve reading performance, (b) masking easily confusable letters will impair performance more than masking easily recognizable letters in a normal reading task, and (c) some higher-level effects in reading may be at least partially accounted for by low-level word characteristics like letter complexity, diagnosticity, and confusability.

Further studies could extend our method to other factors related to reading. For example, the study could be extended to letter features either by decreasing the spatial scale of the noise spread in a subsequent study, or by using de-blurring techniques on the current data. Moreover, identification of noise resistance (and legibility in general) at a letter- or word-sized spatial scale would be useful in the design of text enhancement systems.

Keywords: visual saliency, reading, visual noise, attention

Acknowledgments

This research was supported by a CIHR grant awarded to Martin Arguin, Frédéric Gosselin, and Dan Bub. Portions of this paper were presented at VSS 2008 (Sarasota, Florida), and Cernec 2008 (Montréal, Canada). We thank anonymous reviewers for comments.

Commercial relationships: none.
Corresponding author: Frédéric J. A. M. Poirier.
Email: jamfpo@yahoo.com; frederic.poirier@umontreal.ca.
Address: Centre de Recherche en Neuropsychologie et Cognition, Université de Montréal, Montréal, Québec, Canada.

References

Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15,* 802–814.

Andrews, S. (1992). Frequency and neighborhood size effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18,* 234–254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulleting & Review, 4,* 439–461.

Appleman, I. B., & Mayzner, M. S. (1982). Application of geometric models to letter recognition: Distance and density. *Journal of Experimental Psychology: General, 111,* 60–100.

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53*(6), 452–471.

Beauvillain, C. (1996). The integration of morphological and whole-word form information during eye fixations on prefixed and suffixed words. *Journal of Memory, & Language, 35,* 801–820.

Bernard, M., Lida, B., Riley, S., Hackler, T., & Janzen, K. (2002). A comparison of popular online fonts: Which size and type is best? *Usability News, 4*(1). Available at http://usabilitynews.org/a-comparison-of-popular-online-fonts-which-size-and-type-is-best/.

Blais, C., Fiset, D., Arguin, M., Jolicoeur, P., Bub, D., & Gosselin, F. (2009). Reading between eye saccades. *PLoS ONE, 4*(7), e6448, doi:10.1371/journal.pone.0006448.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*(241), 177–178.

Bouma, H. (1971). Visual word recognition of isolated lower-case letters. *Vision Research, 11,* 459–474.

Bradley, D. C. (1979). Lexical representation of deprivational relation. In M. Aronoff & M. L. Kean (Eds.), *Juncture* (pp. 37–55). Sarasota, CA: Anma Libri.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 443–446.

Burani, C., & Caramazza,A. (1987). Representation and processing of derived words. *Language & Cognitive Processes, 2,* 217–227.

Burani, C., Salmaso, D., & Caramazza, A. (1984). Morphological structure and lexical access. *Visible Language, 18,* 342–352.

Carreiras, M., Perea, M., & Grainger, J. (1997). Orthographic neighborhood effects on visual word recognition in Spanish: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23,* 857–871.

Changizi, M. A., Zhang, Q., Ye, H., & Shimojo, S. (2006). The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *American Naturalist, 167,* E117–E139.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior, 12,* 335–359.

Clark, J. J., & O'Regan, J.K. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research, 39,* 843–857.

Colé, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory & Language, 28,* 1–13.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). London: Academic Press.

Coltheart, V., Patterson, K., & Leahy, J. (1994). When a ROWS is a ROSE: Phonological effects in written word comprehension. *Quarterly Journal of Experimental Psychology, 47A,* 917–955.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256.

Conrad, M., Carreiras, M., & Jacobs, A.M. (2008). Contrasting effects of token and type syllable frequency in lexical decision. *Language & Cognitive Processes, 23*(2), 296–326, doi:10.1080/01690960701571570.

Daneman, M., & Reingold, E. (1993). What eye fixations tell us about phonological recording during reading. *Canadian Journal of Experimental Psychology, 47,* 153–178.

