

Reading nonwords aloud: Results requiring change in the dual route cascaded model

DEREK BESNER and MARTHA ANNE ROBERTS
University of Waterloo, Waterloo, Ontario, Canada

The time to name a nonword increases monotonically as letter length increases. The leading computational model of basic processes in reading (Coltheart, Rastle, Perry, Langdon & Ziegler's *dual route cascaded* model) simulates this, because its nonlexical route assigns phonemes to letters serially, left to right, and arguably, this corresponds to what humans do. New simulation work shows that (1) this letter length effect interacts with the effect of slowing the rate of early processing, and (2) the model produces a qualitatively different pattern from that observed with university-level readers. The contrast between simulation and human performance thus illuminates a problem with how the nonlexical route operates in the model, and constrains accounts that can be provided for the human data. Consideration is given to thresholding the output of the letter-level module as a way to modify the model so as to make it possible to simulate the human data.

A number of highly successful computational models of basic processes in reading implement multiple levels of representation (modules) that get activated when a letter string is presented (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981). A central feature of these models is that activation across different modules is *cascaded*. In systems that operate by cascaded processing, there are no thresholds within modules. As soon as there is even a small amount of activation in an early module this flows on to later modules. The present work demonstrates a previously unappreciated consequence when serial processing follows cascaded parallel processing. More specifically, a simple manipulation of processing rate in the model produces a qualitatively different pattern from that produced by university-level readers. This effect is illustrated in the context of nonlexical processing by Coltheart and colleagues' computational model.

The Dual Route Cascaded Model

The *dual route cascaded* model (hereafter DRC) is very successful at accounting for a wide variety of findings in naming and lexical decision tasks. Indeed, its authors list

18 phenomena from the naming task that the model simulates successfully. Coltheart et al. (2001) comment that

the set of phenomena that the DRC model can simulate is much larger than the set that any other current computational model of reading aloud can simulate; and, to the best of our knowledge, there is no effect seen in reading aloud that any of these other models can simulate but that DRC cannot. (p. 251)

Coltheart et al. thus consider DRC to be the most successful computational model of reading aloud.

DRC consists of three routes; the lexical semantic route, the lexical nonsemantic route, and the nonlexical GPC (grapheme-phoneme conversion) route (see Figure 1). A semantic system is not yet implemented in the model. The knowledge bases in the two remaining routes differ; the lexical system is based on word-specific knowledge, whereas the nonlexical system is based on a set of sublexical spelling-sound correspondence rules. Briefly, the correct pronunciation for all words known to the model can be produced by the lexical route's operating in isolation, and the correct pronunciation of a nonword can be produced by the nonlexical route's operating in isolation. The lexical route's operating in isolation can not produce the correct pronunciation of a nonword. The nonlexical route's operating in isolation produces the correct pronunciation of all words that obey the spelling-sound correspondence rules, but regularizes the pronunciation of words that are exceptions to these rules (e.g., PINT is pronounced /pɪnt/). When the intact model is operating, both the lexical and the nonlexical routines affect the time to construct a pronunciation for both words and nonwords. A more detailed discussion of the architecture and how the full model operates is provided by Coltheart et al. (2001). For our present purposes, only the operation of the nonlexical route need

This work was supported by Grant A0998 from the Natural Sciences and Engineering Research Council of Canada to D.B. We are particularly grateful to Max Coltheart for extended discussion. We also thank Mike Reynolds, Rym Bougala, and Ernie MacKinnon for comments and technical assistance; Jay McClelland, Jonathan Grainger, Guy Van Orden, and Jeff Miller for their reviews of an earlier version of the manuscript; and particularly Ken Forster for editorial assistance. Address correspondence to D. Besner, Psychology Department, University of Waterloo, Waterloo, ON, N2L 3G1 Canada (e-mail: dbesner@watarts.uwaterloo.ca).

