2010

A New Perspective on Visual Word Processing Efficiency

Joseph W. Houpt
Wright State University - Main Campus, joseph.houpt@wright.edu

James T. Townsend

A New Perspective on Visual Word Processing Efficiency

Joseph W. Houpt (jhoupt@indiana.edu) and **James T. Townsend (jtownsen@indiana.edu)**
Indiana University, Department of Psychological and Brain Sciences
1101 E. Tenth Street, Bloomington, IN 47401 USA

Abstract

As a fundamental part of our daily lives, visual word processing has received much attention in the psychological literature. Despite the well established perceptual advantages of word and pseudoword context using accuracy, a comparable effect using response times has been elusive. Some researchers continue to question whether the advantage due to word context is perceptual. We use the capacity coefficient, a well established, response time based measure of efficiency to provide evidence of word processing as a particularly efficient perceptual process to complement those results from the accuracy domain.

Keywords: Visual word perception; Efficiency; Workload capacity.

As a fundamental part of our daily lives, visual word processing has received much attention in the psychological literature. However, the interest in visual word perception extends beyond its value in communication. The written word is a complex stimulus with which most adults have a large amount of experience. Unlike faces, there is no reason to believe we have any innate ability to perceive words. Thus, word perception may represent the limit of perceptual learning in the absence of innate ability.

Due to the relative ease with which most adults read, it is reasonable to assume that word perception is an efficient process. This is further supported by the intuition that with more experience with a process we become more efficient and we are quite experienced with the written word. Often, the efficiency is measured using single letter perception as a base line. When word context offers an advantage in the accuracy or processing time of perceiving a letter, this supports the claim that word perception is efficient.

From the early days of experimental psychology, researchers have been interested in the value of a word context for perceiving letters. In one study, letters were displayed sequentially to participants at faster and faster rates until they could no longer correctly identify the letters. They found that participants maintained accuracy with shorter durations when the letters were presented as part of a word compared with random letter sequences (Cattell, 1886).

One problem with studies of this nature is that they do not control for the constraint on possible letters that a word context puts on the possible letters. Hence it is not clear from those early results whether the advantage is a perceptual advantage or a decisional advantage. In the late 1960's an alternative task was designed to eliminate the decisional advantage of word context so as to examine the perceptual effects. In this task a letter or word was tachistoscopically displayed to a participant. They then chose from two possible choices, one of which was correct. In the letter condition, the choices were letters. In the word condition, both choices were words that differed in only a single letter. Since both alternatives were words, the word context was no longer informative as to the identity of the letter. Participants were still more accurate at perceiving letters in the word condition than the letter condition (Reicher, 1969). Furthermore, they found that participants are also more accurate with word contexts than random letter sequence contexts. This is known as the word superiority effect. An efficiency gain of context over letters alone is not unique to words though. If a sequence of letters conformed to the pronunciation rules of English, strings referred to as pseudowords, then participants were again more accurate than letters alone (e.g., McClelland & Johnston, 1977). This is known as the pseudoword superiority effect.

Despite the robustness of the word and pseudoword superiority effects, a comparable effect using response times (and controlling for decisional information due to context) has been elusive. This may be in part explained by the possibility that people will read an entire word even if the task does not require it. Indeed, this has been put forth as further evidence that word perception is special (LaBerge & Samuels, 1974). One of the goals of this paper is to demonstrate a response time based word superiority effect, and possibly a pseudoword superiority effect as well.

Even in the accuracy domain, some researchers continue to question whether there is a *perceptual* advantage due to word context. For example, Pelli, Farell, and Moore (2003) demonstrated evidence for a model of word perception in which letters are perceived independently and with separate detection decisions on each letter. Their evidence comes from comparing the efficiency of word perception as the number of letters in the word increases. Depictions of longer words have more information about their identity, since the more letters that are known, the fewer possibilities there are for the others. Hence, if a person is able to take advantage of this global information, they should need less per letter information as the number of letters increases. However, a model of word perception based on independent, separate decisions on the letters predicts that as the word length increases, the reader will still need the same amount of information per letter to maintain accuracy. In fact participants did need roughly the same amount of per letter information as the number of letters increased, supporting the latter model.

In the next section we describe the capacity coefficient, a response time based measure of efficiency. We propose that this measure, along with a task that controls for both the available information and possibly mandatory word reading, provides evidence of word processing as a particularly efficient process to complement those results from the accuracy domain.

