Reading Aloud: Qualitative Differences in the Relation between Stimulus Quality and Word Frequency as a Function of Context

By
Shannon Marie O'Malley

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Psychology

Waterloo, Ontario, Canada, 2008

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Shannon O'Malley

Abstract

Virtually all theories of visual word recognition assume (typically implicitly) that when a pathway is used, processing *within* that pathway always unfolds in the same way. This view is challenged by the observation that simple variations in list composition are associated with qualitative changes in performance. The present experiments demonstrate that when reading aloud, the joint effects of stimulus quality and word frequency on RT are driven by the presence/absence of nonwords in the list. Interacting effects of these factors are seen when only words appear in the experiment whereas additive effects are seen when words and nonwords are randomly intermixed. One way to explain these and other data appeals to the distinction between cascaded processing (or interactive-activation) on the one hand versus a thresholded mode of processing on the other, with contextual factors determining which mode of processing dominates.

Acknowledgements

I would like to thank Derek Besner for his time and assistance in this endeavor. Without his help, thoughtful consideration and encouragement this project would not have been possible.

Dedication

To my family

Table of Contents

List of Tables

List of Figures

Introduction

Word frequency is probably the most well studied psycholinguistic factor over the last forty years or so. All theories of visual word recognition accommodate the effect of this factor in various ways (among others, see Adelman, Gordon, & Quesada, 2006; Balota & Chumbly, 1984; Besner, 1983; Becker, 1976; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; McCann & Besner, 1987; Morton, 1969; Murray & Forster, 2004; Norris, 2006; Perry, Ziegler, & Zorzi, 2007; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989). It is not surprising that there are so many different accounts of how word frequency exerts its effect(s): a main effect does not place strong constraints on theory building. In contrast, the joint effects of multiple factors are much more constraining in this regard. The focus of the current paper is on the joint effects of word frequency and stimulus quality (how easily a word is taken up by the processing system) in the context of reading aloud. The results reported here speak to a number of core issues in visible language processing. In particular, (1) the extent to which processing is dynamic or static, (2) thresholded or cascaded (or engaged in interactive activation), and (3) automatic or context dependent.

Skilled readers are remarkably adept at reading words that have been distorted or rendered difficult to take up in various ways (e.g., by reducing stimulus quality, cAse mIxInG, or masking). Stanners, Jastrembski, and Westbrook (1975) were the first to report that the joint effects of stimulus quality (reduced contrast generated by covering the screen with a sheet of acetate) and word frequency were additive on RT in the context of lexical decision. In the ensuing decades, this observation has been replicated a number of times, using a number of different ways to reduce stimulus quality (Becker & Killion, 1977; Borowsky & Besner, 1993;

Balota & Abrams, 1995; Norris, 1984; O'Malley, Reynolds, & Besner, 2007; Plourde & Besner, 1997; Wilding, 1988; Yap & Balota, 2007).[1] Curiously, *computational* accounts of visual word recognition have to date largely ignored this pattern. The singular exception is Plaut and Booth's (2000; 2006) computational PDP model which purports to simulate this additive pattern. However, Besner, Wartak, and Robidoux (2008) demonstrate that the Plaut and Booth model actually fails in this regard because the joint effects of stimulus quality and word frequency yield a nonmonotonic function (underadditivity, additivity, and overadditivity) depending on the size of the stimulus quality effect whereas skilled readers yield additivity across a wide range of stimulus qualities. Theories of visual word recognition which do not address basic findings like the joint effects noted above (obviously) require development (indeed, such effects ought to be among the "benchmark" phenomena for computational models).

Before considering how such theories might be modified to accommodate these results, several other closely related findings merit consideration. In particular, Yap and Balota (2007) and O'Malley et al. (2007) reported additive effects of stimulus quality and word frequency in lexical decision but an interaction between these factors in reading aloud.[2] Yap and Balota also reported an interaction between stimulus quality and word frequency in the context of a semantic categorization task when the words were not members of the target category. In short, the pattern of joint effects between stimulus quality and word frequency is clearly more complex than generally appreciated to date.

These new findings raise a basic empirical question that should be addressed before re-considering various theoretical accounts. What role is played by the presence/absence of nonwords in the relation between stimulus quality and word frequency, given that task (lexical decision, reading aloud, and semantic categorization) and the presence/absence of nonwords are confounded? That is, when stimulus quality and word frequency interact (reading aloud and semantic categorization) nonwords were not part of the stimulus set, whereas when stimulus quality and word frequency have additive effects (lexical decision when the nonwords are orthographically legal) nonwords are part of the stimulus set.

Lexical decision, by definition, involves discriminating between letter strings that spell a word and letter strings that do not; the presence of nonwords is intrinsic to the task. Of course, it is possible to add nonwords to the semantic categorization task. However, doing so invites the criticism that this changes the nature of the task in ways that are not well understood. Thus, it might be difficult to convince various theorists that such a manipulation is important in the context of this task (but see Forster & Hector, 2002). In contrast, there is a long history of experiments on reading words aloud in which nonwords are sometimes present and sometimes not (e.g., Andrews, 1992; Forster & Chambers, 1973; Fredriksen & Kroll, 1976; McCann & Besner, 1987; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Reynolds & Besner, 2005; 2008). One way to investigate the confounding of task and the presence/absence of nonwords is to have subjects read aloud in an experiment where word frequency and stimulus quality are manipulated and nonwords are randomly intermixed, as compared to when only words appear in the experiment. Three such experiments are reported here.[3]

Before reporting these new experiments, we first provide a brief review of a basic distinction in the way that psycholinguists interested in the processing of visible language think about how the special purpose modules that underlie visual word recognition (e.g., feature level, letter level, word level) communicate with each other. In large part, researchers have typically assumed that how such processing unfolds is fixed—for example, that *how* the processing of high versus low frequency words unfolds over time is not affected by the experimental context (in this case the presence versus absence of nonwords). It is proposed here instead that several different processing modes operate, but *when* and *where* in the processing sequence each does so depends on the context. Following this brief review, we turn to a new contextually based prediction (the lexicalization hypothesis) concerning the effect that the presence of nonwords has on the joint effects of stimulus quality and word frequency when reading aloud.

The thresholded–cascaded/interactive activation distinction

Sternberg (1969) proposed that many mental processes occur in a discrete series, one beginning when another ends. For example, process B starts only after process A finishes. If Factor 1 affects process A but not process B, and Factor 2 affects process B but not process A, then additive effects of these factors on RT should be observed. In this formulation, additive effects of two factors on RT are the signature of distinct processes that occur sequentially.

McClelland (1979; see also Ashby, 1982) proposed a different account in which mental processes are cascaded. In this formulation, processes overlap in time. For example, as soon as process A starts, it sends activation to process B which begins without awaiting the completion of process A. This idea was extended by McClelland and Rumelhart (1981) such that ongoing activation of process B feeds back to process A (interactive activation).

Computational accounts of visible language processing have typically ignored the idea of discrete mental processes. Instead, these models are almost invariably cascaded, and often engaged in interactive activation between various levels. One central question here concerns whether interactive activation can produce systematically additive effects of two factors on RT. To date, we are aware of no existence proof to this effect (see Besner, 2006; Borowsky & Besner, 2006; Besner et al., 2008). In contrast, very simple cascade models are, to a first approximation, able to produce additive effects on *mean* RT under certain conditions (Ashby, 1982; McClelland, 1979; Roberts & Sternberg, 1993).

Despite widespread acceptance of the idea that processing in visual word recognition is cascaded, there are circumstances in which such an account is not easy to reconcile with the data produced by skilled readers. For example, Besner and Roberts (2003) reported that when reading nonwords aloud, RT increased as letter length increased, a reduction in stimulus quality also increased RT, and the joint effect of these two factors was additive. In contrast, simulations with the most successful computational model at that time (Coltheart and colleagues' Dual Route Cascaded model [DRC]) yielded an interaction in which longer letter strings were *less* affected by low stimulus quality than were shorter letter strings. To fix this problem, Besner and Roberts proposed that the DRC model be modified such that the letter level is thresholded rather than allowing it to cascade. This way, the effect of reduced stimulus quality would not affect the model beyond the letter level and, given that the letter length effect arises from subsequent serial left to right assignment of phonemes to letters, the joint effects of stimulus quality and letter length would be additive on RT. Unpublished simulation work in our laboratory confirms that changing the model in this way is successful in that it now produces additive effects of letter length and stimulus quality when reading nonwords aloud.

Another computational account of reading aloud is Perry, Ziegler and Zorzi's (2007) connectionist dual process account (CDP+). The lexical route in this model is taken from the DRC model, whereas the non-lexical route starts with the graphemic buffer, and then uses a two layer assembly network. For present purposes the important characteristic of this model is that the connection from the letter level to the non-lexical route is functionally thresholded which should result in additive effects of stimulus quality and letter length when reading nonwords aloud (see pg. 283 of Perry et al. 2007). Conceptually, this suggests that Sternberg's notion of discrete processing stages that has been ignored by psycholinguists for almost three decades has been rediscovered by some computational modelers.

A second example where cascaded processing appears to be problematic concerns the joint effects of neighborhood density and stimulus quality. Neighborhood density refers to the number of words (N) that can be generated by changing one letter at a time in a letter string (Coltheart, Davelaar, Jonasson, & Besner, 1977; but see Mulatti, Reynolds & Besner, 2006 who report data suggesting that N reflects phonemes rather than letters). As N increases, the time to read both words and nonwords aloud decreases (e.g. Andrews, 1992; McCann & Besner, 1987, among others). Reynolds and Besner (2004) reported that, when reading nonwords aloud, the joint effects of stimulus quality and N were additive on RT whereas simulations with the DRC model produced an interaction in which low stimulus quality slowed low N nonwords more than high N ones. Reynolds and Besner suggested that the same modification to DRC as proposed earlier: threshold the letter level rather than allowing it to cascade.

By way of summary then, thresholding the letter level rather than allowing it to cascade provides a simple way to allow two otherwise very successful computational models of visual word recognition to simulate the performance of skilled readers with respect to the joint effects of stimulus quality and letter length, and stimulus quality and N, both when reading nonwords aloud.

There is, however, a problem with thresholding the letter level in DRC and CDP+. Doing so would produce additive effects of stimulus quality and word frequency when reading aloud, but as we noted earlier, these two factors interact in this context (O'Malley et al., 2007; Yap & Balota, 2007) and Reynolds and Besner (2004) showed that the DRC model also produces an interaction between stimulus quality and word frequency when reading aloud. At first blush then, there is a contradiction across the three findings we have discussed so far. Thresholding the letter level so that the effect of stimulus quality does not affect processing beyond that level is sufficient when considering the joint effects of stimulus quality and letter length, and stimulus quality and N, but fails in the case of the joint effects of stimulus quality and word frequency, all in the context of reading aloud.

However, this contradiction may be more illusory than real. When the joint effects of stimulus quality and a second factor (letter length; N) were additive, participants were reading *nonwords* aloud. When the joint effects of stimulus quality and word frequency interacted, only *words* appeared in the experiment. The proposal advanced here is that when reading nonwords aloud the letter level is thresholded, but when reading only words aloud the letter level cascades through to the word level where performance is affected by word frequency, and hence yields an interaction (as reported both in the data from skilled readers (O'Malley et al, 2007; Yap & Balota, 2007, and in the DRC model as reported by Reynolds & Besner, 2004).

The obvious theoretical question concerns why early processes would be so flexible in terms of their configuration. What benefit does this confer on the process of reading aloud? The hypothesis advanced here is that cascaded processing risks lexical capture when stimulus quality is low; a nonword may activate a word sufficiently strongly that the reader mistakenly reads it as a word instead of the nonword. To reduce this problem participants can threshold the letter level. Rather than attempt to alter the parameter settings on the fly, it is easier to set them (unconsciously) for a block of trials so that processing is either in cascaded mode or in thresholded mode.