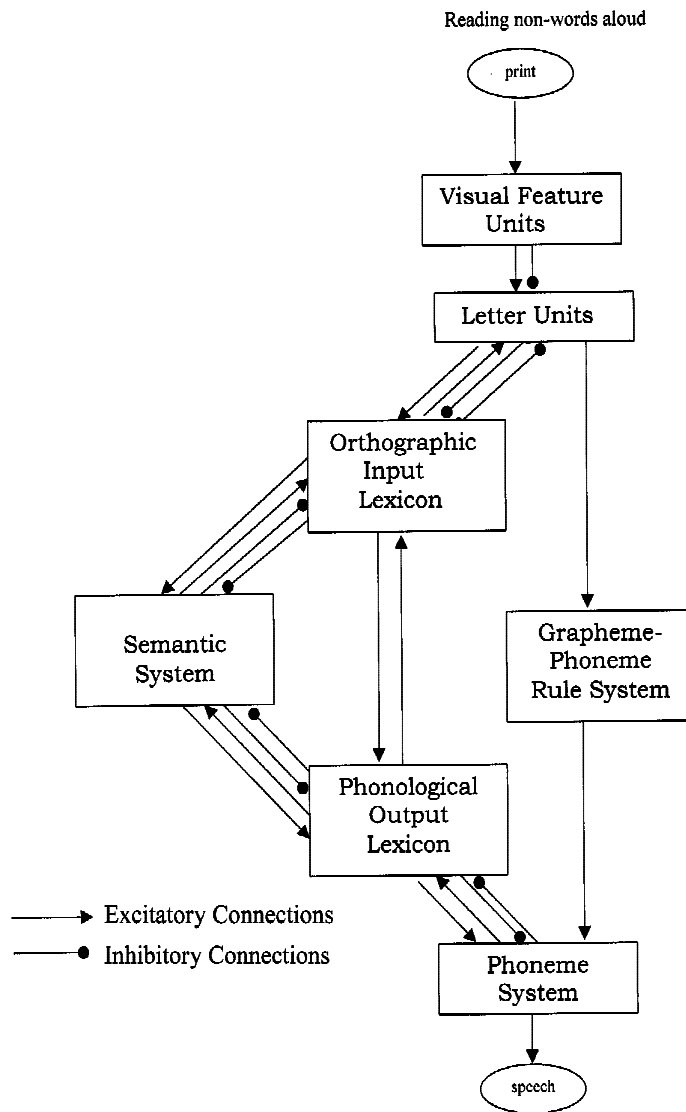


Figure 1. The DRC model of visual word recognition.

be considered. Here, cascaded processing occurs across features, letters, and the grapheme–phoneme translation system. The *phoneme system* is engaged in interactive activation with the *phonological output lexicon*, but for our present purposes this fact can also be ignored without compromising understanding of the central thesis here.

The essentials of the feature- and letter-level processing modules are minor extensions of McClelland and Rumelhart’s (1981) seminal work. Features are extracted in parallel across the stimulus array; letters are also identified in parallel. The phoneme units in the phoneme system are similar in organization to the letter units. The details of the implementation are discussed at length by Coltheart et al. (2001; see pp. 213–218). Coltheart et al. also discuss many details associated with the GPC system that can be ignored

without compromising the main point here: letters activate phonemes in a serial, left to right fashion.

As Coltheart et al. (2001) note, a set number of cycle updates occur across feature and letter levels before the GPC system begins to operate on the first letter in an array. The rule base is searched until an appropriate rule is found to convert that letter to a phoneme, and that phoneme’s unit in the phoneme system then receives some activation. On each subsequent processing cycle, activation is contributed to that phoneme unit in the same way. For our present purposes, the critical point is that processing at this level is serial. *Activation of the second phoneme does not start until a constant number of cycles later, and this is true for each subsequent phoneme.*¹ The model is said to have determined the pronunciation of a monosyllabic letter string

when it has activated all the phonemes of that string to some criterion.²

How Does Cascaded Parallel Processing at the Letter Level Affect Subsequent Serial Processing in the Nonlexical Routine?

Skilled readers' naming reaction time (RT) to nonwords increases monotonically as the number of letters increases (see, e.g., Weekes, 1997). One argument is that this is consistent with serial application of phonemes to letters, and that overt naming starts only when the last phoneme has been identified, just as in DRC (see, e.g., Coltheart et al., 2001; Rastle, Harrington, Coltheart, & Palethorpe, 2000). As we shall see below, this letter length effect turns out to be a good vehicle for examining the effect that cascaded parallel processing at the letter level has on subsequent serial processing in the model's nonlexical route.