Daneman, M., Reingold, E., & Davidson, M. (1995). Time course of phonological activation during reading: Evidence from eye fixations. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 884–898.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J.T. (1978). Phonological recording and lexical access. *Memory & Cognition, 6,* 618–632.

Ferrand, L. (2000). Reading aloud polysyllabic words and nonwords: The syllabic length effect reexamined. *Psychonomic Bulletin & Review, 7*(1), 142–148.

Ferrand, L. (2001). *Cognition et Lecture. Processus de base de la reconnaissance des mots écrits chez l'adulte*. Bruxelles: De Boeck.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 696–713.

Gordon, B., & Caramazza, A. (1985). Lexical access and frequency sensitivity: Frequency saturation and open/closed class equivalence. *Cognition, 21*(2), 95–115.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory & Language, 29,* 228–244.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565.

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics, 45,* 189–195.

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception & Psychophysics, 51,* 49–56.

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition. *Perception & Psychophysics, 47,* 191–198.

Grainger, J., Spinelli, E., & Ferrand, L. (2000). Effects of baseword frequency and orthographic neighborhood size in pseudohomophone naming. *Journal of Memory & Language, 42*(1), 88–102.

Haberlandt, K. (1994). Methods in reading research. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1–31). New York: Academic Press.

Holmes, V. M., & O'Regan, J.K. (1992). Reading derivationally affixed French words. *Language & Cognitive Processes, 7,* 163–192.

Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from proofreading and eye movements. *Journal of Experimental Psychology: General, 128,* 219–264.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120,* 358–394.

Legge, G. E., Cheung, S.-H., Yu, D., Chung, S. T. L., Lee, H.-W., & Owens, D. P. (2007). The case for the visual span as a sensory bottleneck in reading. *Journal of Vision, 7*(2):9, 1–15, http://journalofvision.org/7/2/9, doi:10.1167/7.2.9. [PubMed] [Article]

Legge, G. E., Ross, J. A., Luebker, A., & LaMay, J. (1989). Psychophysics of reading. VIII. The Minnesota Low-Vision Reading Test. *Optometry & Vision Science, 66*(12), 843–853.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*(1), 149–157. doi:10.1037/0278-7393.16.1.149.

Mansfield, S., Legge, G. E., & Ban, M. C. (1996). Psychophysics of reading XV: Font effects in normal and low vision. *Investigative Ophthalmology & Visual Science, 37*(8), 1492–1501, http://www.iovs.org/content/37/8/1492. [PubMed] [Article]

McCusker, L. X., Hillinger, M. L., & Bias, R. G. (1981). Phonological recoding and reading. *Psychological Bulletin, 89,* 217–245.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Lawrence Erlbaum.

Nazir, R. A., Jacobs, A. M., & O'Regan, J. K. (1998). Letter legibility and visual word recognition. *Memory & Cognition, 26,* 810–821.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers, 36*(3), 516–524.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique, 101,* 447–462, http://www.lexique.org.

Nothdurft, H. C. (1993). The conspicuousness of orientation and motion contrast. *Spatial Vision, 7,* 341–363.

Nothdurft, H.-C. (2000). Salience from feature contrast: Additivity across dimensions. *Vision Research, 40,* 1183–1201.

O'Regan, J. K. (1990). Eye movements and reading. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes*. (pp. 395–453). New York: Academic Press.

O'Regan, J. K. (1992). Facteurs sensoriels et moteurs dans la lecture: La position optimale de regard. In P. Lecocq (Ed.), *La lecture: Processus, apprentissage, troubles* (pp. 11–28). Lille, France: Presses Universitaires de Lille.

O'Regan, J. K., & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception & Performance, 18,* 185–197.

O'Regan, J. K., & Levy-Schoen, A. (1987). Eye movement strategy and tactics in word recognition and reading. In M. Coltheart (Ed.), *The psychology of reading: Attention and performance XII* (pp. 363–383). Hillsdale, NJ: Lawrence Erlbaum.