At this juncture a reader might reasonably object that the account offered here is consistent with the data, but post hoc. What new predictions, if any, does this account make? If, as suggested above, including nonwords in the experiment serves to modulate the way that processing unfolds, a straightforward prediction is that the joint effects of stimulus quality and word frequency will be additive on RT rather than interact when nonwords are randomly mixed together with the words. This is because the letter level will now be thresholded rather than cascaded so as to avoid the problem of lexical capture in response to nonwords when stimulus quality is low. Two experiments are reported here that test this prediction. A third experiment directly compares the condition in which words and nonwords were mixed together to one in which only words were presented.

Experiment 1

Method

Participants. Thirty-two undergraduate students from the University of Waterloo were each paid $4.00 for their participation. All were native English speakers and reported normal or corrected-to-normal vision.

Stimuli. The stimulus set consisted of two hundred words and two hundred nonwords. The one hundred high frequency words (mean count per million = 411.6) and one hundred low frequency words (mean count per million = 17.3) were taken from Yap and Balota (2007). The mean number of letters in the words was 4.8 (range 3–7). The mean orthographic neighborhood size (N values: see Coltheart, Davelaar, Jonasson, & Besner, 1977) for the high frequency words was 4.8, and the mean summed bigram frequency was 6370. For low frequency words, the mean orthographic neighborhood size was also 4.8, and the mean summed bigram frequency was 6149. The nonwords, taken from O'Malley et al. (2007), were matched to the words for length (mean = 4.8, range 3–7), and the mean orthographic neighborhood size was 8.9. The nonwords were split into 4 lists and rotated through conditions using a partial Latin square such that each nonword list was presented with each word list equally often across participants, resulting in 8 lists.

The stimuli were rotated through stimulus quality conditions across participants, who were assigned to a counterbalancing condition based on order of arrival in the laboratory, with words and nonwords randomly intermixed. The letter strings were displayed in 16 point Times New Roman font on a black background (RGB 0, 0, 0). In the bright condition, the letter strings appeared in RGB (120,120,120); in the dim condition, they appeared in RGB (36, 36, 36). The lighting in the room was dim (a measure of luminance at the level of the screen would have been preferable, but this laboratory lacks this expensive piece of equipment).

Apparatus. The data were collected on a Pentium 4 computer running E-Prime 1.1 (Schneider, Eschman, & Zuccolotto, 2001). Stimuli were displayed on two 17” monitors: One monitor presented stimuli to the participants. The other monitor allowed the experimenter to observe what letter string was presented without disturbing the participant. Vocal responses were collected using a Plantronics LS1 microphone headset and a voice key assembly.

Procedure. Participants were tested individually and were seated approximately 50 cm from the screen. At this distance, 3-letter words subtended approximately 1.2° of visual angle and 7-letter words subtended approximately 3.1° of visual angle. Participants were instructed that when a letter string appeared on the screen their task was to pronounce it as quickly and as accurately as possible. Responses were coded as correct, incorrect, or mistrial (e.g., voice key error) by the experimenter. Each trial consisted of a fixation symbol (+) at the center of the screen for 250 ms followed by a blank screen for 56 ms after which the word was presented at fixation until a vocal response was detected. A set of 20 practice trials (10 words and 10 nonwords) served to familiarize the participant with the task and allowed the experimenter to adjust the microphone sensitivity to minimize spoiled trials (i.e., trials in which either the microphone failed to respond or it responded prematurely).

Results

RTs and errors were analyzed across participants and items, with both stimulus quality and word frequency as within-subject factors in the subject analysis. In the item analysis, stimulus quality was a within-item factor and word frequency was a between-item factor. To remove individual subject variance, the item data were z-scored prior to the analysis (e.g., see Reynolds & Besner, 2004). The subject data can be seen in Table 1. 95% Confidence Intervals (CI) for the difference scores were calculated with the Masson and Loftus (2003) within-subjects procedure. The variance for stimulus quality was greater than for word frequency in all experiments; the confidence intervals therefore were calculated using the mean standard errors for the interaction.

Analysis of only the mean RTs is potentially misleading. For example, Yap, Balota, Tse and Besner (2008) using a lexical decision task, found opposing interacting effects in a distributional analysis, leading to additivity of two factors in the means. Here, if early processing is thresholded, we would also expect that the joint effects of word frequency and stimulus quality would be additive through much of the distribution. Specifically, the size of the word frequency effect should be the same for bright and dim words across the distribution. We therefore report vincentile plots for the joint effects of stimulus quality and word frequency.

RTs. Trials on which there was a voice key error (1.4%) or an incorrect response (2.9%) were removed prior to RT data analysis. The remaining RTs were submitted to a recursive data trimming procedure in which the criterion for outlier removal was established based on the sample size in that cell (Van Selst & Jolicœur, 1994), resulting in an additional 1.9% of the data being removed. Mean RTs and Errors can be seen in Table 1, item means can be seen in Appendix A. Words presented brightly were read aloud faster than dim ones, $F_1(1, 31) = 66.5$, $MSE = 2384.0$, $p < .001$, $F_2(1, 198) = 834.2$, $MSE = .032$, $p < .001$. High frequency words were read aloud faster than low frequency words, $F_1(1, 31) = 48.9$, $MSE = 405.4$, $p < .001$, $F_2(1, 198) = 36.6$, $MSE = .126$, $p < .001$. Critically, there was no interaction between the effects of stimulus quality and word frequency (*F*s < 1).

Errors. There was a main effect of stimulus quality, $F_1(1, 31) = 7.1$, $MSE = 3.5$, $p < .05$, $F_2(1, 198) = 7.8$, $MSE = 56.3$, $p < .01$. More errors were made to low frequency words than to high frequency words, $F_1(1, 31) = 15.2$, $MSE = 3.9$, $p < .001$, $F_2(1, 198) = 53.1$, $MSE = 15.6$, *p*

< .001. There was a 1.5% interaction between stimulus quality and word frequency in which low stimulus quality affected low frequency words more than high frequency ones, $F_1(1, 31) = 6.2$, $MSE = 2.9$, $p < .05$, $F_2(1, 198) = 5.7$, $MSE = 9.8$, $p <$ 05.

Table 1

Mean Response Times (RTs in ms), 95% Confidence Intervals (CI), and Mean Percentage Errors (%E) in Reading Aloud as a Function of Word Frequency and Stimulus Quality in Experiment 1.

	Clear			Degraded		
	RT	CI	%E	RT	CI	%E
Low Frequency	543		1.3	615		2.9
High Frequency	520		0.7	589		0.8
Difference	**23**	**± 4**	**0.6**	**26**	**± 4**	**2.1**
Nonwords	575		5.3	648		6.4

Vincentile analysis. A vincentizing procedure was used in which the response time distributions for individual participants are averaged across participants to produce the response time distribution (Vincent, 1912). Ten vincentiles (the mean of observations within a given percentile range) were first computed for each participant. The individual vincentiles were then averaged across participants and the mean vincentiles plotted. The vincentile plots reported here were computed in R (R Development Core Team, 2004).

The mean vincentiles are plotted as a function of word frequency and stimulus quality in Figure 2. The difference scores (high frequency – low frequency) for clear and degraded items are plotted in Figure 3. The frequency effect increased across vincentiles for both clear and degraded items. For present purposes, the key result is that the overlap in the size of the frequency effects for clear and degraded items is consistent with additivity between stimulus quality and word frequency throughout the distribution.

Figure 1

Experiment 1: Vincentile means for participants' reading aloud times as a function of word frequency and stimulus quality.

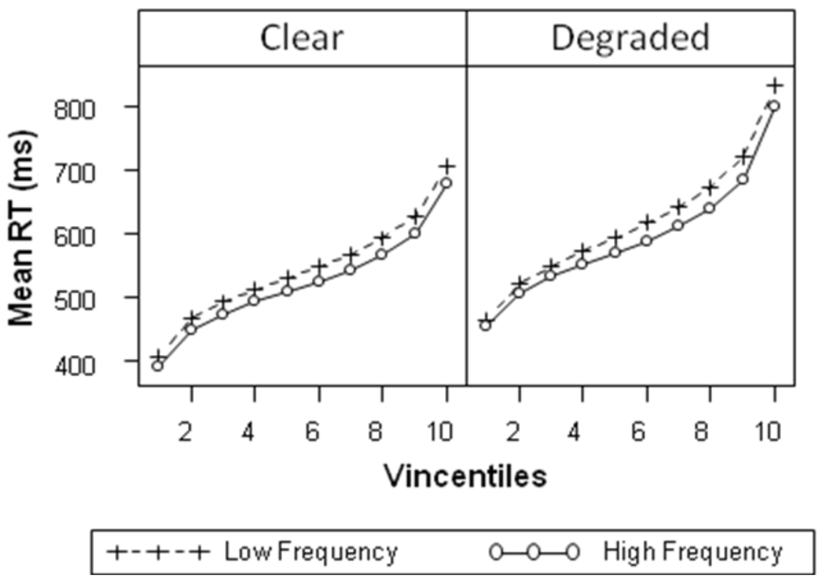
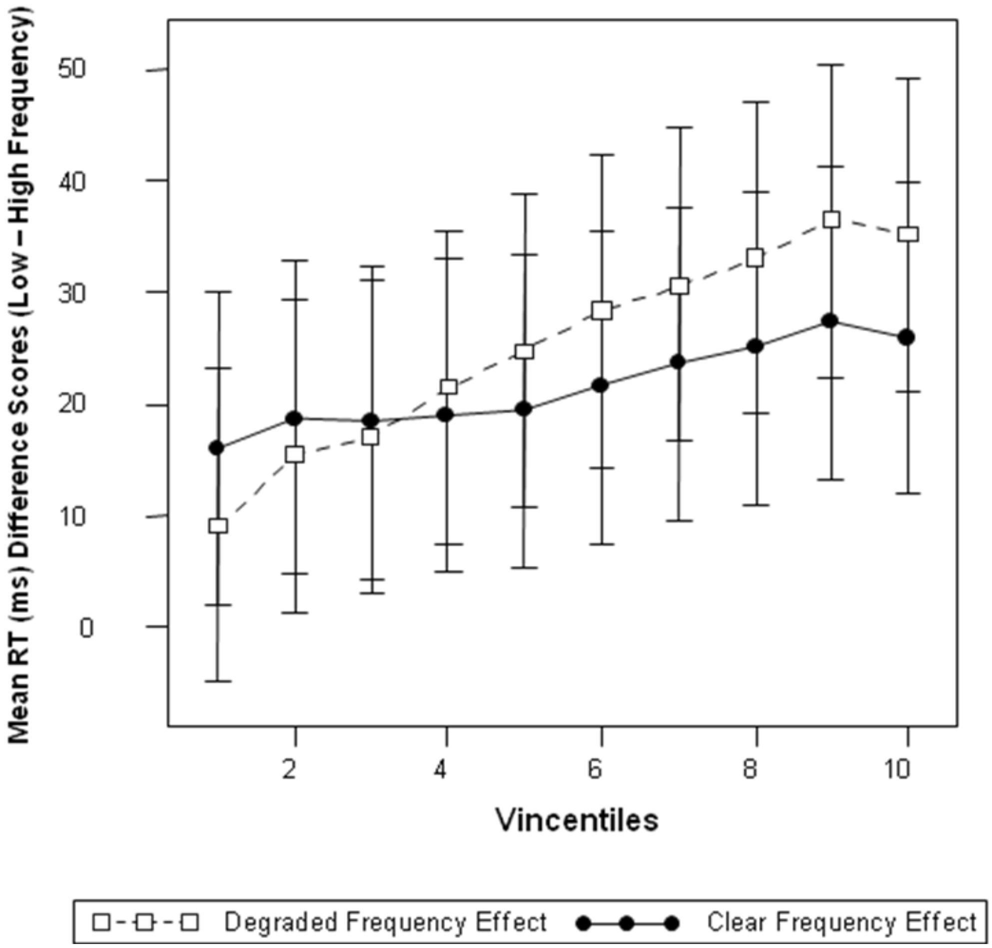


Figure 2

Experiment 1: The difference in the vincentile means for low versus high frequency items for participants' reading aloud times and 95% confidence intervals as a function of stimulus quality.