Factorial experiments in which a factor that affects the rate of processing (e.g., stimulus quality) is varied in conjunction with another factor (e.g., word frequency, semantic priming, or repetition priming) have been used for over a quarter of a century to evaluate different noncomputational accounts of visual word recognition (e.g., Balota & Abrams, 1995; Becker & Killion, 1977; Besner & Smith, 1992; Besner & Swan, 1982; Borowsky & Besner, 1993; Meyer, Schvaneveldt, & Ruddy, 1975; Plourde & Besner, 1997; Stanners, Jastrzembski, & Westbrook, 1975; Stolz & Neely, 1995). To date, virtually none of the published computational models of visual word recognition has explored whether such results can be successfully simulated (but see Plaut & Booth, 2000). In the present work, we depart from this tradition by exploring the impact of slowing the rate of early processing on the letter length effect in nonword naming performance by the DRC model, and by humans. What predictions can we make here? Because letter-level processing is parallel and cascaded, and because subsequent phonemic processing is serial, it is not easy to predict the outcome with any confidence (at this juncture the reader might like to try to predict the outcome of such a simulation). Running the simulation is the only way to find out what happens in the model. We then want to know whether the results of the simulation correspond to how human readers perform.

Stepping Back

Before we proceed further, it will be necessary to step back and consider how the current version of the implemented model behaves at the feature level. Neither the original model (McClelland & Rumelhart, 1981) nor DRC attempts to simulate how visual input activates feature units. This simply occurs earlier than what the (implemented) front end of these models includes, and so it remains outside their scope. Instead, the DRC model begins its simulation when all the feature units have reached an activation of 1.0. Currently, the feature units are clamped, and reach an activation of 1.0 in one cycle. In short, the feature level is not cascaded, but *thresholded*. In real life, of course,

these units will not go from 0 to 1.0 instantaneously; this will take time. If reducing the rate of processing at this level means that the asymptote is not affected, the effect of such a manipulation would simply delay the time to reach threshold and therefore add a time constant to performance. This would render the model incapable of producing anything but additive effects of factors on RT, where one factor would affect the rate constant at the feature level and a second factor would affect, for example, the letter level, the orthographic input lexicon, the semantic system, or the phonological output lexicon. This outcome is problematic, given the fact that the joint effects of repetition and degradation on RT are not additive but instead interact in lexical decision (words presented for the first time are more affected by degradation than words presented for the second time, even when multiple stimuli intervene between presentations; Besner & Swan, 1982). It is also problematic given the repeated observation that the effect of semantic priming on RT is larger when the target is degraded as compared with when it is clear in both lexical decision and naming (Besner & Smith, 1992; Borowsky & Besner, 1993; Meyer et al., 1975; Stolz & Neely, 1995).

Clearly, then, the ability to produce interactions like those described above requires that slowing the rate of processing must affect the feature level in some other way besides simply delaying the time to reach threshold. One way is to assume that the asymptote of activation at the feature level is affected by a factor such as stimulus quality (i.e., the asymptotic level of activation will be smaller for degraded than for clear stimuli). This can be modeled quite simply by reducing the weights on the feature to letter level in the implemented model, thus returning us to the domain where the model measures performance.

To recapitulate, we considered the effect of slowing the rate of processing both on human performance and in DRC. The implemented model does not attempt to simulate how visual input activates the feature level; consequently, that level is simply thresholded for convenience. The difficulty with thresholding the feature level is that differences in processing rate will not affect anything beyond the feature level, unless the activation asymptote is affected. Data from experiments with college-level readers strongly suggest that a reduction in rate of processing as exemplified by a reduction in stimulus quality serves to affect processes beyond the feature level. We therefore assumed that the asymptote is affected, and modeled a reduction in the rate of processing by reducing the connection weights between the feature and letter levels in DRC.

To return to our original question then, How does cascaded parallel processing at the letter level affect subsequent serial processing in the nonlexical routine? This question was addressed in two simulations where slowing the rate of processing was accomplished by arbitrarily reducing the weights on the connections between the feature and letter levels in the model by 40% (weight reductions of 20% produce the same statistically significant pattern;

the effects are numerically smaller). The stimulus set consisted of 64 monosyllabic nonwords³ whose letter length varied from short (3 and 4 letters) to long (5 and 6 letters). The left-hand panel of Figure 2 shows the results of a simulation in which the model is intact, thus allowing lexical activation to contribute to performance. The results depicted in the middle panel of Figure 2 are from the model's nonlexical route operating in isolation. This was accomplished by zeroing out ("lesioning") the connections between the letter level and the orthographic input lexicon, and zeroing out the connections from the phoneme system to the phonological output lexicon.⁴