O'Regan, J. K., Levy-Schoen, A., Pynte, J., & Brugaillère, B. (1984). Convenient fixation location within isolated words of different lengths and structure. *Journal of Experimental Psychology: Human Perception & Performance, 10,* 250–257.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46,* 4646–4674.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12):12, 1136–1169, http://journalofvision.org/content/8/15/14, doi:10.1167/4.12.12. [PubMed] [Article]

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision, 7*(2):20, 1–36, http://journalofvision.org/content/7/2/20, doi:10.1167/7.2.20. [PubMed] [Article]

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception & Performance, 24*(3), 767–779.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Peter, M., & Turvey, M. T. (1994). Phonological codes are early sources of constraint in visual semantic categorization. *Perception & Psychophysics, 55*(5), 497–504.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56–115.

Poirier, F. J. A. M., Gosselin, F., & Arguin, M. (2008). Perceptive fields of saliency. *Journal of Vision, 8*(15):14, 1–19, http://journalofvision.org/content/8/15/14, doi:10.1167/8.15.14. [PubMed] [Article]

Rastle, K., & Coltheart, M. (1998). Whammy and double whammy: Length effects in nonword naming. *Psychonomic Bulletin & Review, 5,* 277–282.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychonomic Bulletin, 124*(3), 372–422.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Rayner, K., & Sereno, S. C. (1994). Eye movements in reading: Psycholinguistic studies. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 57–81). New York: Academy Press.

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition, 68,* B71–B80.

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition, 74,* 1–12.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. (1971). Evidence for phonemic coding in visual word recognition. *Journal of Verbal Learning & Verbal Behavior, 10,* 645–657.

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review, 81,* 99–118.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 21,* 876–900.

Segui, J., Frauenfelder, U. H., Lainé, C., & Mehler, J. (1987). The word frequency effect for open- and closed-class items. *Cognitive Neuropsychology, 41*(1), 33–44.

Seidenberg, M. S. (1985a). Constraining models of word recognition. *Cognition, 20,* 169–190.

Seidenberg, M. S. (1985b). The time course of information activation and utilization in visual word recognition. In D. Besner, T. Waller, & G. Mackinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 199–252). London: Academic Press.

Senécal, M.-J. (2001). *Version française du Minnesota Low Vision Reading Test* (Unpublished master's thesis). Université de Montréal, Montréal; Institut Nazareth et Louis-Braille, Longueuil.

Senécal, M. J., Gresset, J., & Overbury. O. (2002). Version française du Minnesota Low-Vision Reading Test. In J. Gresset, & P. Rondeau (Eds.), *Mesures évaluatives: Du développement aux applications: 5e Symposium scientifique sur l'incapacité* *visuelle et la réadaptation* (pp. 31–34). Montréal: Université de Montréal; Longueuil: Institut Nazareth et Louis-Braille.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7,* 263–272.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory & Cognition, 15,* 181–198.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14,* 371–386.

Van Orden, G. C., Stone, G. O., Garlington, K. L., Markson, L. R., Pinnt, G. S., Simonfy, C.M., et al. (1992). "Assembled" phonology and reading: A case study in how theoretical perspective shapes empirical investigation. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 249–292). North Holland: Elsevier Science Publishers.

Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception & Psychophysics, 47*(6), 583–600.

Ziegler, J. C., & Jacobs, A. M. (1995). Phonological information provides early sources of constraint in the processing of letter strings. *Journal of Memory & Language, 34,* 567–593.

Ziegler, J. C., Van Orden, G. C., & Jacobs, A. M. (1997). Phonology can help or hurt the perception of print. *Journal of Experimental Psychology: Human Perception & Performance, 23*(3), 845–860.

Zipf, G. K. (1935). *The psycho-biology of language*. Boston: Houghton Mifflin.

Appendix A

Details of factors

All frequency data was taken from Lexique 3.55 (New, Pallier, Brysbaert, & Ferrand, 2004; New, Pallier, Ferrand, & Matos, 2001).

Frequencies

Frequency factors are often log-transformed before analyses, for the purpose of reducing the weight of high-frequency items. Here, frequency factors were only log-transformed when doing so increased stability of results across analyses (i.e., showed fewer changes when entering or removing other factors) and/or decreased individual variability.