Discussion

Experiment 1 yielded clear additivity in the RT data and in the vincentiles (but there was a small (1.5%) interaction in the error data). This result in the RTs contrasts with the results reported by both Yap and Balota (2007) and O'Malley et al (2007) that used the same word set. Both sets of investigators found that stimulus quality and word frequency interacted on RT in the context of reading aloud. The primary difference between those experiments and the present one is the absence of nonwords in the prior experiments, and their presence here. We discuss these results further after reporting Experiments 2 and 3.

Experiment 2

Given that the results of Experiment 1 are novel and surprising to many colleagues, we report a replication. In Experiment 2 we used the DMDX software (Forster & Forster, 2003) which has the advantage of recording the vocal responses. Using this software in conjunction with CheckVocal (Protopapas, 2007), allows one to determine RTs using the waveform and hence serves to reduce measurement error associated with voice key timing (Rastle & Davis, 2002) as well as possible experimenter bias associated with determining errors online. In using the DMDX software other small adjustments were necessary in the method and procedure to accommodate the new software, specifically (1) the brightness of the stimuli, and (2) the timing of presentation of stimuli.

Method

Participants. A new set of thirty-two undergraduate students from the University of Waterloo were each paid $4.00 for their participation. All were native English speakers and reported normal or corrected-to-normal vision.

Stimuli. Experiment 2 used the same items as in Experiment 1. The stimuli were rotated through stimulus quality conditions across participants, who were assigned to a counterbalancing condition based on order of arrival in the laboratory. Words were again displayed in 16 point Times New Roman font on a black background (writing color 000, 000, 000). In the bright condition, the letter strings appeared in writing color (255, 255, 255); in the dim condition, they appeared in writing color (075, 075, 075). These values differ from Experiment 1 because e-prime and DMDX software have different parameters for RGB settings, (the color white is set as 120, 120, 120 in e-prime and 255, 255, 255 in DMDX) making it difficult to set the exact same brightness across experiments. However, this difference adds to the strength of a replication of the findings in the sense of generalizing across more than one brightness level.

Apparatus. The data were collected using the DMDX software and RTs and errors were determined using CheckVocal software.

Procedure. The procedure was identical to that of Experiment 1 except for two small changes: (1) responses were coded offline as correct, incorrect, or mistrial (e.g., voice key error) by the experimenter using the CheckVocal software, (2) the fixation symbol (+) appeared for 56 ms, followed by a blank screen for 150 ms, after which the stimulus was presented at fixation until a response was detected.

Results

RTs. Trials on which there was a mistrial (1.1%) or an incorrect response (4.5%) were removed prior to RT analysis. The remaining RTs were submitted to the same recursive data trimming procedure as in Experiment 1, resulting in an additional 1.6% of the data being removed. These data can be seen in the middle part of Table 2. Words presented brightly were read aloud faster than those in the dim condition, $F_1(1, 31) = 275.3$, $MSE = 402.7$, $p < .001$, $F_2(1, 198) = 1030$, $MSE = .013$, $p < .001$. High frequency words were read aloud faster than low frequency words, $F_1(1, 31) = 78.1$, $MSE = 137.2$, $p < .001$, $F_2(1, 198) = 17.3$, $MSE = .226$, $p < .001$. There was no interaction between the effects of stimulus quality and word frequency (*F*s < 1).

Errors. There was a main effect of stimulus quality, $F_1(1, 31) = 14.4$, $MSE = 4.9$, $p < .01$; $F_2(1, 198) = 18.3$, $MSE = 12.5$ $p < .001$. More errors were made to low frequency words than to high frequency words, $F_1(1, 31) = 22.3$, $MSE = 4.4$, $p < .001$, $F_2(1, 198) = 15.1$, $MSE =$

18.9, $p < .001$. There was no interaction between stimulus quality and word frequency, $F_1(1, 31) = 1.58$, $MSE = 3.9$, $p = .22$, $F_2(1, 198) = 1.4$, $MSE = 12.5$, $p = .22$.

Table 2

Mean Response Times (RTs in ms), 95% Confidence Intervals (CI), and Mean Percentage Errors (%E) in Reading Aloud as a Function of Word Frequency and Stimulus Quality in Experiment 2.

	Clear			Degraded		
	RT	CI	%E	RT	CI	%E
Low Frequency	509		2.1	568		4.0
High Frequency	491		0.8	550		1.9
Difference	**18**	**± 4**	**1.3**	**18**	**± 4**	**2.1**
Nonwords	551		7.3	612		10.9

Vincentile analysis. The mean vincentiles were again plotted as a function of word frequency and stimulus quality and appear in Figure 4. The difference scores (low frequency – high frequency) for clear and degraded items are plotted in Figure 5. It is clear that the frequency effect increases across vincentiles for both clear and degraded items. The overlap in the size of the frequency effects for clear and degraded items is consistent with additivity between stimulus quality and word frequency throughout the distribution.

Discussion

Experiment 2 replicated the RT results observed in Experiment 1. When nonwords are randomly intermixed with words, and the task is to read items aloud, stimulus quality and word frequency have additive effects on both mean RT and throughout the distribution. Experiment 2 also produced additive effects of these factors on the errors (thus failing to replicate the small interaction in the error data observed in Experiment 1).

Figure 3

Experiment 2: Vincentile means for participants' reading aloud times as a function of word frequency and stimulus quality.

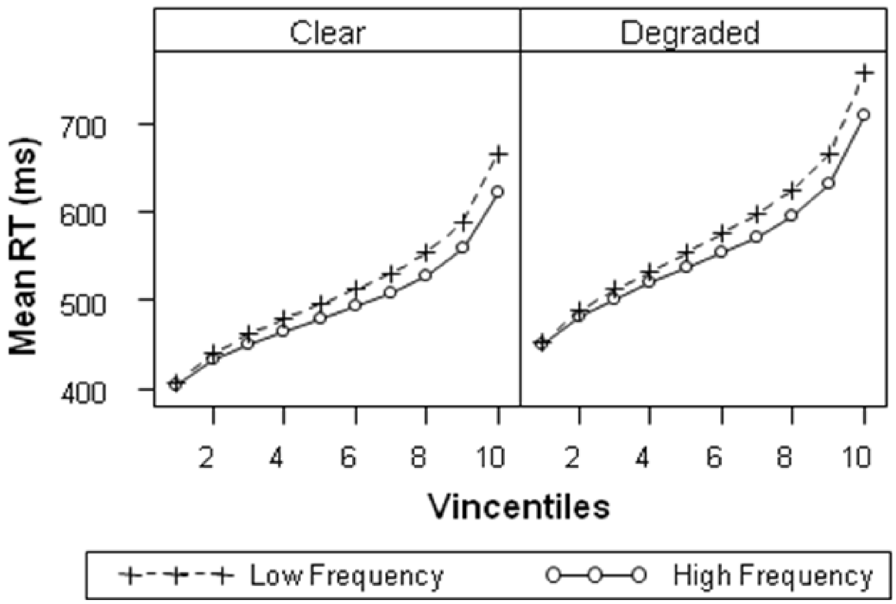
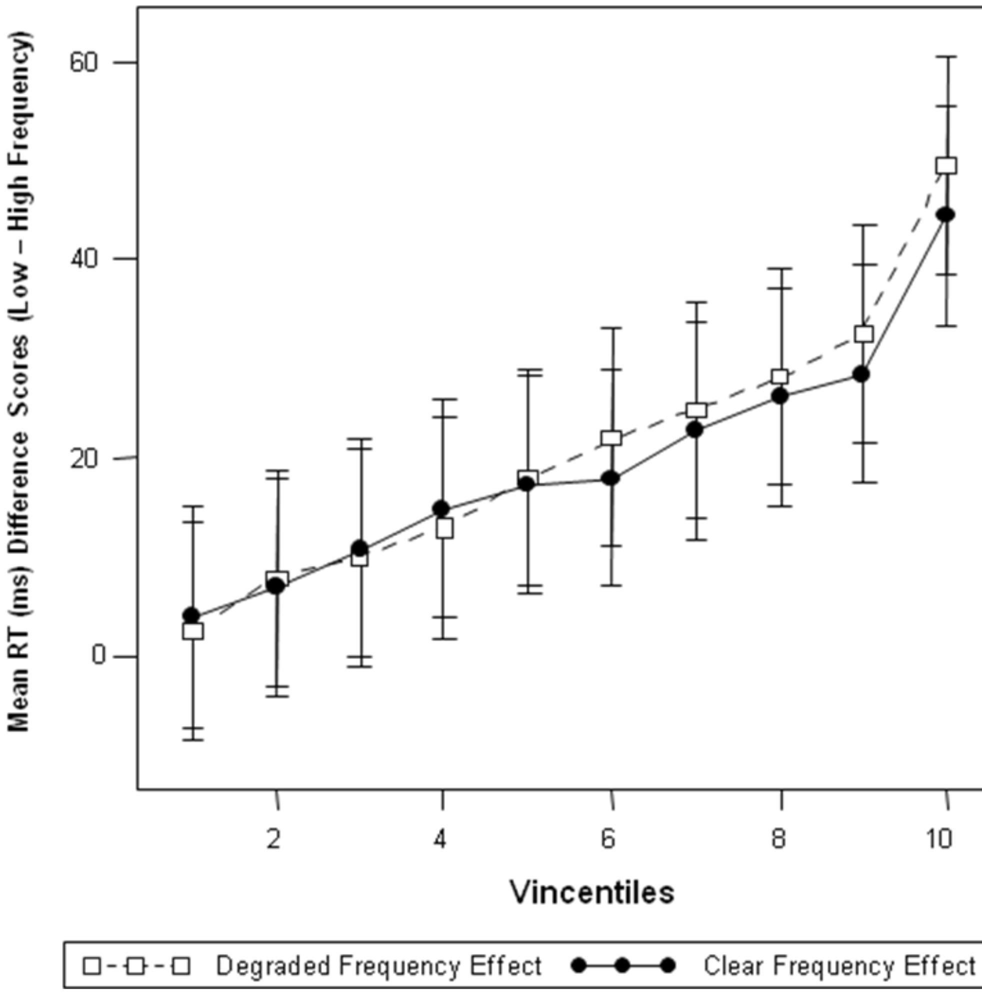


Figure 4

Experiment 2: The difference in the vincentile means for low versus high frequency items for participants' reading aloud times and 95% confidence intervals as a function of stimulus quality.



Experiment 3

The results of Experiment 1 and 2 differ from the results reported by O'Malley et al. (2007) where the task was to read *words* aloud, using the same word set and the same kind of stimulus quality manipulation as used here. Nonwords did *not* appear in the O'Malley et al (2007) experiment, and an interaction between stimulus quality and word frequency was observed (see also Yap & Balota, 2007). Given the importance of this change in the relation between stimulus quality and word frequency as a function of the presence/absence of nonwords, we sought to further strengthen the case by replicating the previous experiments using a new word set (at the request of the editor), and by including a condition in which only words appear. We expected this experiment to produce a three way interaction in which an interaction between stimulus quality and word frequency is observed for subjects who are only presented with words, whereas subjects who are presented with words and nonwords mixed together yield additive effects of stimulus quality and word frequency.

Method

Participants. Fifty-six undergraduate students from the University of Waterloo were each paid $4.00 for their participation. Thirty-two of them read words and nonwords aloud, the other twenty-four students read only words aloud. All were native English speakers and reported normal or corrected-to-normal vision.

Stimuli. Experiment 3 used a new set of two hundred words and two hundred nonwords. The one hundred high frequency words (mean count per million = 666.2) and one hundred low frequency words (mean count per million = 16.1) both had a mean of 4.8 letters (range 3–7). The mean orthographic neighborhood size for the high frequency words was 4.9, and the mean summed bigram frequency was 7093. The mean orthographic neighborhood size was also 4.9 for the low frequency words, and the mean summed bigram frequency was 5954. The nonwords matched the words in length (mean = 4.8, range 3–7), and the mean orthographic neighborhood size was 5.9. The items were rotated through conditions using a partial Latin square such that each nonword list was presented with each word list equally often across participants in the words and nonwords condition, resulting in 8 lists. The stimuli were rotated through stimulus quality conditions across participants in both the words only and the words and nonwords conditions. These were assigned to a counterbalancing condition based on order of arrival in the laboratory, with the nonword condition alternating between participants.

