Journal of Experimental Psychology:
Human Perception and Performance
2006, Vol. 32, No. 5, 1120–1135

0096-1523/06/$12.00 DOI: 10.1037/0096-1523.32.5.1120

Word Length and Lexical Activation: Longer Is Better

Mark A. Pitt
Ohio State University

Arthur G. Samuel
Stony Brook University, State University of New York

Many models of spoken word recognition posit the existence of lexical and sublexical representations, with excitatory and inhibitory mechanisms used to affect the activation levels of such representations. Bottom-up evidence provides excitatory input, and inhibition from phonetically similar representations leads to lexical competition. In such a system, long words should produce stronger lexical activation than short words, for 2 reasons: Long words provide more bottom-up evidence than short words, and short words are subject to greater inhibition due to the existence of more similar words. Four experiments provide evidence for this view. In addition, reaction-time-based partitioning of the data shows that long words generate greater activation that is available both earlier and for a longer time than is the case for short words. As a result, lexical influences on phoneme identification are extremely robust for long words but are quite fragile and condition-dependent for short words. Models of word recognition must consider words of all lengths to capture the true dynamics of lexical activation.

Keywords: spoken word recognition, lexical activation, word length, phoneme identification

In verbal communication, the spoken message must have sufficient clarity on multiple linguistic dimensions (e.g., articulation, appropriate word choice, proper word order) for comprehension to succeed. Consider just one stage of the communication chain, that of recognizing spoken words. The listener must correctly recognize each word intended by the talker, but doing so requires overcoming the high degree of variability in the speech signal. Phonetic, phonological, talker, and speech-rate changes can add a great deal of variability to the speech signal, potentially making it difficult to recover the intended word. Flexibility in word processing seems to be essential for successful recognition.

Connectionist models of speech processing achieve some of this flexibility through two complementary mechanisms: excitation and inhibition. In localist varieties (TRACE: McClelland & Elman, 1986; Shortlist: Norris, 1994; Merge: Norris, McQueen, & Cutler, 2000), each word is represented as a node in a network that is excited by the degree to which it matches the incoming speech signal; nodes are inhibited by other word nodes as a function of how well they also match the signal. Similarly, sublexical units (e.g., phonemes) are also excited by matching information in the input and inhibited by competing sublexical units. At the lexical level, the combination of excitation and inhibition controls a lexical entry's activation level, which represents how well it matches a spoken word relative to all other words in the lexicon. Distributed connectionist models (e.g., the distributed cohort model [DCM]: Gaskell & Marslen-Wilson, 1997, 1999) use the same mechanisms but, through training, form shared representations of words in their hidden units.

Activation has proven to be a productive metaphor for contemplating the dynamics of recognition within this modeling tradition. In particular, the gradedness of activation nicely captures the recognition system's sensitivity to the quality of the speech signal and the composition of the lexicon, as a host of studies have demonstrated (Connine, Blasko, & Titone, 1993; Davis, Marslen-Wilson, & Gaskell, 2002; Gaskell & Marslen-Wilson, 2002; Vroomen & de Gelder, 1995; Warren & Marslen-Wilson, 1987; Whalen, 1981).

From this perspective, a central goal of modeling word recognition is to specify the dynamic changes in lexical activation as a word is being heard. Word recognition models generate a hypothetical activation contour, which can be compared with observed performance under various experimental conditions. As noted, the two most important factors affecting these contours will be (a) the accumulation of evidence for a lexical entry, based on its match to the signal, and (b) inhibitory effects of competing lexical entries. To test models, then, one needs to find manipulations that will affect these two factors strongly. In the current study, we focus on one such manipulation that has been surprisingly neglected: word length.

The effect of word length on these two factors should be relatively straightforward. Consider the contrast between a short word like "bet" and a long word like "dangerous." If lexical activation depends on the accumulation of bottom-up evidence, then the longer word has an inherent advantage over the shorter one: There are more phonemes to support the lexical activation of "dangerous" than of "bet." There should also be more lexical competition for "bet" than for "dangerous," because there are many words that are phonetically similar to "bet" (e.g., "bit," "met," "bed") but none that are equally close to "dangerous."

This work was supported by Research Grants DC004330 from the National Institute on Deafness and Other Communication Disorders and MH51663 from the National Institute of Mental Health. We thank Erik Tracy, James Ashford, Paul Shilling, Sally Blechschmidt, Jessica Snider, Nicholas Altieri, and Anastasiya Lysenko for help in testing participants. We thank Donna Kat for her help in conducting the research.