Word frequency statistics included several factors: (a) lexical frequency (M = 1,243 per million; SD = 3,788; min. < 1; max. = 38,929; power law: R^2 = 90.5%; exponent = –2.34), (b) cumulative word frequency, i.e., lexical frequency summed across words with the same lemma (M = 1,833 per million; SD = 4,370; min. < 1, max. = 38,929; power law: R^2 = 86.7%; exponent = –2.30), (c) number of homographs, i.e., other words with same spelling as target (M = 1,688; SD = 0,827; max. = 5), (d) number of homophones, i.e., other words with same pronunciation as the target (M = 5,038; SD = 3,628; max. = 24), (e) orthographic neighbor frequency, i.e., summed frequency of other words with the same number of letters as the target that differ by a single letter (M = 4,974; SD = 5,254; max. = 25), and (f) phonological neighbor frequency, i.e., summed frequency of other words with the same number of graphemes as the target that differ by a single grapheme (M = 25,052; SD = 32,458; max = 127).

Complexity

The factor complexity (or perimetric complexity) was defined as contour length squared divided by ink area (Attneave & Arnoult, 1956; Pelli, Burns, Farell, & Moore-Page, 2006), with contour length simplified to only include intensity changes due to horizontal and vertical edges, because line thickness was one pixel.

Letter width

Letter width was the horizontal extent of letters, in pixels, plus one pixel for interletter spacing.

Confusability

Confusability measures (i.e., in noise or low contrast) were taken from a study conducted in our lab (Fiset et al., unpublished observations), which used the same font as used here, although they used a finer spatial sampling. Confusability corresponds to the proportion of trials on which a particular letter, briefly exposed (16.67 ms), was reported incorrectly. Average accuracy was maintained at 50% across letters by either adding noise or lowering contrast.

Letter frequency

Letter frequency was taken from a published dataset for written text in French (Lexique 3.55; New et al., 2001; New et al., 2004).

Letters that appeared fewer than 10 times in the 30 sentences (i.e., "k" and "w" to "z") were not included in correlations with noise resistance. Those letters are included in all other correlations, and correlations with or without them were similar.

Diagnosticity

Diagnosticity was a measure of word confusability increase when a given letter cannot be recognized. Diagnosticity was calculated using the Lexique 3.55 database and an algorithm that: (1) selected a target letter and a replacement letter, (2) changed every instance of the target letter in the word list by the replacement letter, simulating letter confusion, (3) counted the number of identical words in the modified list, simulating word confusion due to letter confusion, (4) repeated Steps 1 through 3 for every target and replacement combination possible, and stored the results into a matrix, (5) subtracted from the matrix the number of words that are identical if no letter was changed, thus isolating word confusion due to letter confusion, (6) summed the matrix for a target letter across replacement letters to give the diagnosticity of that letter, i.e. the extra confusability due to the target letter being misread, and (7) normalized to the 0 to 1 range for the regression analysis. Diagnosticity calculated by factoring in word frequencies or letter frequencies produced similar results, thus only the unweighted diagnosticity is discussed here.

Sublexical factors

Sub-lexical statistics included the following factors (statistics in our samples shown in parentheses): (1–2) letters per grapheme ("grapheme length"; $M = 1.44$; $M = 0.36$; min. = 1; max. = 3) or syllable ("syllable length"; $M = 3.6$; $SD = 1.2$; words had an average of 1.638 syllables, $SD = 0.758$; min. = 1; max. = 4), and (3–4) syllable frequency by type, i.e., the number of syllabic neighbors, and by token, i.e., the sum of the lexical frequencies of all syllabic neighbors ($M = 1{,}096$ and 4,646 per million; $SD = 1{,}524$ and 7,478; max. = 8,160 and 49,857 for token and type, respectively).

Word type

Word type included the following categories, using sentence context to determine category (frequency of occurrence in our stimuli indicated in parentheses): (a) auxiliary verbs (1.7%); (b) conjunctions (0.9%); (c) articles (2.6%); (d) pronouns (5.5%); (e) verbs (19.2%); (f) prepositions (4.1%); (g) nouns (42.6%); (h) determiners (3.2%), including quantifiers, demonstrative adjectives, and possessive adjectives; (i) adverbs (5.0%); and (j) adjectives (15.2%). These categories were mutually exclusive.