Procedure. The procedure was identical to that of Experiment 2.

Results

As in the previous analyses, RTs and errors were analyzed across participants and items, with both stimulus quality and word frequency as within-subject factors in the subject analysis. In the item analysis, stimulus quality was a within-item factor and word frequency was a between-item factor. The item data were z-scored prior to the analysis to reduce the impact of individual subject variance; the z-scores were calculated by collapsing across all conditions. The subject data can be seen in Table 3. Trials on which there was a voice key error (0.9%) or an incorrect response (4.5%) were removed prior to RT data analysis. The remaining RTs were submitted to the same recursive data trimming procedure used in Experiments 1 and 2, resulting in an additional 1.8% of the data being removed.

Three-way interaction. The critical three way interaction between stimulus quality, word frequency and experimental condition (presence/absence of nonwords) was significant in the RT analysis, $F_1(1, 54) = 4.5$, $MSE = 125$, $p = .04$, $F_2(1, 198) = 5.9$, $MSE = .038$, $p = .02$, but not in the error analysis, $F_1(1, 54) = .9$, $MSE = 6.5$, $p = .34$, $F_2(1, 198) = 1.1$ $MSE = 20.3$, $p = .30$.

Further analysis treated the data for the list conditions (presence/absence of nonwords) separately.

Words only. Brightly presented words were read aloud faster than dimly presented ones, $F_1(1, 23) = 138.2$, $MSE = 2222$, $p < .001$, $F_2(1, 198) = 2850$, $MSE = .054$, $p < .001$. High frequency words were read aloud faster than low frequency words, $F_1(1, 23) = 22.7$, $MSE = 365$, $p < .001$, $F_2(1, 198) = 16.2$, $MSE = .213$, $p < .001$. The interaction between the effects of stimulus quality and word frequency was significant, $F_1(1, 23) = 11.3$, $MSE = 123$, $p < .01$, $F_2(1, 198) = 10.1$, $MSE = .024$, $p < .01$.

Errors. There was a main effect of stimulus quality, $F_1(1, 23) = 16.3$, $MSE = 15.9$, $p < .01$, $F_2(1, 198) = 424.4$, $MSE = 25.6$, $p < .001$. There was no main effect of frequency, $F_1(1, 23) = 2.9$, $MSE = 8.9$, $p = .10$, $F_2(1, 198) = 3.4$, $MSE = 32.3$, $p = .07$, and there was no interaction between stimulus quality and word frequency ($Fs < 1$).

Words when mixed with nonwords. Brightly presented words were read aloud faster than dimly presented ones, $F_1(1, 31) = 327$, $MSE = 1178$, $p < .001$, $F_2(1, 198) = 2698$, $MSE = .039$, $p < .001$. High frequency words were read aloud faster than low frequency words, $F_1(1, 31) = 14.1$, $MSE = 168$, $p = .001$, $F_2(1, 198) = 3.2$, $MSE = .158$, $p = .07$. There was no interaction between the effects of stimulus quality and word frequency, ($Fs < 1$).[4]

Errors. There was a main effect of stimulus quality, $F_1(1, 31) = 30.2$, $MSE = 7.8$, $p < .001$, $F_2(1, 198) = 44.1$, $MSE = 16.8$, $p < .001$. More errors were made to low frequency words than to high frequency words, $F_1(1, 31) = 7.1$, $MSE = 6.7$, $p < .05$, $F_2(1, 198) = 5.8$, $MSE = 25.7$, $p < .05$. The interaction between stimulus quality and word frequency was marginal, $F_1(1, 31) = 2.9$, $MSE = 6.7$, $p = .09$, $F_2(1, 198) = 3.6$, $MSE = 16.8$, $p = .06$.

Vincentile analysis. The mean vincentiles are plotted as a function of word frequency and stimulus quality, for the condition in which only words appeared, (see Figure 6), and for the condition in which both words and nonwords where mixed together (see Figure 8). The difference scores (low frequency – high frequency) for clear and degraded items are plotted in Figures 7 and 9 for the condition in which only words were presented, and when words mixed with nonwords respectively. It is clear that the frequency effect increases across vincentiles for both clear and degraded items. In Figure 7 there is a clear divergence in the size of the frequency effect such that as reaction times increased, the size of the frequency effect for degraded items increased more than for the clear items. In contrast, inspection of Figure 9 reveals that when words and nonwords were mixed together the size of the frequency effect was approximately the same throughout the distribution for bright and dim words.

Table 3

Mean Response Times (RTs in ms) and Mean Percentage Errors (%E) when Reading Words Aloud for the Combined Analysis of Experiments 1 – 3 along with a median split based on the average errors.

	Clear			Degraded		
	RT	CI	%E	RT	CI	%E
Words Only						
Low Frequency	481		1.7	602		5.1
High Frequency	470		0.8	576		3.9
Difference	**11**	**± 5**	**0.9**	**26**	**± 5**	**1.2**
With Nonwords						
Low Frequency	513		1.8	624		5.3
High Frequency	506		1.4	614		3.3
Difference	**7**	**± 4**	**0.4**	**10**	**± 4**	**2.0**
Nonwords	562		7.7	677		12.9

Discussion

Experiment 3, with a new stimulus set, provides a replication of the interaction between word frequency and stimulus quality when reading only words aloud, as in Yap and Balota (2007) and in O'Malley et al. (2007), totalling 3 such demonstrations. It also provides a second replication of the null interaction between these two factors on RT when nonwords are present, totalling 3 demonstrations.

Figure 5

Experiment 3: Vincentile means for participants' reading aloud times as a function of word frequency and stimulus quality when only words are read aloud.

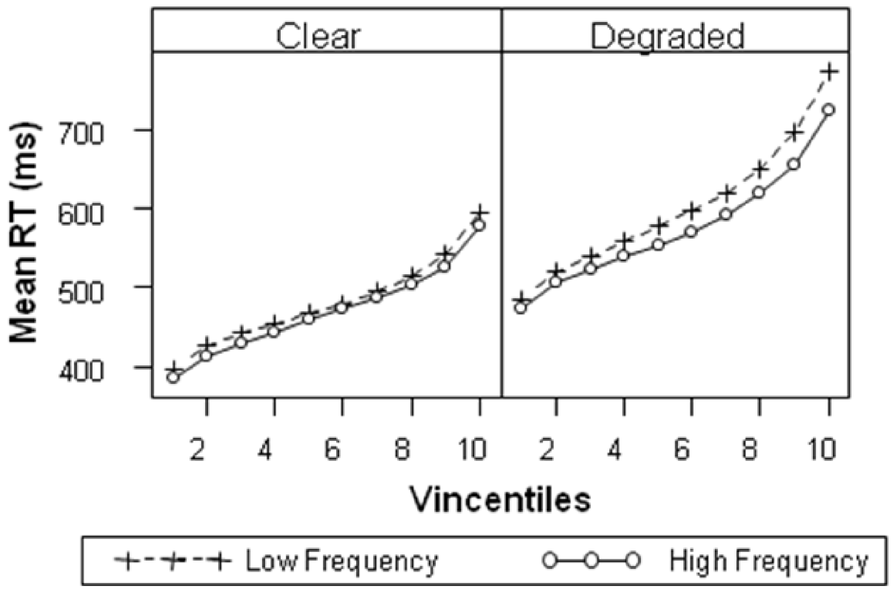


Figure 6

Experiment 3: The difference in the vincentile means for low versus high frequency items for participants' reading aloud times and 95% confidence intervals as a function of stimulus quality when only words are read aloud.

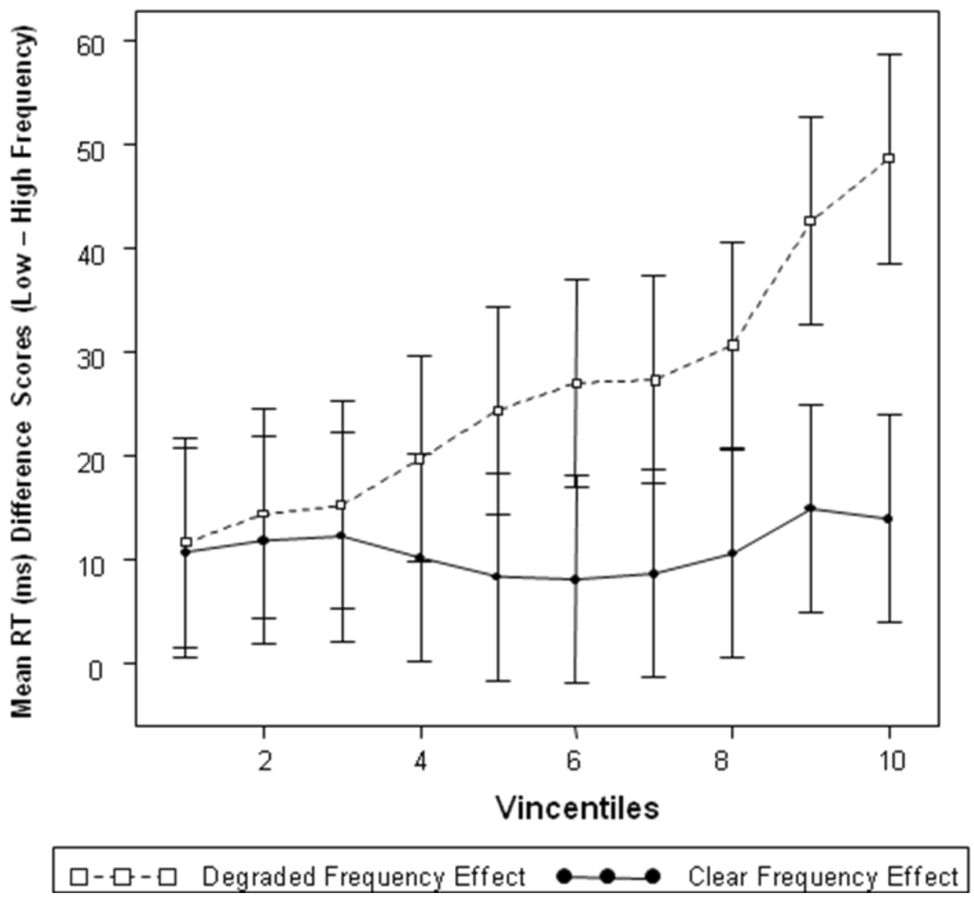


Figure 7

Experiment 3: Vincentile means for participants' reading aloud times as a function of word frequency and stimulus quality when words and nonwords are read aloud.