The results of these two simulations are very similar. They both show that slowing the rate of processing affects short stimuli *more* than long stimuli relative to a faster rate. More formally, with the model intact there was a main effect of increasing letter length [longer stimuli needed more processing cycles to reach threshold; $F(1,27) = 56.83$, $MS_e = 142.10$, $p < .001$], a main effect of rate [the slower rate yielded more processing cycles; $F(1,27) = 4,460.68$, $MS_e = 1.71$, $p < .001$], and an interaction in which there was a *smaller* effect of letter length when processing rate was slowed [$F(1,27) = 23.81$, $MS_e = .90$, $p < .001$]. With the nonlexical route tested in isolation, there was, as before, a main effect of increasing letter length [$F(1,27) = 53.29$, $MS_e = 163.87$, $p < .001$], a main effect of rate [$F(1,27) = 8942.93$, $MS_e = 0.75$, $p < .001$], and an interaction in which there was a *smaller* effect of letter length when processing rate was slowed [$F(1,27) = 211.68$, $MS_e = 0.21$, $p < .001$].

Why does slowing the rate of processing affect short nonwords more than long ones? In terms of the model, an explanation for this seemingly counterintuitive result is

quite straightforward. The model's engagement in serial processing at the phoneme level means that activation of each successive phoneme only *starts* when the prior phoneme has been receiving activation for a constant number of cycles. Meanwhile, activation at the letter level (operating in parallel across all letter positions) continues. Consequently, while activation of the first phoneme is affected by slowing the rate of processing, the delay associated with the start of activation of each additional phoneme allows ongoing letter-level activation to move closer to asymptote. In the limit, there would be no effect of a slower rate of processing on the naming of a long stimulus given that naming does not start until all phonemes reach threshold (Max Coltheart, personal communication, June 2001).

The idea that a skilled reader would be affected by a processing rate reduction when naming short nonwords but less so when naming long ones strikes many people as counterintuitive. If skilled readers do not produce this result, then some component of the DRC model needs modification. However, retooling such interactive-activation models is a potentially large-scale enterprise, in part because there is no guarantee that modifying the model would preserve the ability to successfully simulate all the phenomena that it currently does. Before attempting to modify the model then, its proponents would likely want to be convinced that human readers produce a pattern that differs from what the model produces. In the experiment reported here, university-level readers named the same set of nonwords tested in the simulation work under both clear and degraded conditions. Our working assumption was that reducing stimulus quality reduces the rate of processing early in the system (see, e.g., Pashler, 1994; Pashler & Johnston, 1989).

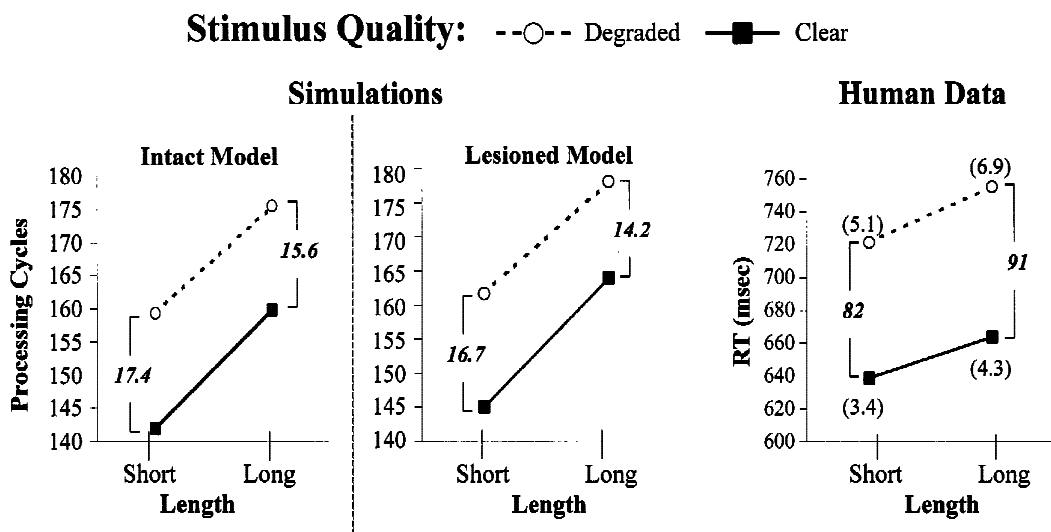


Figure 2. Mean number of processing cycles as a function of stimulus length and stimulus quality for the simulations, and mean reaction time (in milliseconds) and percent error for humans.