Appendix B

	Confusability in					
	Noise (%)	Contrast (%)	Complexity	Log letter frequency	Noise resistance	Diagnosticity
a	67.0	71.8	128.8	8.73	11.15	14,293
b	45.3	33.0	134.6	1.77	10.78	7,482
c	44.5	44.5	117.6	3.70	11.11	9,699
d	42.8	45.5	134.6	2.39	10.61	7,345
e	65.5	85.3	130.7	10.56	10.87	25,110
f	69.0	58.8	101.9	1.30	11.24	6,450
g	36.3	21.0	165.1	1.97	10.43	6,179
h	41.3	39.5	121.5	1.94	10.43	2,552
i	84.8	93.3	73.1	8.49	11.14	12,310
j	68.8	73.3	100.0	0.25	10.91	1,526
k	23.8	31.3	134.6	0.53		651
l	85.0	90.8	72.3	4.72	11.28	10,935
m	18.0	12.8	145.2	2.92	10.67	9,602
n	40.0	47.3	109.7	7.07	11.13	12,556
o	67.8	64.5	115.2	6.26	10.99	6,259
p	28.8	20.0	142.4	2.48	10.39	9,305
q	47.5	45.0	142.4	0.55	10.98	489
r	60.8	66.3	98.0	7.78	11.42	20,186
s	48.8	58.8	149.3	7.23	11.31	34,204
t	86.8	85.8	94.1	6.69	11.48	19,596
u	43.5	42.3	101.9	4.17	11.05	5,520
v	51.5	56.0	135.5	1.20	11.31	6,921
w	12.5	13.5	167.5	0.22		115
x	46.8	55.0	171.4	0.35		845
y	13.3	14.0	171.4	0.73		1,731
z	27.5	27.8	128.8	0.46		4,819

Letter statistics. *Notes*: Letter statistics used in the regression model, as well as mean noise resistance for lowercase letters that appeared at least 10 times in the experiment.

Appendix C

Implications for letter processing

The pattern of correlations between letter-based variables (Table 3) suggests that the alphabet is more optimal for writing than for reading. This argument is presented below in three parts: (a) the results are reviewed from the perspective of *reading* constraints, (b) the results are reviewed from the perspective of *writing* constraints, and (c) the results are contrasted to those of Changizi, Zhang, Ye, and Shimojo (2006) who argue, based on topology distributions, that alphabets are developed based on visual constraints rather than motor constraints.

An alphabet designed to be easy to *read* would focus on making diagnostic letters (especially frequent ones) easier to read, i.e., more efficient and less confusable. Supporting this, we found that frequent and diagnostic letters were more efficiently perceived as measured by noise resistance (see also Pelli et al., 2006). However, in French at least, frequent and diagnostic letters are unfortunately easily confused with other letters (Table 3), making the alphabet suboptimal for reading. We propose that efficiently perceived simple letters are more easily confused together because they lack distinctive features (see below; Attneave & Arnoult, 1956; Pelli et al., 2006).

An alphabet designed to be easy to *write* would focus on making frequent letters simple, so that writing speed would be improved. Supporting this, frequent letters were less complex. Although not directly relevant to writing performance, frequent letters tend to be more diagnostic and easily confusable (e.g., "i," "l," compared to "m," "w," "y"). That is, by meeting writing constraints, reading constraints were not optimized. Thus in designing an alphabet that is both easy to write *and* to read, the emphasis should be on making diagnostic and frequent letters both easy to encode (e.g., high noise resistance, low complexity) and easy to discriminate from each other (e.g., low confusability).