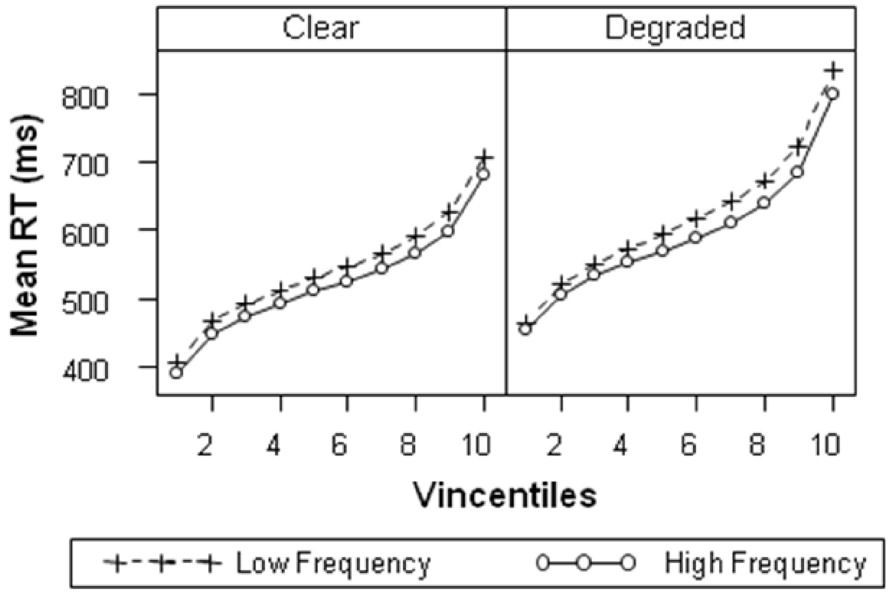
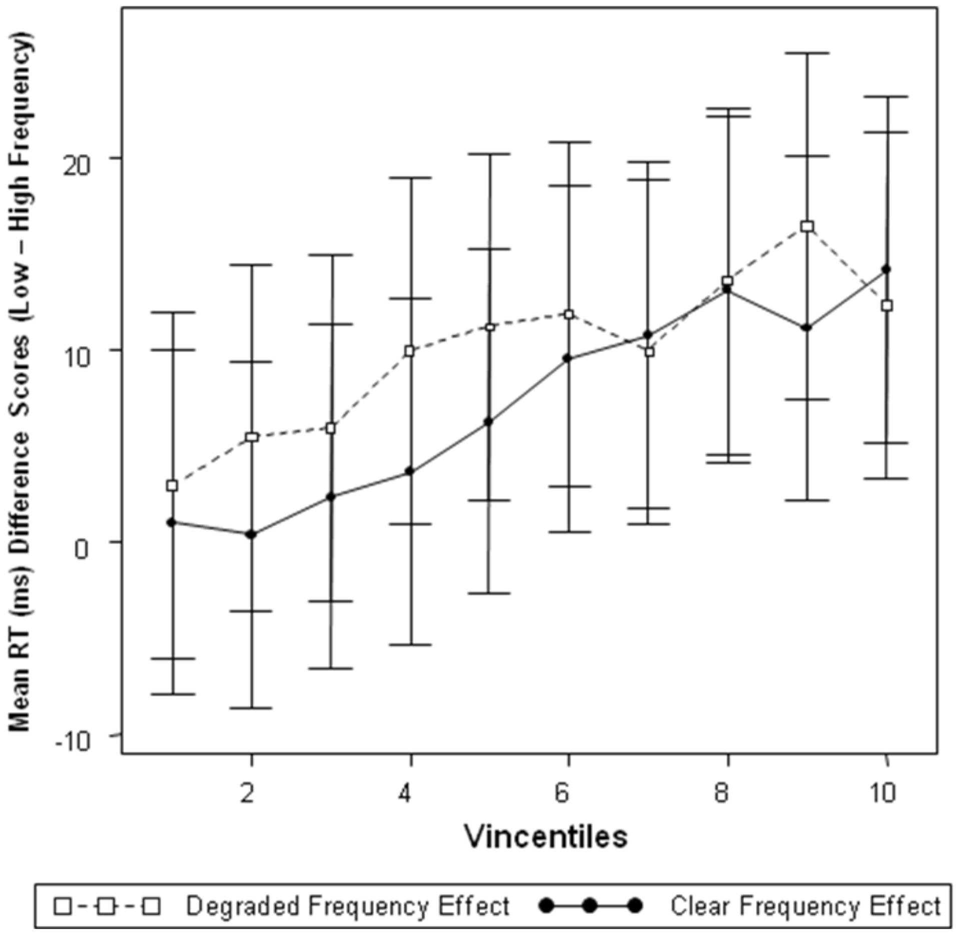


Figure 8

Experiment 3: The difference in the vincentile means for low versus high frequency items for participants' reading aloud times and 95% confidence intervals as a function of stimulus quality when words and nonwords are read aloud.



Combined Analysis of Participants across Experiments

The three experiments reported here yielded no significant interaction between word frequency and stimulus quality in either the mean RTs or the vincentiles when both words and nonwords were mixed in the same block and read aloud. In the first experiment a small interaction was observed in the error data but not in the following two experiments, although there was a trend towards an interaction in the third experiment. In order to further explore these findings and increase power, we combined the three experiments in one analysis. These data can be seen in Table 2.

For RTs, there was a main effect of stimulus quality, $F_1(1,95) = 344$, $MSE = 131.3$, $p < .001$, $F_2(1,398) = 2817$, $MSE = .05$, $p < .001$ a main effect of word frequency, $F_1(1,95) = 105$, $MSE = 280$, $p < .001$, $F_2(1,398) = 26$, $MSE = .156$, $p < .001$, but no interaction ($Fs < 1$).

For errors, there were main effects of both stimulus quality, $F_1(1,95) = 48.7$, $MSE = 5.8$, $p < .001$, $F_2(1,398) = 65.2$, $MSE = 11.7$, $p < .001$ and word frequency, $F_1(1,95) = 37.9$, $MSE = 5.0$, $p < .001$, $F_2(1,398) = 18.9$, $MSE = 20.75$, $p < .001$. There was a significant interaction in the errors, $F_1(1,95) = 8.7$, $MSE = 4.4$, $p < .01$, $F_2(1,398) = 7.9$, $MSE = 11.7$, $p < .01$.

The fact that there is a reliable interaction in the errors might be seen as undermining the additivity observed in the RT data, but this is not the only interpretation. Plourde and Besner (1997) suggested that when early processing is thresholded, some participants may not always have completely finished the cleanup operation (i.e., activation is passed on before the effect of degradation is fully resolved) resulting in a tendency to produce an interaction in the error data. Participants who make many errors are arguably those who unduly emphasize speed and are thus likely to terminate the cleanup process early on some proportion of the trials. To investigate this issue here we did a median split on the participant data based on the average percentage errors across conditions.

For participants with few errors (average = .87%) there was a main effect of stimulus quality on RTs $F_1(1,47) = 138$, $MSE = 1894$, $p < .001$, and of word frequency, $F_1(1,47) = 49.8$, $MSE = 253$, $p < .001$, but no interaction, $F_1(1,47) = 1.6$, $MSE = 122.9$, $p = .21$. In the error analysis there were main effects of stimulus quality, $F_1(1,47) = 12.6$, $MSE = 1.4$, $p < .01$, and of word frequency, $F_1(1,47) = 10.8$, $MSE = 1.0$, $p < .01$, but no interaction, $F_1(1,47) = .26$, $MSE = 2.0$, $p = .61$.

For participants with more errors (average = 3.5%) there was a main effect of stimulus quality on RTs, $F_1(1,47) = 218.4$, $MSE = 1662$, $p < .001$, and of word frequency, $F_1(1,47) = 55.1$, $MSE = 310$, $p < .001$, but no interaction $F_1(1,47) = .02$, $MSE = 140.8$, $p = .89$. However, there was an interaction in the errors, $F_1(1,47) = 10.2$, $MSE = 6.3$, $p = .002$. The results of this analysis are therefore consistent with the suggestion that the interaction in the error data is driven by those participants who generate more errors and are likely to have prematurely terminated the cleanup process on some proportion of trials.

Table 4

Mean Response Times (RTs in ms) and Mean Percentage Errors (%E) when Reading Words Aloud for the Combined Analysis of Experiments 1 – 3 along with a median split based on the average errors.

	Clear		Degraded	
	RT	%E	RT	%E
Combined Analysis (N = 96)				
Low Frequency	522	1.7	604	4.1
High Frequency	506	0.9	585	2.0
Difference	**16**	**0.8**	**19**	**2.1**
Nonwords	562	6.8	646	10.1
Few Errors Group (N = 48)				
Low Frequency	523	0.8	599	1.5
High Frequency	509	0.4	580	0.9
Difference	**14**	**0.4**	**19**	**0.6**
High Errors Group (N = 48)				
Low Frequency	522	2.7	609	6.7
High Frequency	503	1.5	590	3.2
Difference	**19**	**1.2**	**19**	**3.5**

General Discussion

The results of the present three experiments can be summarized as follows. Stimulus quality and word frequency have additive effects on reading aloud RTs when words and nonwords are randomly intermixed (Experiments 1, 2 and 3). However, when only words appear in the experiment, stimulus quality and word frequency interact such that high frequency words are less affected by low stimulus quality than are low frequency words (Experiment 3; see also O'Malley et al., 2007; Yap & Balota, 2007).

Three issues merit discussion here. First, *how* do the same factors (stimulus quality and word frequency) produce both additive and interacting effects on RT as a function of the presence/absence of nonwords in the list? Second, *why* does this happen? And third, *what* general implications, if any, do these results have for our understanding of visual word recognition processes?

*The **how** of **additive** effects of stimulus quality and word frequency*

One way to understand how additive effects of stimulus quality and word frequency arise is in terms of Sternberg's (1969) proposal that additive effects of two factors on RT reflect serially organized processes in which some process only starts after the prior process has finished, and each of two manipulated factors affects a separate process. There is both broad and deep support for this seemingly implausible proposal (see Sternberg, 1998; Roberts & Sternberg, 1993). This account has been suggested before in the context of exactly these factors (albeit in the context of lexical decision). Stimulus quality affects feature and letter processing but not subsequent orthographic lexical activation, whereas word frequency affects lexical activation but not feature and letter level processing. (e.g., Borowsky & Besner, 1993; Plourde & Besner, 1997; O'Malley et al, 2007; Yap & Balota, 2007; Yap et al., 2008).

It is also known that cascaded processing, provided certain constraints are satisfied, can produce additive effects of two factors on mean RT (Ashby, 1982; McClelland, 1979; Roberts & Sternberg, 1993). For example in a localist model, if stimulus quality affects the feature level and not the letter level, and word frequency affects the lexical level, additivity could be observed in a cascaded model provided the feature and lexical levels are relatively fast, while the letter level is relatively slow (see Roberts & Sternberg, 1993). Additivity could presumably also be generated by a PDP model if there is a relatively slow hidden unit level between relatively fast feature and letter levels with the feature level being affected by stimulus quality and the letter level by word frequency.

That said, we are aware of no implemented model of visual word recognition in its current form that produces additivity of stimulus quality and word frequency in reading aloud. Indeed, exploration of this issue by Reynolds and Besner (2004) failed to yielded additivity of these factors in the context of the DRC model (even when feedback was eliminated by zeroing out the connections between levels). Simulating additive effects of these factors (and others) in the context of such models may be less easy to accomplish than implied by prior work described by McClelland, and by Roberts and Sternberg, given that the parameter constraints that need to be satisfied may not be so easily reconciled with the architecture and processing dynamics currently implemented that play a critical role in simulating other phenomena. In particular, and as noted earlier, cascaded processing, at least in the context of the DRC model where it feeds a serial process in the nonlexical route, leads to an unusual outcome in which a factor that slows processing (letter length) when combined with another factor that also slows processing (stimulus quality) yields an interaction in which the effect of stimulus quality *decreases* as letter length *increases* (Besner & Roberts, 2003). The human data do not yield this pattern.

Interactive-activation between various levels is also a central assumption in many computational accounts of visual word recognition, reading aloud and perceptual identification (e.g., McClelland & Rumelhart, 1981; McClelland, 1987; Coltheart et al 2001; Perry et al, 2007). We are aware of no demonstration to date that any IA model can produce all the effects currently considered benchmarks, and also produce systematically additive effects of word frequency and stimulus quality**.** Proponents of such models might therefore take the additivity of stimulus quality and word frequency (along with a number of other examples; see Besner, 2006) as an issue that merits attention. To be absolutely clear, we are not claiming that these models are incapable of producing additivity of factor effects, rather that they do not do so in their current form.[5]

*The **how** of an **interaction** between stimulus quality and word frequency*

The how of the interaction between stimulus quality and word frequency is likely to be uncontroversial given that interactive activation (as in the DRC model) produces an interaction between stimulus quality and word frequency when reading aloud (Reynolds & Besner, 2004). As well, cascaded processing (as in the DRC model when IA is prevented by lesioning feedback) also produces an interaction between stimulus quality and reading aloud (Reynolds & Besner, 2004). We do not expect that producing such an interaction in the context of a PDP model would be difficult either, but that of course remains to be demonstrated.