	Target	Distractors				Single Character							
Word	care	bare	cure	cave	card	c	b	a	u	r	v	e	d
Pseudoword	lerb	nerb	larb	lemb	lerf	l	n	e	a	r	m	b	f
Non-Word	rlkf	vlkf	rtkf	rlhf	rljk	r	v	l	t	k	h	f	k
Upside-down	ɹlʞɟ	ʌlʞɟ	ɹʇʞɟ	ɹlɥɟ	ɹlɾʞ	ɹ	ʌ	l	ʇ	ʞ	ɥ	ɟ	ʞ
Katakana	サイクオ	ヘイクオ	サナクオ	サイフオ	サイクノ	サ	ヘ	イ	ナ	ク	フ	オ	ノ

Table 1: Stimuli used for capacity analysis.

The Capacity Coefficient

The capacity coefficient, $C(t)$ is an established response time based measure of the effect of increased load on processing efficiency (Townsend & Nozawa, 1995; Townsend & Wenger, 2004). Specifically, $C(t)$ is a measure of the change in processing rates as the task requires attention to more targets, or possibly more dimensions of a single target. The basic idea of the measure is to compare response times when reading the full string to the times that would be predicted if each character took the same amount of time, whether or not it was in a string.

The capacity function for an exhaustive task is defined using the natural log of the cumulative distribution function, $K(t) = \ln F(t); F(t) = \Pr\{\mathrm{RT} \leq \mathrm{t}\}$, and is similar to the cumulative hazard function used in survival analysis. If K_{c1} is the cumulative hazard for the first character response times, K_{c2} is the cumulative hazard for the second character, etc., and K_S is the cumulative hazard for the string condition, the capacity coefficient is given by $C(t) = \left[\sum_{i=1}^{4} K_{c_i}\right] / K_S$.

This formulation is based on the predictions of the unlimited capacity, independent, parallel (UCIP) model. The assumptions of the UCIP model are sufficient conditions for there to be no change in the rates of processing with increased load. If these assumptions hold then the relationship between the processing times of the string to the processing times of the individual characters is as follows:

$$\Pr\{RT_S \leq t\} = \Pr\{RT_{c_1} \leq t\}\Pr\{RT_{c_2} \leq t\}\Pr\{RT_{c_3} \leq t\}\Pr\{RT_{c_4} \leq t\}$$

By taking the natural log of both sides of this equation, then dividing by the left hand side, we see that the UCIP model predicts $C(t) = 1$ for all $t \geq 0$. This gives us a baseline for comparison. If a person performs better than the baseline model, $C(t) > 1$, their performance is referred to as super-capacity. There are multiple ways performance could be super-capacity. For example, if there is facilitation between the characters, or in more extreme cases if the information from the characters is accumulated together toward a single decision (Townsend & Wenger, 2004).

Performance worse than the baseline model, $C(t) < 1$, is limited-capacity. In contrast to the case of super-capacity, inhibition between characters could result in limited-capacity. A standard serial model (independent and unlimited capacity) also predicts limited capacity. Furthermore, if the system only has a certain amount of resources to dedicate to the task, limited capacity performance would be expected. For example, a fixed capacity parallel model (a finite amount of resources is evenly divided up among the current processes) also predicts limited capacity.

When performance is about the same as the baseline model, $C(t) \approx 1$, then we refer to it as unlimited capacity. Of course this would happen if all of the assumptions of the baseline model were met. Alternatively some blend of features that lead to limited capacity and features that lead to super-capacity could balance out and result in capacity around 1. It is not likely that people are truly unlimited-capacity or super-capacity in the extreme case of very long words, but it is reasonable for shorter words.

The capacity coefficient measures processing efficiency in isolation by comparing the capacity coefficient to predicted values of unlimited capacity, independent, parallel models. Thus, this measure also allows us to compare the efficiency of a variety of processes despite any possible differences in difficulty due to component processes. In particular, we are able to draw conclusions about the efficiency of word processing relative to pseudoword, non-word, upside-down non-word, and unfamiliar character string processing.