Correspondence concerning this article should be addressed to Mark A. Pitt, Department of Psychology, Ohio State University, 1835 Neil Avenue, Columbus, OH 43210-1222, or to Arthur G. Samuel, Department of Psychology, Stony Brook University, State University of New York, Stony Brook, NY 11794-2500. E-mail: pitt.2@osu.edu or asamuel@ms.cc.sunysb.edu

Thus, both factors would be expected to lead to greater lexical activation of the longer word than of the shorter one.[1]

There has been very little investigation of word-length effects on lexical activation, either from a modeling approach or an empirical one. In fact, a remarkably large portion of the word recognition literature has been entirely based on the shortest possible words: monosyllables. This is especially true for visual word recognition, but it is also the case for spoken word recognition. On the visual side, leading examples of monosyllable-only models include the dual route cascaded model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), Plaut, McClelland, Seidenberg, and Patterson's (1996) distributed model, Harm and Seidenberg's (1999) model, and the multiple read-out model (Grainger & Jacobs, 1996). For spoken word recognition, monosyllables have also received the most attention. The neighborhood activation model (Luce & Pisoni, 1998) is based solely on monosyllables. The behavior of the DCM has been extensively explored with monosyllables (Gaskell & Marslen-Wilson, 1997; but see Gaskell & Marslen-Wilson, 1999). In many instances, these and other models often focus on monosyllables because the empirical results that they are trying to simulate were obtained in experiments that were restricted to monosyllables (e.g., Goldinger, Luce, & Pisoni, 1989; Kessler & Treiman, 2002; Marslen-Wilson & Warren, 1994; Vitevitch & Luce, 1999) or because monosyllables were considered adequate to simulate the basic finding of interest. Note that if the dynamics of lexical activation for long words are quite different than those for short ones, then there is a gaping hole in our understanding of word recognition.

There has been at least one modeling investigation of the effect of word length. In a series of TRACE simulations, Frauenfelder and Peeters (1990) showed that peak (asymptotic) activation is positively associated with the number of phonemes in a word. The longer the word, the more highly it is activated. This result is consistent with the prediction outlined above. Thus, TRACE has the desirable property of using all acoustic–phonetic information to maximize successful recognition.[2] A model's sensitivity to word length also provides a means by which longer words can diverge from their shorter and more frequent monosyllabic competitors (e.g., *car* in *carpenter*) and, thus, be recognized.

There have been surprisingly few empirical examinations of the effect of word length on lexical activation. Samuel (1981, 1996) used the phonemic restoration paradigm, in which listeners have to discriminate words in which a phoneme has been replaced by noise from one in which noise was added to the phoneme. In this paradigm, discrimination should *decrease* if lexical activation increases with word length. That is, it should be increasingly difficult to distinguish between noise-added and noise-replaced stimuli. This is because as lexical activation grows, lexical influences on responding will increase, causing listeners to report hearing the phoneme as being present (i.e., restoring it) even when noise replaced the phoneme. The noise-added and noise-replaced versions of the words should become more indistinguishable as word length increases. Words varied in length from two to four syllables in Samuel's (1981, 1996) studies. Across multiple experiments, discrimination decreased as word length increased, suggesting that longer words do in fact generate greater lexical activation.

Although these results are consistent with the theoretical expectations, this issue was not the primary focus of the restoration studies, and as such the experiments did not provide an ideal test. For example, the shortest words, monosyllables, were not included, and other variables (e.g., word duration) were not controlled. The prediction of poorer performance in longer words could be attributed to other, less interesting factors (e.g., increasing masking effects). In addition, phonemic restoration is not as sensitive as other paradigms (Samuel, 1997), and it might therefore have underestimated effects of word length.

One other recent study has provided evidence that longer words generate more lexical activation than short ones, although again this issue was not the primary focus. Davis et al. (2002) investigated the potential problems for word recognition that occur when a shorter word is embedded in a longer word. They contrasted cases such as "The soldier saluted the flag with his cap tucked under his arm" and "The soldier saluted the flag with his captain looking on." Note that at the point that the listener hears "cap" (either as a stand-alone word or as part of "captain"), the same phonetic information has been provided. Davis et al. investigated whether there are nonetheless cues in the signal that could guide lexical selection in resolving the ambiguity, and they determined that there are (notably, differences in duration). However, also visible in Davis et al.'s data is an effect of word length. Across four cross-modal priming experiments, longer words ("captain") consistently produced much stronger priming effects than shorter words ("cap"). This result suggests that longer words inherently produce greater lexical activation than their shorter counterparts.