It is not immediately obvious to us how serially organized processes as in Sternberg's proposal can produce an interaction between stimulus quality and word frequency given that these same factors are additive when nonwords are intermixed with the words. One might suppose that feature and letter processing fail to completely clean up the internal representation of the stimulus before passing it for lexical processing when only words appear in the list (relatedly, see Sternberg, 1967, session 1 vs. session 2). Or, there might be reasons (unidentified to date) why stimulus quality affects both feature/letter processing *and* lexical processing under these conditions. Until there is some plausible proposal as to why this might be the case we are inclined to the view that discrete processes are problematic when only words appear in the list).

In summary, it is easy to produce *additive* effects of stimulus quality and word frequency when the processes affected by these factors are serially arranged, discrete and doubly dissociated in the sense that factor A affects the first process and not the second, and factor B affects the second process, but not the first. Cascading processes (feed-forward) and cascading processes combined with feed-back face rather more difficulties. This situation is reversed when considering the *interaction* between these same factors of stimulus quality and word frequency when the background context changes such that nonwords are no longer present in the list. Now it is difficult to see how a Sternbergian arrangement of processes can produce the observed outcome, whereas cascaded processing and/or IA produces the observed pattern with ease (at least in the context of DRC (see Reynolds & Besner, 2004).

One resolution to this conundrum, as proposed in the introduction, is that serially arranged and discrete processes are in play when additive effects are observed, whereas cascaded processing and/or IA are in play when the interaction is observed. This proposal is simply not a post hoc account generated to explain these data; rather, it is a hypothesis generated to explain other data, also discussed in the introduction, and it predicted the outcome of the experiments reported here. That said few psycholinguists are likely to find the explanation offered here appealing. Be that as it may, the empirical pattern of data appears clear.

The lexicalization hypothesis: looking forward

The lexicalization account is not without its own issues. First, in the current experiments the neighborhood density of the nonwords was relatively high level. One implication of the lexicalization account as expressed here is that when nonword N is low (i.e. the nonwords have few neighbors) the probability of lexical capture (pronouncing a nonword as a word) would decrease. This should therefore reduce the probability of subjects using a thresholded mode of processing. This line of reasoning leads to the expectation of an interaction between word frequency and stimulus quality when the nonwords mixed with the words are low N.

Second, Blais and Besner (2007) reported a three-way interaction between repetition, lexicality (words versus nonwords) and stimulus quality when reading aloud (the lag between repetitions was 16 items). Repetition and stimulus quality interacted for the words, but had additive effects for the nonwords. The Blais and Besner results are not surprising on their own, but the interaction between repetition and stimulus quality in the nonword context is unexpected given the account suggested here. If the letter level is thresholded when nonwords are present (as the lexicalization hypothesis assumes) then repetition should also have been additive with stimulus quality. It remains to be seen what insights can be gleaned from an experiment which replicates the Blais and Besner experiment and also manipulates word frequency.

An alternative account

Can CDP+ simulate these data by emphasizing the non-lexical route, which Perry et al. (2007) claim is functionally thresholded? Ziegler (personal communication) suggested that:

> "CDP+ can produce both an interaction between stimulus quality and frequency as well as an additive effect of these two variables. Whether one or the other is obtained does seem to depend on the strength of the nonlexical route. If these arguments are correct, it should be the case that the size of the frequency effect is reduced in the mixed list compared to the pure list. The size of the frequency effect would therefore provide an important marker for the strategic shift from

> lexical to nonlexical processing. This information would provide a crucial constraint for further simulations of these effects."

We are unable to follow exactly how Ziegler's account could work, but then we are not modelers. Our comprehension failure here may only mean that there are subtleties associated with this computational model that we do not yet appreciate. A demonstration that CDP+ can simulate the present results would therefore be important, but a critical aspect of the computational modeling enterprise surely involves understanding *why* the models behave the way they do.

Our second point is that Ziegler is very clear that the size of the word frequency effect across the presence/absence of nonwords must differ according to his route shift emphasis account. We note, however, that the data do not support this claim given that in Experiment 3 the magnitude of the word frequency effect in the *bright* condition was the same size when nonwords were present as when they were absent ($F_1 < 1$).

*The **why** question*

Why do the processing dynamics appear to vary so dramatically across the present contexts? The proposal advanced here is that cascaded processing increases the probability of lexicalizing the pronunciation of a non-word when stimulus quality is low, something that participants should wish to avoid given the typical emphasis on accuracy in these kinds of experiments. Hence, they engage in discrete processing at an early level throughout the experiment, leading to additive effects of stimulus quality and word frequency in the presence of nonwords. When only words are present in the reading aloud task then cascaded processing is adopted because in this context lexicalization is not a potential problem.

The ***what*** *question*

What general implications, if any, do the results discussed here have for understanding visual word recognition? One major account of visual word recognition is that many of the subprocesses are "automatic" in one way or another (e.g., Brown, Gore & Carr, 2002; see also the long list of investigators noted in Reynolds & Besner, 2006). A strong view of such automaticity is that it is context independent. This claim is problematic given the present results (and many others). The typical response to this point is to claim that automaticity is context dependent. However, assuming that processing is automatic but context *dependent* in an unspecified way is unappealing to us because it is too theoretically vacuous and potentially circular at present. A more profitable direction is to look for additional examples where the joint effects of a pair of factors change depending on the level of a third factor (see Brown, Stolz & Besner, 2006; Ferguson, Robidoux & Besner, 2008; Stolz & Neely, 1995). At the very least, this will serve to broaden the empirical base that will need a theoretical perspective. More generally, we currently lack any broad theory of context effects that will help guide such a search.

Conclusions

The lexicalization account proposed here should be viewed as tentative; it clearly needs to be explored further. Whatever ones theoretical predilections, the central implication of the results discussed here is that the processing underlying aspects of visual word recognition are rather more dynamic than widely assumed. This conclusion is neither particularly welcome nor especially appealing to the extent that it makes theorizing about mental performance more difficult. Nonetheless, it reflects a direction (e.g., see also Balota,Yap, Tse & Besner, 2008) that the field at large will need to take into account when attempting to explain skilled "reading" in particular and mental performance more generally.

References

Adelman, J.S., Brown, G.D., & Quesada, J.F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17,* 814-823.

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

Ashby, G.F. (1982). Deriving exact predictions from the cascade model. *Psychological Review, 89,* 599-607.

Baayen, R.H. (in press) Analyzing linguistic data. A practical introduction to statistics. Cambridge University Press.

Balota, D.A., & Abrams, R.A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning,Memory, and Cognition, 21,* 289-1302.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 340-357.

Becker, C.A. (1976) Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 556-566.

Becker, C.A., & Killion, T.H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 389-401.

Besner, D. (1983). Basic decoding components in reading: Two dissociable feature extraction processes. *Canadian Journal of Psychology, 37,* 492-438.

Besner, D. (2006). Visual language processing and additive effects of multiple factors on timed performance: A challenge for the interactive activation framework? *PsyCrit* website: http://psycrit.com/

Besner, D., & Roberts, M.A. (2003). Reading nonwords aloud: Results requiring change in the dual route cascaded model. *Psychonomic Bulletin & Review, 10,* 398-404.

Besner, D., Wartak, S., & Robidoux, S. (2008). Constraints on computational models of basic processes in reading. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 242-250.

Blais, C., & Besner, D. (2007). Reading aloud: When the effect of stimulus quality distinguishes between cascaded and thresholded components. *Experimental Psychology, 54*, 215-224.

Borowsky, R., & Besner, D. (1993) Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813-840.

Borowsky, R., & Besner, D. (2006). Parallel distributed processing and lexical-semantic effects in visual word recognition: Are a few stages necessary? *Psychological Review, 113*, 181– 195.

Brown, T.L., Gore, C.L., & Carr, T.H. (2002). Visual attention and word recognition in Stroop color naming: Is word recognition automatic? *Journal of Experimental Psychology: General, 131*, 220-240.

Brown, M., Stolz, J.A., & Besner, D. (2006) Dissociative effects of stimulus quality on semantic and morphological contexts in visual word recognition. *Canadian Journal of Experimental Psychology, 60,* 190-199.

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI* . Hillsdale, NJ. (pp. 535-555).

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.

Ferguson, R., Robidoux, S. & Besner, D. (2008). Reading aloud: Evidence for contextual control over lexical activation. *Journal of Experimental Psychology: Human Perception and Performance* (under revision).

Forster, K.I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627-635.

Forster, K.I., & Forster, J.C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35,* 116-124.

Forster, K.I., & Hector, J. (2002) Cascaded versus noncascaded models of lexical and semantic processing: The turple effect. *Memory & Cognition, 30,* 1106-1117.

Frederiksen, J.R., & Kroll, J.F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 361-379.

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

Masson, M. & Loftus, G.R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57,* 203-220.

McCann, R.S. & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 14-24.

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

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287-330.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 3-36). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Monsell, S., Patterson, K.E., Graham, A., Hughes, C.H., & Milroy, R. (1992). Lexical and sub-lexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory & Cognition,18*, 452-467.

Morton, J. (1969) Interaction of information in word recognition. *Psychological Review, 76,* 165-178.

Mulatti, C., Reynolds, M. G., & Besner, D. (2006). Neighborhood effect in reading aloud: New findings and new challenges for computational models. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 799-810.

Murray, W.S., & Forster, K.I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111*, 721-756.

Norris, D. (1984). The effects of frequency, repetition and stimulus quality in visual word recognition. *Quarterly Journal of Experimental Psychology, 36A,* 507-518.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113,* 327-357.

O'Malley, S., Reynolds, M.G. & Besner, D. (2007). Qualitative differences between the joint effects of stimulus quality and word frequency in reading aloud and lexical decision: Extensions to Yap and Balota. *Journal of Experimental Psychology: Learning Memory & Cognition,* 33, 451-458.

Perry, C., Ziegler, J.C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114,* 273-315.

Plaut, D.C., & Booth, J.R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107,* 786-823.

Plaut, D.C., & Booth, J.R. (2006). More modeling but still no stages: Reply to Borowsky and Besner. *Psychological Review, 113,* 196-200.

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

Plourde, C.E., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology, 51,* 181-194.

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behaviour Research Methods*, 39, *859-862.*

R Development Core Team (2004). R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rastle, K., & Davis, M.H. (2002). On the complexities of measuring naming**.** *Journal of Experimental Psychology: Human Perception and Performance, 28*, 307-314.

Reynolds, M.G., & Besner, D. (2004). Neighborhood density, word frequency, and spelling-sound regularity effects in naming: Similarities and differences between skilled readers and the dual route cascaded computational model. *Canadian Journal of Experimental Psychology, 58*, 13-31.

Reynolds, M.G., & Besner, D. (2005). Basic processes in reading: A critical review of pseudohomophone effects in reading aloud and a new computational account. *Psychonomic Bulletin & Review*, 12, 622-646.

Reynolds, M.G., & Besner, D. (2006). Reading aloud is not automatic: Processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology: Human Perception and Performance*, *32,* 1303-1323.

Reynolds, M.G., & Besner, D. (2008). Contextual effects on reading aloud: Evidence for pathway control. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 50-64.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer and S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611-653). Cambridge, MA: MIT Press.

Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-Prime user's guide.* Pittsburgh: Psychology Software Tools Inc.

Seidenberg, M.S., & McClelland, J.L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523-568.

Stanners, R.F., Jastrzembski, J.E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning and Verbal Behavior*, *14*, 259-264.

Sternberg, S. (1967). Two operations in character recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics, 2,* 45-53.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 267-315.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues: An invitation to cognitive science* (pp. 703-863). Cambridge, MA: MIT Press.

Stolz, J.A., & Neely, J.H. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 596-611.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology, 47A*, 631-650.

Vincent, S.B. (1912). The function of vibrissae in the behavior of the white rat. *Behavioral Monographs, 1*, (Whole No. 5).

Wilding, J.M. (1988). The interaction of word frequency and stimulus quality in the lexical decision task: Now you see it, now you don't. *Quarterly Journal of Experimental Psychology,40A*, 757-770.

Yap, M.J., & Balota, D.A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 274 -296.