Method

To properly compare perceptual efficiency across words, pseudowords, non-words, upside-down words and unfamiliar characters, our task must eliminate the extra information available given a word context. Furthermore, the possibility that words are exhaustively processed automatically may lead to a disadvantage for words on response time measures. To address these issues, we adapted a task from Blaha, Busey, and Townsend (2009) which forces exhaustive processing of the characters in a string. This experiment consists of two components. First, we measure the participants response times to correctly identifying the target string. To ensure that participants base their identification on the entire string and not any subset, we include a distractor of a string with a single character different in each position in the string. For example if the target is ”care” then “bare,” “cure,” “cave” and “card” are used as distractors (see Table 1). Second, the participants distinguish between letters in isolation. Whereas in the exhaustive case the participant needed to distinguish between “bare” and “care,” we now only require them to distinguish between “b” and “c.” The response times on these tasks are used for computing the predicted performance of the UCIP model.

Stimuli Stimuli were created using GIMP version 2.2. For each stimulus, the character or characters were written in black in 29pt Courier onto a gray (200) background. Then each stimulus was doubled in size. There were five types of stimuli used: words, pronounceable non-words (pseudowords), unpronounceable non-words, upside-down unpronounceable non-words, and strings of Katakana characters. All strings used were four characters long. Words were chosen so that the frequency of the target was roughly equal to the average frequency of the distractors. Pseudowords were taken from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). Table 1 summarizes the stimuli used for both the single character and exhaustive trials for each type.

Participants Participants were recruited from the Indiana University population. Eight females and two males participated in this study, all of whom were native English speakers and reported that they did not read or speak Japanese. Their ages ranged from 19-34. All participants reported having normal or corrected to normal vision, no difficulty reading English, and no prior diagnoses of a reading disorder. Participants were paid \$8 per session, and received a \$20 bonus upon completion of all sessions. Each session lasted between 45 and 60 minutes.

Procedure All experimental conditions were run using DMDX version 2.9.06 developed at Monash University and at the University of Arizona by K.I.Forster and J.C.Forster. Stimuli were presented on a 17 Dell Trinitron CRT monitor running in 1024x720 mode. Participants used a two-button mouse for their responses.

Each session was dedicated to a type of stimuli and there were ten total sessions, two sessions for each type. At the beginning of each session, we read the participant the general instructions for the task while those instructions were presented on the screen. The instructions encouraged participants to respond as quickly as possible while maintaining a high level of accuracy. Each session was divided into five blocks, one block of string stimuli and a block for each of the corresponding single character stimuli. Each block began with a screen depicting the button that corresponded to each of the categories. Participants had 40 practice trials, 20 of each category. Next participants were given 240 trials divided evenly between the two categories, the first 40 of which were not used in the analysis. Each trial began with a 30 ms presentation of a fixation cross. After a random delay (300-600 ms), the stimulus was presented for 80 ms. Participants had a maximum of 2500 ms to respond. If the participant responded correctly, the next trial started after a 400 ms delay. If the participant responded incorrectly, a tone was played then the next trial started after a 400 ms delay. The session order was counterbalanced among the participants so that participants completed the different types on different days and in different orders.

Analysis All data was analyzed using R. Analysis was limited to the correct responses on the target category. We computed a repeated measures ANOVA of the response times in each condition. We then calculated the AND capacity coefficient, $C(t)$ for each participant and each condition. For each capacity coefficient, bootstrapped confidence intervals (95%) were calculated to determine if the function was reliably above or below 1 for any length of time.

To facilitate comparison between conditions, we analyzed the capacity functions using functional principal component analysis (fPCA, Ramsay & Silverman, 1997). In fPCA, each capacity function is treated as a weighted linear combination of a common set of basis functions. The set of weights is specific to each function and are therefore an alternative representation of the individual function. Similar to multivariate PCA, the basis functions each explain some percentage of the variance in the data. By treating those basis functions that explain only small amounts of variance as noise, we can achieve a concise representation of our data in terms of just the weights. The justification for applying fPCA is essentially the same as standard PCA; further details can be found in Ramsay and Silverman (1997).

To apply fPCA to capacity coefficient functions, we first calculated the empirical $K(t)$ by taking the natural log of the empirical cumulative distribution functions. Each of those functions were then interpolated using monotone cubic interpolation. The capacity coefficient for each condition was then calculated for each condition using those estimated K functions, then registered by aligning the median of each participants response time data in each condition. Functional principal components was then applied to the smoothed and registered functions, with the data weighted by the overall density function of the response time data and factor scores were computed based on a varimax rotation.