Although the studies of Samuel (1981, 1996) and Davis et al. (2002) are suggestive, there is a clear need for a more focused investigation that directly examines length effects on lexical activation. This was the goal of the current study. In place of phonemic restoration, we used the Ganong (1980) paradigm, in which listeners classify items from a test continuum as being one of two alternatives (e.g., /s/ vs. /ʃ/). Two word contexts (e.g., /mI/ and /wI/) are prepended to the fricatives to bias labeling of the perceptually ambiguous (middle) steps along the continuum toward the lexically consistent endpoint (e.g., *miss* or *wish*). The contextual bias in labeling is a measure of lexical influences on phoneme perception and is referred to as a *lexical shift*. We used the size of the shift (i.e., the extent of the bias) as a measure of the degree of lexical activation, with a larger shift indicating greater activation. A very useful advantage of the Ganong paradigm is that it is possible to partition responses on the basis of reaction time (RT) to examine changes in lexical activation over time (Fox, 1984; Pitt & Samuel, 1993). We capitalized on this analysis in comparing the dynamics of lexical activation of short and long words.

Our investigation of the dynamics of lexical activation is organized into three parts. In Part 1, we examine the accumulation of phonemic evidence for a word and lexical competition during recognition. The primary manipulation is word length, with the prediction that longer words will produce a stronger lexical shift

[1] Of course, a model can be designed to counteract such differences, by, for example, making inhibition strength a function of word length, as in TRACE.

[2] Frauenfelder and Peeters (1990) showed that all phonemes do not contribute equally to word-node activation in TRACE. Rather, each additional phoneme contributes a smaller amount than its predecessor. This is in part due to the sequentiality of processing in the model and the need to provide an upper bound on total activation.

than shorter words. As noted, this pattern would be expected both because there is greater bottom-up evidence available for the longer words and because there should be less lexical competition with them. We included a second, related manipulation in these experiments: Longer words either had relatively early lexical uniqueness points or relatively late ones. As Marslen-Wilson and Welsh (1978) and others have shown, the point at which a word diverges from all others in the language is a good predictor of performance on a number of tasks, all of which suggest that early divergence generally leads to faster lexical access due to a reduction of lexical competition. In this sense, we can consider the early-unique long words to be most different from the short words, whereas the late-unique long words should be somewhere in between; the latter have all of the extra bottom-up activation that comes with being a long word, but they have less of an advantage because of extended competition from lexical competitors.

Part 1 includes one additional manipulation. When words differ in number of syllables, they also typically differ in duration (e.g., monosyllabic words are usually shorter than trisyllabic words). In the above discussion, we focus entirely on the informational differences, not on durational ones. However, particularly if the goal is to examine the dynamics of activation, temporal properties are potentially quite important. For example, it seems quite plausible not only that longer words might have higher asymptotic activation levels than shorter ones (as shown in the Frauenfelder & Peeters, 1990, simulations) but that a high level of activation might be more sustained, because the supporting information is coming in over a longer period of time. To examine this issue, we include a pair of experiments in Part 1: In Experiment 1, durations of the short and long words were kept at their natural (and therefore differing) values. In Experiment 2, durational differences were eliminated while informational differences were maintained.

The results of Part 1 provide clear evidence regarding the effects of accumulating phonemic evidence and lexical competition. In Part 2, we use RT-based partitioning of the data from the first two experiments to clarify the changes in lexical activation over time for short versus long words.

In Part 3, we build on the results in Parts 1 and 2 to better understand how the experimental setup (e.g., instructions and stimuli) can alter lexical influences in phonetic categorization. Consider a situation in which the participant must indicate whether a stimulus ended in /s/ or in /ʃ/. If the stimulus is the word "miss," the participant could base the response (directly or indirectly) either on the developing sublexical representation of the /s/ itself or on the developing lexical representation that contains a final /s/ (Norris et al., 2000). One or the other of these representations may be better developed at a given point in time, and word length might alter listeners' reliance on each. Part 3 includes two experiments that used time-altered stimuli and a response-deadline procedure to measure the extent of lexical influences in phonetic categorization. Collectively, the results of the four experiments provide a window into the time course of the lexical activation process.

Part 1: Accumulation of Phonemic Evidence and Lexical Competition

Experiment 1

Lexical shifts generated by short (one-syllable) and long (three-syllable) words were compared, with the former expected to yield smaller shifts. Three-syllable words were chosen as the comparison length because Samuel and Pitt (2003) found that they generated large lexical shifts and because there were too few pairs of even longer words that met the necessary criteria for inclusion. Four pairs of words of each length were used to minimize the chance that some idiosyncratic property of a word might drive the results.