METHOD

Subjects

The subjects were 24 undergraduate students from the University of Waterloo. All individuals had normal or corrected-to-normal vision and were native English speakers. Each individual was paid for his/her participation.

Design

The experiment consisted of a 2 (short vs. long items) \times 2 (clear vs. degraded presentation) within-subjects design. Schneider's (1988) MEL 2 software was used. Clear stimuli were displayed in white (RGB values in MEL: 63, 63, 63) on a black background. Stimuli in the degraded presentation were displayed in gray (RGB: 5, 5, 5) on a black background. Stimulus quality and letter length varied randomly throughout a single block of trials.

Materials

Stimuli consisted of 32 pairs of monosyllabic nonwords, each with a neighborhood size of 2, and were matched for initial phoneme and consonant cluster.⁵ There were thus 32 stimuli at each of two levels of length (short vs. long). Items that were 3 and 4 letters in length represented the short length condition; items that were 5 and 6 letters long represented the long length condition.

The items were counterbalanced across levels of stimulus quality in such a way that half the subjects saw an individual item under the clear (degraded) condition, and the remaining subjects saw that item under the degraded (clear) condition. The stimuli were displayed in lowercase in standard MEL 2 software font (system72.fnt) and were presented on a black background.

Apparatus and Procedure

Stimuli were displayed on a 17-in. View Sonic P775 color monitor. Stimulus presentation and data recording were controlled by Micro Experimental Laboratory 2 (MEL 2) software (Schneider, 1988), which was run on a 486 Vault PC. Responses were collected via a Playtronics LS1 microphone headset and voice-key assembly.

The subjects were tested individually and were seated approximately 50 cm from the computer monitor. They were instructed to name each item aloud as quickly and accurately as possible. Each trial began with a fixation rectangle. This rectangle was slightly larger than the largest stimulus presented (0.12° higher and 0.69° wider). The rectangle was displayed in the middle of the screen until the subjects initiated the presentation of the test stimulus by pressing the "z" key, at which time the rectangle was immediately replaced with the item to be named. The display terminated when a pronunciation was made. The experimenter then coded the pronunciation as correct on the basis of the standard set of grapheme-phoneme rules, incorrect, or spoiled (cough, stutter, or voice key failure). Immediately after the experimenter had coded the pronunciation, the fixation rectangle appeared and the next trial began. The experiment began with 24 practice trials consisting of items not contained in the test set, followed by 64 experimental trials presented in a different random order for each subject.

RESULTS

Responses classified as incorrect pronunciations (4.8%) or voice key failures (2.7%) and their matched pairs were removed from the individual subject's RT analysis. RTs for correct responses were subjected to a recursive data trimming procedure in which the criterion cut-off for outlier removal is established by the sample size in each condition for each subject (Van Selst & Jolicœur, 1994). Outlier

removal resulted in 1.8% of the data being discarded overall. Mean outlier values across conditions ranged from 1.5% to 2.1%. A 2 (length) \times 2 (stimulus quality) repeated measures analysis of variance (ANOVA) treating outlier values per subject as data points yielded no significant effects (all $F_s < 1$).

Correct RTs and percentage error for subjects (F_1) and items (F_2) were each analyzed in a 2 (length: short vs. long) \times 2 (stimulus quality: clear vs. degraded) repeated measures ANOVA. Owing to extremely high error rates (in excess of 37%), four items and their matched partners were removed from all analyses. These items are identified in the Appendix. The ANOVA on RTs yielded a main effect of length, such that long items were named more slowly than short items [$F_1(1,23) = 33.61, MS_e = 568.28, p < .001; F_2(1,27) = 7.81, MS_e = 2,990.47, p < .01$], and a main effect of stimulus quality, in which degraded stimuli took longer to name than clear ones [$F_1(1,23) = 37.22, MS_e = 3,671.28, p < .001; F_2(1,27) = 32.51, MS_e = 6,401.70, p < .001$]. No interaction between length and stimulus quality was detected in either subject or item analyses ($F_s < 1$).

Analysis of the errors yielded no main effect of length [$F_1(1,23) = 1.20, MS_e = 39.97, p = .29; F_2 < 1$], but a main effect of stimulus quality such that more errors were made to degraded than clear stimuli [$F_1(1,23) = 5.38, MS_e = 18.17, p < .05; F_2(1,27) = 2.96, MS_e = 43.74, p = .10$]. No interaction between length and stimulus was detected in either subject or item analyses ($F_s < 1$). The RT and error data can be seen in the right-hand panel of Figure 2.