Results

Using a repeated measures ANOVA we found significant effects of condition $[F(4,18713) = 937.75, p < 2.2e-16]$ and whether the stimulus was a target for each of the string stimuli $[F(1,18713) = 10.75, p = 0.001]$, along with a significant interaction effect $[F(4,18713) = 57.73, p < 2.2e-16]$. Eight of the ten participants were faster on words and pseudowords than the other two conditions. Participant 6 was fastest on non-words while Participant 1 was fastest with words, but faster with non-words and upside-down non-words than pseudowords. Eight of the ten participants were slowest with Katakana, while Participant 7 was slowest with non-words and Participant 9 was slowest with upside-down words. While these results are interesting, this level of analysis does not account for the varied difficulty of processing each of the components. Hence, we turn to the capacity coefficient.

The results for the capacity analysis are shown in Figure 1. Bootstrap confidence intervals were used to determine significance, but are not included due to space limitations. Significance in comparisons against the UCIP model was determined by overlap of 99% confidence intervals with $C(t) = 1$.

In the word condition, nine participants had capacity coef-

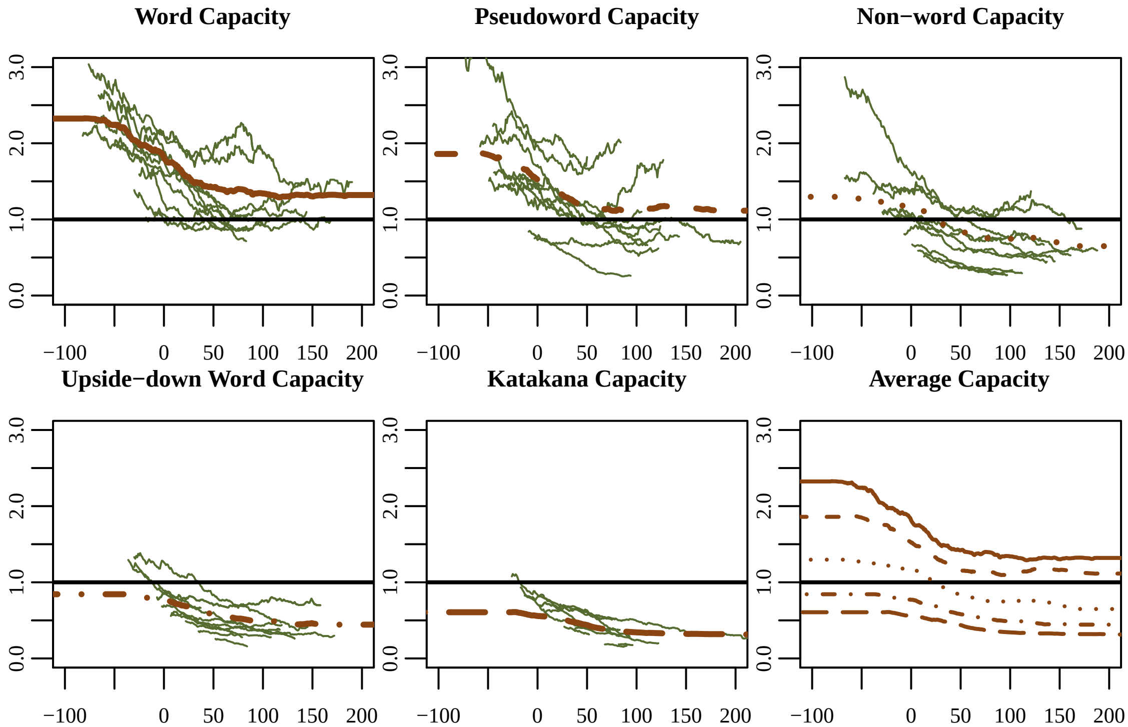

Figure 1: Capacity Coefficient values across time. Thin lines represent individual participant data and thick lines are the mean function across participants. The bottom right panel contains the mean function for each condition together.

ficient values significantly above one for at least some time (1-3, 5-10) and no participant had capacity values below 1 for any time. In the pseudoword condition, eight participants were super-capacity for some time (1, 2, 5-10), two of whom were limited capacity at later times (7, 8). The other two were limited capacity most of the time (3, 4). Three participants were super-capacity for some time in the non-word condition (1, 6, 8), one of whom was limited capacity for later times (8). The other seven participants were limited capacity for some time (2-5, 7, 9, 10), six of whom were limited capacity most of the time (3-5, 7, 9, 10). In the upside-down non-word condition, all participants were limited capacity for most times, with only three participants significantly super-capacity for early times (6-8). All participants were limited capacity for all times in the Katakana condition.