In connectionist models, the number of lexical competitors and the extent to which they overlap a word will affect the word's activation level. TRACE simulation studies performed by Frauenfelder and Peeters (1990, 1998) demonstrated the model's sensitivity to such properties. A positive relationship was found between a word's recognition point and its lexical uniqueness point. It took longer (more cycles) for TRACE to recognize a word as its uniqueness point moved further into the word, and this held true for words of differing lengths. Only after a word diverges from its competitors can its activation level begin to increase above those competitors and inhibit them enough to eventually be recognized. This takes longer for late-unique than for early-unique words.

In a recent behavioral study, Gaskell and Marslen-Wilson (2002) found clear evidence in support of this simulation result. Using a cross-modal repetition priming paradigm, they found significantly larger priming effects with early-unique than late-unique words, indicating that activation level is directly related to the point in a word at which it diverges from its competitors. In Experiment 1, we tested whether a word's length and its cohort size have opposite effects on lexical activation. Increases in the former should raise activation levels, whereas increases in the latter should lower them.

Method

Stimuli. In the Ganong (1980) paradigm, test items are generated in pairs. For example, if listeners are to categorize each test item as ending in either /s/ or /ʃ/, then a pair such as "miss" and "wish" can provide the necessary context. This pair is appropriate because for one item ("miss") there is a word only on the /s/ side ("mish" is not a word), whereas the reverse is true for the other member of the pair ("wish" is a word, and "wiss" is not).

Eight pairs of words constituted the contexts, with one member of each pair ending in /s/ and the other in /ʃ/. The words and their durations are listed in Appendix A. Both members of each pair had the same final vowel. The monosyllables were on average 3.25 phonemes in length. The trisyllables ranged from 6 to 10 phonemes in length and had primary stress on the second syllable. The early-unique trisyllables became unique at the 4th or 5th phoneme, and the late-unique words became unique at the 6th through 8th phonemes. Six of the context word pairs and the /s/–/ʃ/ continuum (described next) were from Samuel and Pitt (2003).

The word-final /s/–/ʃ/phonetic continuum was created by blending clear tokens of each fricative in various proportions (Pitt & McQueen, 1998). An eight-step test series was chosen that yielded good labeling of the endpoints and had a sizable ambiguous region. The most /s/-like endpoint had a blending ratio of .80 /s/ and .20 /ʃ/. The most /ʃ/-like endpoint had a corresponding ratio of .45/.55. Steps in between were equally spaced, with a .05 change in ratios (see Samuel & Pitt, 2003, for further details).

Each word context was spoken in isolation with a final /θ/ so that formant transitions from the final vowel would not bias perception toward one of the fricatives. /θ/ was then spliced off after the final pitch period of the vowel, and each step on the fricative continuum was appended to a copy of each context word to yield the 16 continua.

Procedure. Participants were tested in groups of up to 4 at a time in a sound-dampened booth. Stimuli were presented over headphones at a comfortable listening level, and listeners were instructed to categorize as quickly and as accurately as possible the final fricative of each word as either “s” or “sh” by pressing the corresponding button on a response box. A computer recorded which button was pressed and the time it took to respond, measured from fricative onset. Listeners had 2,500 ms to respond after word offset (placing an upper bound on RTs), and there was a 1,200-ms pause between trials.

The experimental design was completely within subject. Each step on each continuum was presented 20 times. With 16 continua and 8 steps on each, there were 2,560 trials. These were split equally over 2 days of testing. Stimuli were presented in blocks of 128 randomly ordered trials, with each step on each continuum occurring only once per block. A 20-trial practice session preceded the test session.

Participants. Twenty-eight students enrolled in introductory psychology at the Stony Brook University, State University of New York, served as listeners. All identified themselves as native English speakers, and none reported hearing difficulties.

Results and Discussion

We focused our analyses first on performance differences across the three word types (monosyllables, early-unique trisyllables, and late-unique trisyllables), collapsed over word pairs, to assess the effects of word length and uniqueness. For each listener, the percentage of /ʃ/responses at each step was calculated for the /ʃ/-biased and /s/-biased contexts for the three word types. RTs shorter than 300 ms were deemed too fast for accurate responding and excluded from this analysis (but see below). These data were then averaged over listeners and are plotted in the left graphs of Figure 1.