Yap, M.J., Balota, D.A., Tse, C., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency on lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning Memory and Cognition.(in press).*

Footnotes

1. Wilding (1988) and Norris (1984) both report experiments in which there was an interaction between stimulus quality and word frequency in lexical decision when there was a very long ITI (over 3 seconds). However, both authors reported additive effects of these factors when the inter-trial interval was considerably shorter. Wilding argues that the interaction seen with a long ITI says little or nothing about reading per se, and more about attention and recovery from long fore-periods.

2. To manipulate stimulus quality, Yap and Balota (2007) rapidly alternated a mask and the letter string, whereas O'Malley et al. (2007) used contrast reduction. Both manipulations yielded an interaction between stimulus quality and word frequency in reading aloud and additive effects of stimulus quality and word frequency in lexical decision.

3. Pollatsek raised an important direction for future research. How would the present claims play out in the context of eye movement studies? One approach might be to have subjects read prose and use a high proportion of very low frequency words that are unknown to the subjects so as to mimic the presence of nonwords.

4. When nonwords were included the main effect of frequency was only marginal in the item analysis. However standard item analyses are generally associated with low power. To investigate this further, we fitted the word data (in the nonword condition) to a linear mixed-effects model as outlined by Baayen (in press), with subject and items as crossed

random effects. There was a main effect stimulus quality, β^ = 115.2, $t(6601) = 15.5$, $p < .001$ and, critically, a main effect of word frequency, β^ = 8.5, $t(6601) = 2.02$, $p < .05$. The interaction did not approach significance. We complete the specification of the model by reporting the standard deviation of the random effects. The standard deviation of the random effect of Word was estimated at 26.9. In this model there were two random effects associated with Subjects. First, the standard deviation of the by-subject adjustments was estimated at 61.7. Additionally, subjects were differentially sensitive to stimulus quality, (log-likelihood ratio = 37896, p < .0001), the standard deviation for the by-subject adjustments to the quality coefficient was 42.5, and the correlation of the by-subject adjustment to intercept and Family size was .007. The residual standard deviation was 71.9.

5. We note with interest that the newest computational version of DRC (version 1.1.4 is available on a website) includes an option for thresholding various modules.

APPENDICES

Appendix A

Items used in Experiments 1 & 2.

High Frequency Words			Low Frequency Words		
air	high	sight	adept	hick	skull
artist	home	sign	anvil	hobby	slab
ball	hotel	six	apron	howl	smash
behind	house	sort	arid	jargon	sock
bottom	human	sound	awe	jolt	spice
carry	index	south	banjo	lass	spin
case	job	stage	bean	loft	spoil
cause	kitchen	start	beggar	lord	spoon
chance	labor	station	boom	lust	spy
chief	land	story	canon	magnet	stack
child	large	student	carve	march	stink
church	later	study	cavern	mask	stool
claim	less	style	cheer	mayor	thorn
close	letter	table	chore	mentor	thrill
cold	life	there	comic	merry	torch
color	like	thing	concede	mesh	traitor
cut	long	top	coral	mint	tramp
desk	loss	total	coward	mule	trout
dinner	machine	view	crate	munch	uncle
doctor	major	voice	dense	ounce	valve
drive	money	wait	dual	pail	veil
eight	motor	well	dummy	peach	vile
father	music	wish	dune	plump	weave
feel	name	woman	dusk	polar	weld
feet	novel	world	edit	queen	witch
film	paid	yellow	embark	rim	wizard
final	party	you	exit	roast	wreck
fire	picture		fare	rude	zoo
food	piece		float	ruler	
force	place		flu	rumor	
free	plane		flute	scoop	
gas	pretty		gaze	scratch	
girl	road		gorge	scrub	
goal	sea		grape	seam	
gun	ship		gravel	servant	
happen	side		haste	shrug	

Appendix A (cont'd)

Non-words					
ard	dort	hout	pote	sline	trought
baunt	drass	jame	pouse	slint	turl
beld	drave	jatch	prease	slock	vares
bez	dridge	jate	prench	sloss	vatch
blinch	dup	jide	pribe	soun	vate
bloss	fape	jight	prige	spack	vead
boke	fatch	jod	pright	spale	vight
bome	faunt	kets	prine	spart	voe
borts	feak	kig	pross	spave	vorn
bouth	feen	lafe	prought	spile	wase
brench	fim	lails	prown	spint	weam
brotch	fitch	laught	rark	splorge	wec
cas	flass	lecs	rell	spludge	wime
chack	flink	lel	ringe	spoot	yeal
chank	foat	lorse	routh	spourge	yide
chone	frange	losh	ruv	sprine	yight
citch	frudge	lut	sark	sprong	yought
clane	fuff	meam	scarch	stame	zear
clase	gatch	meck	scranch	statch	zey
closs	geal	mert	scripe	steet	zight
clunch	gight	mook	shace	stell	
coof	glave	mout	shafe	stook	
couse	goss	mowl	shafe	stope	
cown	gought	nass	shalk	stort	
coys	gowl	neek	shate	stratch	
crame	grabe	noke	shick	strice	
crase	grafe	noot	shink	strofe	
creat	grake	oatch	shint	tarm	
crench	grame	pame	shork	tatch	
crols	grare	parn	shrife	teaf	
crought	gress	pash	shunk	thrine	
crove	grought	pench	sife	touth	
dard	hean	petch	slame	toz	
datch	hest	pight	slank	trass	
dight	hez	pise	slare	trime	
doke	hoil	porse	slass	trine	

Appendix B

Items used in Experiment 3.

High Frequency			Low Frequency		
about	going	science	aloft	hill	stake
above	gone	small	apt	honey	stare
act	great	space	ballot	label	stead
alone	ground	spirit	bird	lath	stealth
along	hall	still	blank	lava	steer
around	hands	street	blob	lessen	stew
back	hard	strong	canoe	luck	stuff
basis	have	system	cast	mall	sword
before	help	taken	cat	metal	taste
began	just	talk	cheek	mirror	tend
black	kind	then	copy	moot	thief
both	late	think	cough	murky	thorn
brown	left	thought	creep	nail	thread
call	level	three	cube	nerve	toad
class	local	type	dent	ocean	track
clear	man	under	dial	owl	trench
control	matter	until	digit	photo	twinkle
could	might	very	doe	plaza	wart
court	much	view	dreamt	pose	wax
day	nature	west	drill	prep	welcome
death	night	what	evoke	pun	wilt
early	north	which	fairy	raid	witty
end	often	while	flame	roach	wonder
every	once	will	flirt	roar	worst
fact	over	work	flood	rowdy	wrath
felt	paper	would	fluent	royal	zeal
few	peace	wrote	fold	saver	zebra
field	period	year	fray	scout	zip
figure	plan		freeze	scream	
first	point		frock	serum	
floor	power		fuse	slip	
found	provide		gable	slope	
from	quite		gene	smuggle	
front	river		glimpse	snow	
general	run		grief	sprout	
get	same		grunt	squat	

Appendix B (cont'd)

Non-words					
agale	flane	hoint	nint	shog	tove
ank	flench	jalve	norb	sholt	trall
beash	flep	jang	nount	shrime	twing
blap	flesk	jave	numble	sirth	ved
bleck	flinge	jench	nurl	slidge	vinch
blit	flonk	juff	oam	sloat	vink
blounce	floy	junce	pedge	slod	vonce
blunk	frash	jush	pemp	slouth	waber
blut	frew	keast	phick	slyth	wacing
bose	frex	kend	pice	smill	winth
brong	fronk	kurp	plaif	smow	witsy
bup	fusk	laint	plail	speam	woast
charp	gake	leck	plang	spletch	wust
chelk	gect	leet	plax	spodge	yine
chesk	gick	lirge	pleg	sprew	yis
chisk	glane	loach	plimpse	sterm	yold
chort	glept	lolt	plur	strawl	yurk
cisque	glip	louth	pode	strax	zark
clast	glunk	loy	polt	streach	zinge
cloat	goam	mant	pount	strise	
clotch	gope	marp	prant	strunch	
clune	graught	maunch	prate	swone	
coe	greash	meap	praught	swun	
creeze	greem	meath	predge	tade	
crope	greft	medge	quast	tander	
cype	gright	mib	raint	tarch	
dast	grile	minge	rasting	tase	
detch	grimpse	mome	rint	tetch	
dobe	grine	moy	salk	tey	
drail	gurst	moz	scame	thock	
drait	gutch	murse	scole	thoice	
drance	hade	narp	screak	thrase	
drine	hect	neak	screlch	throwd	
dripe	hetch	neeve	scrug	tilch	
feant	hig	nim	sharn	tob	
felly	hilch	ninch	shent	toist	

Appendix C

Individual Participant condition RT (ms) means and Percent Errors for Experiment 1.

	High Frequency		Low Frequency		Nonwords	
Subjects	Clear	Deg	Clear	Deg	Clear	Deg
1	728.6	826.3	796.5	904.7	902.3	963.0
2	560.7	598.7	577.2	635.6	630.7	675.5
3	497.2	525.3	493.5	540.8	515.5	552.0
4	435.0	489.6	439.8	483.0	444.2	508.6
5	567.7	648.0	588.2	653.8	676.3	728.0
6	601.8	623.1	582.3	615.7	606.2	627.9
7	458.3	684.8	480.5	724.9	555.6	843.8
8	562.6	612.9	601.6	665.6	677.8	720.7
9	524.8	614.8	579.2	637.2	571.0	656.7
10	591.1	647.2	653.5	671.8	655.7	701.6
11	563.4	744.8	617.4	794.4	729.3	888.3
12	743.9	785.3	797.5	809.6	838.9	878.6
13	474.1	500.1	477.2	513.2	497.4	542.6
14	458.4	486.2	481.6	527.8	489.3	526.4
15	444.8	495.6	448.0	508.6	457.0	506.1
16	485.6	538.6	497.2	539.8	492.3	564.5
17	499.1	532.6	501.7	556.4	505.3	567.1
18	487.7	605.3	525.1	596.2	506.3	641.9
19	528.3	667.3	562.0	744.7	607.3	787.3
20	546.0	583.9	569.3	630.3	604.3	628.2
21	468.7	539.7	474.9	537.6	495.1	541.6
22	490.1	524.5	525.7	545.5	512.6	560.3
23	486.1	561.4	522.2	579.2	532.8	642.5
24	456.5	509.2	451.0	528.3	467.0	537.2
25	507.5	567.1	514.4	556.3	546.7	565.2
26	463.2	537.1	465.7	595.4	477.3	570.0
27	395.6	436.7	396.4	437.0	421.1	460.2
28	524.0	603.4	526.4	633.7	557.4	625.1
29	505.6	554.0	551.8	593.5	599.8	651.7
30	514.8	640.8	542.9	713.5	615.0	810.0
31	554.2	610.2	600.1	635.2	630.8	675.0
32	512.4	539.8	538.7	577.7	576.5	601.0

Appendix C (cont'd)

	High Frequency		Low Frequency		Nonwords	
Subjects	Clear	Deg	Clear	Deg	Clear	Deg
1	4	4	2	14	7	9
2	0	0	2	2	1	6
3	0	0	0	0	8	3
4	2	0	2	2	4	9
5	2	0	0	0	3	4
6	0	0	0	0	3	7
7	2	0	0	2	4	5
8	0	0	2	0	10	6
9	0	4	0	4	5	6
10	0	0	4	0	4	4
11	0	2	0	4	6	3
12	0	0	4	8	14	12
13	0	0	0	0	7	3
14	0	0	0	0	8	3
15	0	2	2	0	3	9
16	0	2	2	4	0	2
17	0	2	0	2	6	4
18	0	0	0	2	4	5
19	0	0	2	0	1	2
20	0	0	0	0	8	5
21	2	2	0	4	3	7
22	4	0	2	0	4	8
23	0	0	0	4	2	6
24	0	0	2	4	5	4
25	2	0	0	4	2	10
26	0	0	2	6	9	16
27	0	2	0	0	4	3
28	0	2	2	2	7	9
29	0	0	0	8	4	5
30	0	0	2	6	7	16
31	2	2	0	4	5	10
32	2	2	10	8	12	5

Appendix D

Analysis of Variance of reaction times (top table) and errors (bottom table) by word frequency and by stimulus quality for Experiment 1.