The bottom right panel of Figure 1 shows the capacity function averaged across participants for each condition. The average word capacity was the highest, followed by pseudoword, non-word, upside-down non-word, and lastly Katakana.

Figure 2 depicts the first principal component function of the capacity coefficients. This demonstrates that the feature that best distinguishes performance is a relatively large change in magnitude at early times, tapering off to a slight opposite change at later times. This first principal component describes 94% of the variance in the capacity functions. Furthermore, the condition was found to be a significant predictor of the factor score on this component in a repeated measures ANOVA $[F(4,36) = 18.56, p < 2e-8]$.

Discussion

Due to space limitations, we limit the majority of our discussion to the word and, to a lesser extent, the pseudoword results. We have demonstrated clear evidence of super-capacity processing of the word stimuli for nine of the ten participants. These participants are efficiently perceiving the whole word in comparison to individual letter perception. As mentioned earlier, evidence for the word superiority effect has been difficult to demonstrate with response times. These findings provide that evidence and thus agree with the majority of the word perception literature based on accuracy results. Based on comparisons across conditions, it is also clear that the word perception was more efficient than non-word, upside-down word, and strings of Katakana perception, findings that again match with the results reported for accuracy (e.g., McClelland & Rumelhart, 1981).

There is also evidence for a pseudoword superiority effect, another well established effect in the accuracy domain (McClelland & Johnston, 1977). Although the evidence was not as consistent as the word results, eight of the ten participants were super capacity for some time, with only two participants showing significantly limited capacity processing for most times.

The functional principal components analysis demonstrates that most of the differences in capacity across participants and types of stimuli occur early in the response time

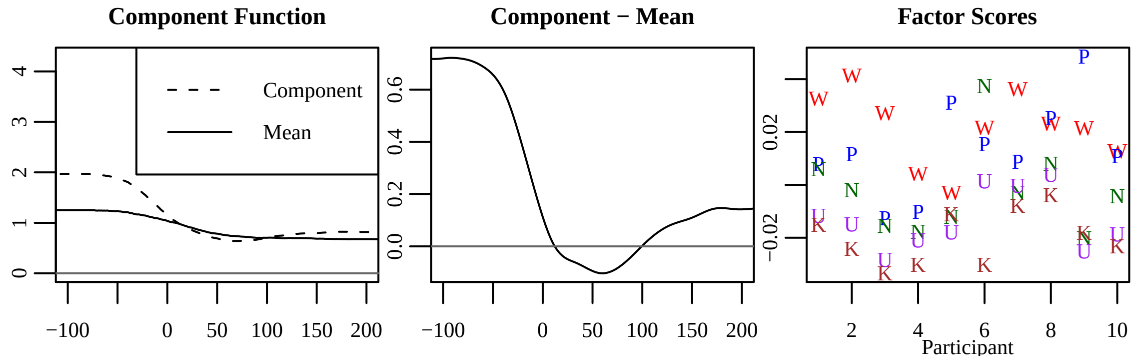

Figure 2: The first principal component of the capacity coefficient functions, which accounts for 94% of the variance. The first panel shows the component function with the mean capacity function for comparison. The second panel shows the difference between the component and the mean functions. The third panel depicts the factor scores on this component for each participant in each condition.

distribution. For all participants except Participant 6, the factor scores are higher for words and pseudowords than any of the other three conditions. This indicates that the word and pseudoword superiority effects are most pronounced in faster response times.

There are multiple plausible explanations for the capacity coefficient results demonstrating particularly efficient processing of words. At least one of the assumptions of the UCIP model must have been violated, so we examine each of those assumptions in turn. Each of these violations have been considered previously for modeling the accuracy based superiority effects.