The only study that looked empirically at whether alphabets were designed based on reading or writing constraints is a study comparing the topology distribution of alphabets, shorthand, trademarks, and natural scenes (Changizi et al., 2006). They measured the relative frequency of different topologies within a stimulus set (e.g., alphabets, trademarks, natural scenes), and compared topology distributions across stimulus sets. Their conclusions can be summarized as follow: (a) alphabets have topology distributions similar to trademarks and natural scenes, suggesting that they are influenced by visual constraints, (b) alphabets have a topology distribution that is dissimilar to that of shorthand, suggesting that motor constraints play a much lesser role in alphabet design, and (c) alphabets have a topology distribution that differs from scribbles or random lines, suggesting that alphabets are not random. Their conclusion conflicts with ours. However, we reconcile the diverging views below.

Changizi et al. (2006) did their analysis over a collection of alphabets, thus they lack single-letter data to compare with. This issue is critical because the assignment of letter shapes to sounds may itself be consistent with motor constraints, consistent with our results. Early alphabet designers likely associated simple symbols to frequent sounds first, and then created symbols of increasing complexity for increasingly infrequent sounds. Although they made symbols generally easy to produce and discriminate from each other, they likely assigned simpler symbols to frequent sounds. Thus, although Changizi's account may be accurate at the alphabet level, ours may be an adequate description at the letter level.

To further this argument, Changizi et al.'s (2006) study did not account or control for the frequency of occurrence of letters. As shown above, simple letters are frequent. Correcting their data for this bias would increase frequencies at low-complexities. This could make their results consistent with alphabets being influenced by motor constraints.

Appendix D

Ecological validity

Every method to study reading has its strengths and weaknesses, and none is perfect (Haberlandt, 1994; Perfetti, 1985). The visual spread method used here is an efficient way of measuring noise resistance at a relatively fine spatial scale. An hour and a half of testing per participant was sufficient to extract all the effects discussed above, including effects related to single-letters, letter position, word type, and syllable frequencies. This technique also required no specialized equipment such as an eye tracker or a voice key. However, caution is advised when generalizing from our task to other reading tasks, for several reasons.

Firstly, we did not ask participants to read sentences, and we did not measure reading comprehension. Reading effects are generally robust across methods, including passive-viewing tasks (e.g., fMRI experiments). Nevertheless, participants may sometimes adjust noise using visual factors, which may be unrelated or indirectly related to reading. For example, our results are consistent with complexity and diagnosticity, but could be more consistent with confusability if reading performance had been measured. In the presence of conflicting factors influencing noise, participants may not always spread noise to improve readability.

Secondly, participants had ample opportunity to extract the full content of a sentence over the duration of the study (average of 3 min spent per sentence). Mechanisms that are time-sensitive may require speeded-response tasks to produce measurable effects. Moreover, our method may not capture well effects that are dependent on surprise or uncertainty (e.g., when participants are asked to search for specific information or spelling errors).

Thirdly, the stimuli were degraded. Polarity was reversed in 12.5% of pixels, effectively changing letter shape. However, our noise levels were intermediate between studies using no noise and those using high levels of noise, such that reading sentences remained easy.

Overall, our results are consistent with the literature. However, both the similarities and differences of results across different methodologies can be informative. Similar results inform of effects that are robust across task demands and stimuli. Different results inform of effects that are sensitive to task demands and stimulus characteristics. By integrating results across methodologies and task demands, we can gain a better understanding of the underlying mechanisms involved.

In particular, our method differs from many other methods in the way the sentences were generated. A common research method is to generate word lists that exaggerate natural differences on few experimental factors while minimizing differences on control factors. These lists can nevertheless include sampling biases, especially with regards to other factors not considered in the selection criteria. The approach taken here respects natural frequencies and word statistics, and uses a statistical control instead. Our word sampling method is representative of normal texts, but it undersamples infrequent word types, word lengths, letters, etc. It is thus possible that some factors above would be significant with a larger sample of low-frequency words. This is not a weakness specific our study, but rather a general concern in research when attempting to generalize across studies using natural or controlled samples.

Statistical techniques do offer advantages, namely: (a) ability to control for many factors, (b) ability to easily include low-level factors such as letter complexity or syllabic length, and (c) higher statistical sensitivity to effects that are normally masked by other factors also influencing performance. Our study raises the possibility that certain effects commonly attributed to word type could in fact be due low-level factors. A follow-up study could confirm whether word type, even controlled for low-level factors, would still impact reading performance.