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	19860.2	1	19860.2	49.0	.000
Error (frequency)	12566.1	31	405.4		
QUALITY	158484.5	1	158484.5	66.5	.000
Error (Quality)	73918.0	31	2384.5		
FREQUENCY x QUALITY	96.3	1	96.3	.6	.429
Error	4643.0	31	149.8		

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	60.5	1	60.5	15.2	.000
Error (frequency)	123.5	31	4.0		
QUALITY	24.5	1	24.5	7.1	.012
Error (Quality)	107.5	31	3.5		
FREQUENCY x QUALITY	18.0	1	18.0	6.2	.018
Error	90.0	31	2.9		

Appendix E

Individual Participant condition RT (ms) means and Percent Errors for Experiment 2.

	High Frequency		Low Frequency		Nonwords	
Subjects	Clear	Deg	Clear	Deg	Clear	Deg
1	600.8	704.9	643.9	741.6	696.6	848.0
2	490.0	601.8	497.0	595.7	584.0	698.5
3	556.1	630.0	540.8	680.0	678.7	754.9
4	492.0	573.0	521.2	603.7	544.0	653.3
5	462.0	542.5	498.2	583.9	590.5	674.7
6	386.1	459.6	398.4	484.8	411.8	492.6
7	506.0	588.1	533.0	592.5	562.9	650.0
8	658.1	715.6	689.1	740.6	749.7	832.8
9	488.4	540.2	502.7	559.0	503.8	584.6
10	495.3	549.0	518.8	578.8	561.5	626.7
11	508.5	537.5	514.1	590.3	549.2	620.2
12	478.3	541.1	493.3	559.0	542.9	590.4
13	402.6	456.1	403.5	468.2	419.4	477.0
14	401.4	445.1	408.1	469.8	416.9	479.2
15	492.3	566.0	524.5	585.0	565.1	597.9
16	502.9	572.7	552.4	603.5	577.2	620.4
17	458.5	513.0	477.6	527.7	491.1	546.4
18	480.2	517.6	505.3	531.5	496.8	588.5
19	444.6	492.9	461.6	532.0	470.9	506.6
20	451.5	488.7	447.3	511.4	469.8	512.0
21	495.8	563.1	525.2	571.7	583.2	606.7
22	568.6	606.3	557.6	618.1	619.4	658.1
23	477.0	511.4	481.0	528.1	507.3	561.0
24	428.1	481.2	442.6	496.4	477.2	504.6
25	453.2	511.9	473.0	506.4	482.6	523.6
26	399.5	447.1	417.0	460.4	433.0	474.1
27	697.2	745.0	704.3	757.2	844.1	872.6
28	474.2	511.4	512.7	543.0	554.3	593.2
29	541.1	601.2	565.4	566.4	573.7	616.3
30	480.8	529.9	492.9	525.7	558.1	561.8
31	406.9	439.2	435.2	461.2	455.2	478.8
32	539.5	610.0	557.1	612.8	616.6	675.8

Appendix E (cont'd)

	High Frequency		Low Frequency		Nonwords	
Subjects	Clear	Deg	Clear	Deg	Clear	Deg
1	4	0	2	4	2	2
2	0	0	0	0	1	12
3	0	0	12	2	12	8
4	2	2	0	2	9	0
5	0	2	0	2	9	11
6	2	0	0	4	10	10
7	2	8	0	8	17	14
8	2	0	4	6	5	9
9	0	6	6	6	9	18
10	0	4	2	6	19	22
11	4	2	6	14	17	38
12	2	2	2	2	3	13
13	0	0	0	4	11	21
14	2	0	0	4	1	5
15	0	2	0	0	2	7
16	0	0	0	2	1	7
17	0	0	0	2	2	5
18	2	2	8	2	11	21
19	0	4	2	2	3	10
20	0	2	4	10	19	27
21	0	4	2	4	9	15
22	0	2	2	4	10	10
23	0	4	2	8	4	8
24	0	2	0	4	4	4
25	0	2	0	2	3	2
26	2	4	4	8	6	4
27	0	0	0	0	2	3
28	0	2	2	4	9	7
29	0	2	0	2	7	12
30	0	2	4	4	4	12
31	0	0	0	4	4	7
32	2	0	4	4	5	7

Appendix F

Analysis of Variance of reaction times (top table) and errors (bottom table) by word frequency and by stimulus quality for Experiment 2.

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	10706.0	1	10706.0	78.1	.000
Error (frequency)	4252.2	31	137.2		
QUALITY	110868.4	1	110868.4	275.3	.000
Error (Quality)	12482.8	31	402.7		
FREQUENCY x QUALITY	2.0	1	2.0	.0	.901
Error	3925.3	31	126.6		

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	98.0	1	98.0	22.3	.000
Error (frequency)	136.0	31	4.4		
QUALITY	72.0	1	72.0	14.5	.001
Error (Quality)	154.0	31	5.0		
FREQUENCY x QUALITY	6.1	1	6.1	1.6	.218
Error	119.9	31	3.9		

Appendix G

Individual Participant condition RT (ms) means and Percent Errors for Experiment 2. Subjects 1 – 32 were in the words and nonwords condition, whereas subjects 33-56 were in the words only condition

	High Frequency		Low Frequency		Nonwords		High Frequency		Low Frequency		Nonwords	
Subjects	Clear	Deg	Clear	Deg	Clear	Deg	Clear	Deg	Clear	Deg	Clear	Deg
1	568	706	561	734	668	815	0	0	0	2	4	9
2	546	611	531	613	575	633	0	4	2	0	5	10
3	463	599	477	596	522	684	2	8	2	6	9	10
4	472	583	489	590	553	649	0	2	2	2	13	5
5	519	571	519	582	569	644	2	2	10	12	13	20
6	493	641	492	653	597	780	0	8	6	12	21	24
7	467	623	487	673	555	688	8	2	4	12	36	45
8	475	601	492	599	504	657	0	2	2	0	2	9
9	479	625	495	627	567	744	2	2	2	6	9	14
10	473	588	479	605	495	601	0	2	2	10	1	7
11	480	598	504	602	528	658	0	4	0	8	8	15
12	442	538	441	560	508	575	0	6	0	6	4	13
13	515	595	534	586	557	640	0	0	2	2	3	6
14	542	650	569	651	599	718	0	0	2	2	3	4
15	385	464	374	459	398	491	6	4	2	2	3	14
16	504	613	498	595	532	644	4	4	2	14	8	18
17	488	566	500	585	540	602	2	2	0	6	5	8
18	374	490	392	510	423	521	4	4	0	2	7	10
19	584	678	586	718	627	733	0	2	0	0	1	8
20	499	684	507	666	523	735	0	4	2	14	4	32
21	455	583	467	649	518	715	0	6	0	8	3	14
22	545	674	553	657	591	731	0	2	2	0	6	7
23	477	577	471	573	533	624	0	4	0	10	16	27
24	499	583	504	577	554	633	2	2	0	2	5	8
25	455	527	454	522	473	567	0	2	2	0	5	8
26	629	796	609	822	738	898	2	6	2	12	19	15
27	525	630	541	644	630	720	6	0	0	6	5	6
28	675	731	682	761	778	811	2	6	4	6	10	10
29	502	581	527	592	551	632	0	8	2	4	6	9
30	639	733	671	774	712	834	0	2	2	0	5	11
31	517	645	523	621	542	698	2	4	2	2	3	6
32	491	565	488	564	507	599	0	2	0	2	3	8

Appendix G (cont'd)

Subjects	High Frequency		Low Frequency		High Frequency		Low Frequency	
	Clear	Deg	Clear	Deg	Clear	Deg	Clear	Deg
33	483	645	484	668	2	6	2	0
34	489	596	506	614	0	0	0	6
35	434	539	467	562	2	4	0	2
36	514	575	511	578	0	2	0	4
37	508	584	506	608	2	2	0	4
38	489	638	509	722	0	2	4	4
39	492	578	494	575	0	8	2	8
40	427	545	447	550	0	12	4	22
41	582	633	576	647	2	0	4	4
42	444	508	445	566	0	4	4	4
43	476	529	462	549	0	0	4	2
44	332	405	351	403	2	8	0	0
45	471	582	494	610	0	0	0	0
46	434	504	434	502	2	4	0	2
47	495	683	533	737	2	12	2	8
48	585	692	589	735	2	0	2	4
49	474	556	481	585	0	8	4	2
50	437	564	444	567	0	4	0	2
51	482	645	480	658	0	2	2	10
52	531	616	564	688	2	0	2	2
53	368	468	382	474	0	2	0	2
54	426	518	423	521	0	2	0	12
55	389	468	393	488	0	4	2	0
56	519	742	567	831	0	8	2	18

Appendix H

Analysis of Variance of reaction times (top table) and of errors (bottom table) by word frequency, stimulus quality and by context (presence/absence of nonwords) for Experiment 3.

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	10047.9	1	10047.9	39.9	.000
FREQUENCY x CONTEXT	1330.3	1	1330.3	5.3	.025
Error (frequency)	13602.8	54	251.9		
QUALITY	680119.4	1	680119.4	418.6	.000
QUALITY x CONTEXT	165.5	1	165.5	.1	.751
Error (Quality)	87731.7	54	1624.7		
FREQUENCY x QUALITY	1036.3	1	1036.3	8.2	.006
FREQUENCY x QUALITY x CONTEXT	574.8	1	574.8	4.5	.038
Error (freq x sq)	6831.2	54	126.5		
CONTEXT	56402.5	1	56402.5	3.0	.088
Error (Context)	1007402.8	54	18655.6		

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	70.1	1	70.1	9.2	.004
FREQUENCY x CONTEXT	.4	1	.4	.1	.813
Error (frequency)	412.4	54	7.6		
QUALITY	495.4	1	495.4	43.9	.000
QUALITY x CONTEXT	4.5	1	4.5	.4	.530
Error (Quality)	609.4	54	11.3		
FREQUENCY x QUALITY	11.3	1	11.3	1.7	.193
FREQUENCY x QUALITY x CONTEXT	5.9	1	5.9	.9	.344
Error (freq x sq)	350.1	54	6.5		
CONTEXT	.5	1	.5	.0	.854
Error (Context)	843.7	54	15.6		

Appendix I

Analysis of Variance of reaction times (top table) and errors (bottom table) by word frequency and stimulus quality in Experiement 3 for the condition in which nonwords are present.

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	8177.0	1	8177.0	22.4	.000
Error (frequency)	8397.5	23	365.1		
QUALITY	306908.2	1	306908.2	138.3	.000
Error (Quality)	51049.3	23	2219.5		
FREQUENCY x QUALITY	1380.2	1	1380.2	11.1	.003
Error	2864.3	23	124.5		

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	26.0	1	26.0	2.9	.101
Error (frequency)	205.0	23	8.9		
QUALITY	260.0	1	260.0	16.3	.001
Error (Quality)	367.0	23	16.0		
FREQUENCY x QUALITY	.4	1	.4	.1	.808
Error	142.6	23	6.2		

Appendix J

Analysis of Variance of reaction times (top table) and errors (bottom table) by word frequency and stimulus quality in Experiement 3 for the condition in which nonwords are present.

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	2371.9	1	2371.9	14.1	.001
Error (frequency)	5205.4	31	167.9		
QUALITY	384454.9	1	384454.9	324.9	.000
Error (Quality)	36682.4	31	1183.3		
FREQUENCY x QUALITY	39.4	1	39.4	.3	.583
Error	3966.9	31	128.0		

Source	Sums of Squares	df	Mean Square	F	*p*
FREQUENCY	47.5	1	47.5	7.1	.012
Error (frequency)	207.5	31	6.7		
QUALITY	236.5	1	236.5	30.2	.000
Error (Quality)	242.5	31	7.8		
FREQUENCY x QUALITY	19.5	1	19.5	2.9	.098
Error	207.5	31	6.7		