One assumption that may have been violated is that of independence. If there is any type of facilitation between the letter processes, each letter would be processed faster within a word which would explain the capacity coefficient values above one. There could be many explanations of this facilitation. For example, word processing mechanisms may in fact take advantage of the considerable amount of co-occurrence between letters in English. As is often observed, there are only a fraction of possible four letter combinations used for words and it would be surprising if we did not take some advantage of this reduction in uncertainty. This correlation between letters is an important part of how connectionist models explain the word superiority effect (McClelland & Rumelhart, 1981; Plaut, McClelland, Seidenberg, & Patterson, 1996; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001).

Another, related component of many visual word processing models is the phonological pathway. If a phoneme is activated as a possible interpretation of some letter combination, then it may in turn send positive feedback to those letters, speeding up their processing. Hence a phonological component of visual word processing could also lead to capacity coefficient values above one. Both the correlation between letters and the lack of a regular pronunciation of the non-words imply that these predictions are consistent with lack of evidence against the UCIP model of non-word processing. The phonological explanation is also supported by the evidence of a pseudoword superiority effect.

Another assumption of the UCIP model is that the letters are processed in parallel, with a separate detection of each letter. An alternative architecture that does predict capacity coefficient values above one is the coactive architecture. By pooling activation from each of the letters when processing a word, the word is processed much faster than if each letter is processed separately. A coactive architecture in this sense can be thought of as an extreme version of a facilitatory parallel model, in which all activation in each of the letters is shared. Many connectionist models of visual word perception assume a type of coactive architecture. In these models the activation accumulated in favor of a letter is immediately passed on to the word level. In this framework the type of parallel model assumed in the UCIP would not pass on any activation until the letter process is complete. There is some middle ground between these two extremes. One example is that of squelching suggested by Pelli et al. (2003). In this case, the activation from the letter process would only be passed on once it is above a certain threshold. It is important to note that these results are not necessarily inconsistent with serial processing, but for a serial model to predict capacity-values above one it would need to include facilitation and/or be super-capacity.

A coactive architecture could also lead to violations of the assumption of unlimited capacity, so that seemingly more resources are available to each component when more components are present. Capacity values above one imply that the participant had more than four times as many resources dedicated to the word task compared to the letter task, so that none of the individual letter processes has less. In this sense the advantage is similar to chunking; when groups of letters are recognized together, fewer resources are needed for each individual letter. Participants probably do not have truly unlimited resources to dedicate to the task, but having enough to act super-capacity with four letters is not so unreasonable.

There were clearly individual differences present in these data, particularly in word and pseudoword processing capacity. This finding mirrors results reported in accuracy based studies (e.g., Reicher, 1969) and it will be an interesting extension of this work to compare the capacity measure to established measures of individual differences in reading.

Finally, we reiterate the importance of going beyond the simple ANOVA analysis of these data. Merely finding an ordering of the means in the string conditions says nothing about the relative processing efficiencies. For example, faster word processing than non-word processing could be due to the letters in "care" being relatively faster to process than the letters "rlkf". Workload capacity analysis, however, takes the processing of the components into account in estimating efficiency.

Summary

We have demonstrated response time based evidence for visual word perception as a particularly efficient process. This includes evidence that words are more efficiently perceived than predicted by the individual letter reading times, and evidence from comparing word perception efficiency to nonword stimuli. Based on the workload capacity analysis, there is also evidence for a pseudoword superiority effect in the response time domain although not as strong as for word superiority. The evidence we present negates models of word processing that assume parallel, independent processing of letters with separate decision thresholds on each channel. This deeper level of understanding of visual word perception required a shift from statistics based on comparing means toward a more theoretically rich, modeling-based approach.

Acknowledgments

This research was supported by NIH-NIMH MH 057717-07 and AFOSR FA9550-07-1-0078 awarded to JTT.

References

Blaha, L. M., Busey, T., & Townsend, J. T. (2009). An LDA approach to the neural correlates of configural learning. In N. A. Taatgetn & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society.* Austin, TX: Cognitive Science Society.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63–65.

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*, 204–256.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.

McClelland, J. L., & Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception and Psychophysics*, *22*, 249–261.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter percpetion: Part 1. an account of basic findings. *Psychological Review*, *88*, 375–407.

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, *423*, 752–756.

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*, 56–115.

Ramsay, J. O., & Silverman, B. W. (1997). *Functional data analysis*. New York: Springer-Verlag.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, *55A*, 1339–1362.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 274–280.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–360.

Townsend, J. T., & Wenger, M. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1035.