DISCUSSION

The results from the simulation and the experiment are clear. There is a qualitative difference between the pattern produced by the simulation and the pattern produced by university-level readers. The DRC model produces an interaction between length and processing rate in which longer nonwords are *less* affected by a decrease in processing rate than are shorter nonwords. In contrast, the humans produced statistically additive effects of length and stimulus quality on RT. These data appear to call for a modification to the way the DRC model processes along the nonlexical route.

At this juncture, a question that might first be raised concerns whether there are other ways of implementing a reduction in stimulus quality in the model. The answer to this question is surely "yes," but the critical issue here is whether some other way of achieving this reduction would fail to have the same effect overall (i.e., slowing the rate of activation in the letter units).

What changes should be made to the implemented parts of the model? One suggestion that has been made to us repeatedly by reviewers and colleagues is that the serial processing component of the nonlexical route should be abandoned, rather than viewing the present results as critically damaging to the model. This view strikes us as odd;

abandoning the serial processing assumption is surely “critically damaging” to the extent that such a major change would make the model a very different one. More importantly, such a major change is best delayed until it can be shown via simulation that it would not prevent the model from simulating other phenomena that its authors attribute to serial processing, or serial processing combined with parallel processing in the lexical route—for example, the serial position \times regularity interaction seen for reading words aloud (see Rastle & Coltheart, 1999; Roberts, Rastle, Coltheart & Besner, 2003), the whammy effect seen for reading nonwords aloud (Rastle & Coltheart, 1998), the serial position \times whammy interaction discussed by Rastle and Coltheart (1998), and the increasing linear effect of number of letters on RT in nonword naming. Indeed, these phenomena are precisely the ones that are not successfully simulated by other currently existing computational models that, among other things, lack a serial processing component.

Modifying the DRC Model: A Proposal

To return to the present results, we have produced an account of why DRC produces the interaction between processing rate and stimulus length in the simulation data. To recapitulate, by hypothesis the delay incurred by serial processing at the phoneme level allows the effect of the reduction in the rate of processing at the letter level to be more fully absorbed for stimuli with more rather than fewer letters. A successful modification to the model must therefore eliminate this interaction given that the human data produced additive effects. One approach would be to *threshold* the output of a module somewhere in the processing sequence (Sternberg, 1969, 1998). One such possibility was examined in the introduction (thresholding the *feature* units) and rejected on the grounds that it would prevent the model’s being able to simulate several well established two-way interactions in word recognition performance. A different possibility would be to threshold the output of the *letter* level. If the letter-level module does not pass activation to the grapheme–phoneme conversion process until a threshold is reached, the interaction in the simulations reported here would likely not occur. Instead, additive effects of processing rate and letter length would be expected. It turns out that this is indeed the case (Besner, Reynolds, & Chang, 2002). Thresholding the letter level may solve other problems as well (i.e., produce additive effects of stimulus quality and word frequency) and, given that there is feedback between the lexical and letter levels, still allow for the standard interaction between stimulus quality and semantic priming.

Conclusions

Several conclusions may be drawn here. One is that the present work serves as a useful reminder that a computational model should be assessed by running the simulation rather than assuming that it will produce a particular outcome.

The two most central conclusions, however, are the following: First, a new result has been reported that any adequate model of visual word recognition processes will need to be able to explain. Second, the qualitative difference seen here between the outcome of the simulations and the results of the experiment on university-level readers suggests that modification to the currently most successful computational model of word recognition and reading is needed. The proposal entertained here is that the output of the letter-level module be thresholded rather than have the model completely cascaded as in the present implementation. It remains to be seen whether such a modification will leave intact the model’s ability to simulate all the phenomena it currently does, or whether further modification will be necessary.

REFERENCES

- BALOTA, D., & ABRAMS, D. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1289-1302.
- BECKER, C. A., & KILLION, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 389-401.
- BESNER, D., REYNOLDS, M., & CHANG, K. (2002). *Basic processes in reading: Evolution of the dual route cascaded model*. (Manuscript under revision)
- BESNER, D., & SMITH, M. (1992). Models of visual word recognition: When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 468-482.
- BESNER, D., & SWAN, M. (1982). Models of lexical access in visual word recognition. *Quarterly Journal of Experimental Psychology*, **34A**, 313-325.
- BOROWSKY, R., & BESNER, D. (1993). Visual word recognition: A multi-stage activation model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 813-840.
- 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.
- GRAINGER, J., & JACOBS, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, **103**, 518-565.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, **88**, 375-407.
- MEYER, D. E., SCHVANEVELDT, R. W., & RUDDY, M. G. (1975). Loci of contextual effects on visual word recognition. In P. M. A. Rabbitt & S. Dornic (Eds.) *Attention and Performance V* (pp. 98-118). San Diego: Academic Press.
- PASHLER, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, **116**, 220-244.
- PASHLER, H., & JOHNSTON, J. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *Quarterly Journal of Experimental Psychology*, **41A**, 19-45.
- 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.
- PLOURDE, C., & BESNER, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology*, **51**, 181-194.
- RASTLE, K., & COLTHEART, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review*, **5**, 277-282.
- RASTLE, K., & COLTHEART, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 482-503.
- RASTLE, K., HARRINGTON, J., COLTHEART, M., & PALETHORPE, S.

- (2000). Reading aloud begins when the computation of phonology is complete. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1178-1191.
- ROBERTS, M. A., RASTLE, K., COLTHEART, M., & BESNER, D. (2003). When parallel processing in visual word recognition is not enough: New evidence from naming. *Psychonomic Bulletin & Review*, **10**, 405-414.
- SCHNEIDER, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, & Computers*, **20**, 206-217.
- STANNERS, R. F., JASTRZEMBSKI, J. E., & WESTBROOK, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning & Verbal Behavior*, **14**, 259-264.
- STERNBERG, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, **30**, 276-315.
- STERNBERG, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Vol. 4. Methods, models, and conceptual issues* (pp. 703-863). Cambridge, MA: MIT Press.
- STOLZ, J. A., & NEELY, J. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 596-611.
- VAN SELST, M., & JOLICŒUR, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, **47A**, 631-650.
- WEEKES, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology*, **50A**, 439-456.
- ZIEGLER, J. C., PERRY, C., JACOBS, A. M., & BRAUN, M. (2001). Identical words are read differently in different languages. *Psychological Science*, **12**, 379-384.

NOTES

1. It seems likely that the nonlexical route in DRC will also have to incorporate units that are larger than single letters (see, e.g., Ziegler, Perry, Jacobs, & Braun, 2001).
2. A scanning process across all phoneme positions evaluates whether activation at each phoneme position is above threshold at the end of each processing cycle. This scanning process occurs between processing cycles and therefore does not contribute to DRC's naming time.
3. Four items and their matched partners were removed from the simulations in order to make them consistent with the analysis of the human data.
4. The implemented model is available at <http://www.macqs.mq.edu.au/~max/DRC/Model/>. Implementing the weight changes at the feature level used here involves changing a single parameter value in the model. Similarly, the "lesions" done here involve nothing more than zeroing out two sets of connection weights.
5. In our initial attempt to address the issue of what pattern is produced by the joint effects of length and stimulus quality, we started with the stimulus set from Weekes (1997). Long items were named slower than short items in both subject and item analyses. However, Weekes's stimuli are confounded with neighborhood size in such a way that longer items have fewer neighbors. When Weekes partialled out *N*, a length effect remained. When the effect of *N* was partialled out of our data, no effect of length remained in either the subject or the item analysis. Whatever the reason for the differences between what we found and what Weekes found, we needed a new stimulus set for the present work. We are grateful to Max Coltheart and Kathleen Rastle for providing us with a stimulus set that varied in length but was controlled for neighborhood size, initial phoneme, and initial consonant cluster.

APPENDIX
Stimuli

Length		Length	
Short	Long	Short	Long
dra	dralt	flis	flist
fla	left	frev	frast
kal	kacts	frin	frolts
plu	plept	grys	gropts
pru	practs	kolf	kolfs
sko	scost	lilm* (42%)	linxed
sma	smast	ninc* (57%)	nisks
spo	spolts	preb	pramps
twa	twept	smem	smums* (38%)
zof	zolks	sneb	snept
bleg	bleds	snem* (43%)	snent
blic	blemps	stiv	stasks
cril	crask	stol	stonds
drek	drapt	twag	twilt
drob	drists	twon	twumps
flif	flents	zics	zimps

*Items removed because of high error rate in any one cell (rates given).