A lexical shift is clearly visible with both word lengths. The percentage of /ʃ/responses is greater in the /ʃ/-biased context than in the /s/-biased context across all of the continuum steps. For the monosyllables, the shift is most visible starting at Step 3, largest in the middle of the continuum (Steps 4–6), and even extends all the way to the /ʃ/ endpoint (Step 8). For the trisyllables, the shift extends across the entire continuum. Visual inspection of the sizes of the lexical shifts (i.e., the separation between the two /s/ and /ʃ/ labeling functions) shows that the shift is larger for the trisyllables than for the monosyllables. In addition, the shift is larger for the early- than for the late-unique trisyllables. Both outcomes support the predictions of the effects of word length and uniqueness on lexical activation.

Statistical analyses comparing shift sizes across conditions confirmed the preceding observations. As in past studies of this type (Pitt & Samuel, 1993; Samuel, 1986), shift size was measured by calculating the difference in labeling between the functions across the four middle steps of the continuum (3–6). Mean labeling of the middle four /s/-biased steps was first calculated for each listener. The same was then done for the corresponding /ʃ/-biased steps. The difference between them constituted the size of the lexical shift when averaged over listeners. Each listener’s means from the two biasing contexts served as the data for the statistical tests to assess the reliability of the lexical shift itself as well as for comparing shifts. Inferential tests were reliable at the .01 level unless otherwise noted.

Mean shifts in the three word-type conditions were submitted to a one-way analysis of variance (ANOVA). The effect was statistically reliable, $F(2, 54) = 13.58$, indicating lexical biases differed as a function of word length and uniqueness. We performed additional analyses on each word type and between word types to determine the source of these differences. The lexical shift for the monosyllables measured 26%, $F(1, 27) = 106.50$. This is a large Ganong effect, as would be expected given that the phonetic contrast was made with fricatives and the to-be-categorized phoneme occurred word-finally. The trisyllables, when collapsed over uniqueness point, yielded an even larger shift (40%), $F(1, 27) = 61.62$. The difference between the two shifts (14%) was reliable, $F(1, 27) = 15.70$. Comparison of the shifts across early- and late-unique words showed that activation strength was affected by a word’s uniqueness point. Both early-unique and late-unique words produced large lexical shifts (early: 46%, $F[1, 27] = 91.58$; late: 34%, $F[1, 27] = 33.64$), but the shift obtained with the early-unique words was reliably larger (12%), $F(1, 27) = 10.37$.

Four pairs of words of each length were used in the experiment to ensure that conclusions were minimally affected by any idiosyncrasies of a single pair of words. Although the results for individual word pairs were of course a bit noisier, the overall pattern was relatively consistent. Appendix B lists lexical shifts for each word pair for this and the subsequent experiments in this study. The monosyllabic pairs generally produced substantially smaller lexical shifts than the trisyllabic pairs. The late-unique pairs generally yielded smaller shifts than those for early-unique pairs, producing more variability for the trisyllables than for the monosyllables.

The results of Experiment 1 suggest that the degree of lexical activation is tied to the amount of perceptual information (i.e., number of phonetic cues or phonemes) contained in the spoken word and its point of divergence from other words. Long words generate more activation than short words, and the earlier a word diverges from its competitors, the more activation it will produce. The lexical uniqueness results converge nicely with those from Gaskell and Marslen-Wilson’s (2002) cross-modal priming study.

Experiment 2

In Experiment 1, we found evidence for much stronger lexical activation in longer words than in shorter ones, a result predicted on the basis of differences in the amount of bottom-up support for the two word types and on the basis of differences in lexical competition. However, there is a natural confound in a comparison of one-syllable to three-syllable words: In addition to the monosyllables having fewer phonemes than the trisyllables, the monosyllables were also shorter in duration by an average of 286 ms (668 vs. 954). Lexical activation could have been greater with the trisyllables because there was more time to process the utterances, not just because they contained more phonemes. The purpose of Experiment 2 was to neutralize the effects of word duration in contributing to the effects of length observed in Experiment 1.

Durational differences between the stimuli were eliminated by time-altering the monosyllables and trisyllables to be equivalent in duration. If the results of Experiment 1 were attributable to the additional time during which the signal supported the activation of longer words, then equating the durations should eliminate the difference in lexical shift size.

Method

The methodology was identical to that of Experiment 1, with the exception that the stimuli were time-altered to be approximately 796 ms. This value, which is halfway between the mean durations of the one- and the three-syllable stimuli, was chosen because it resulted in

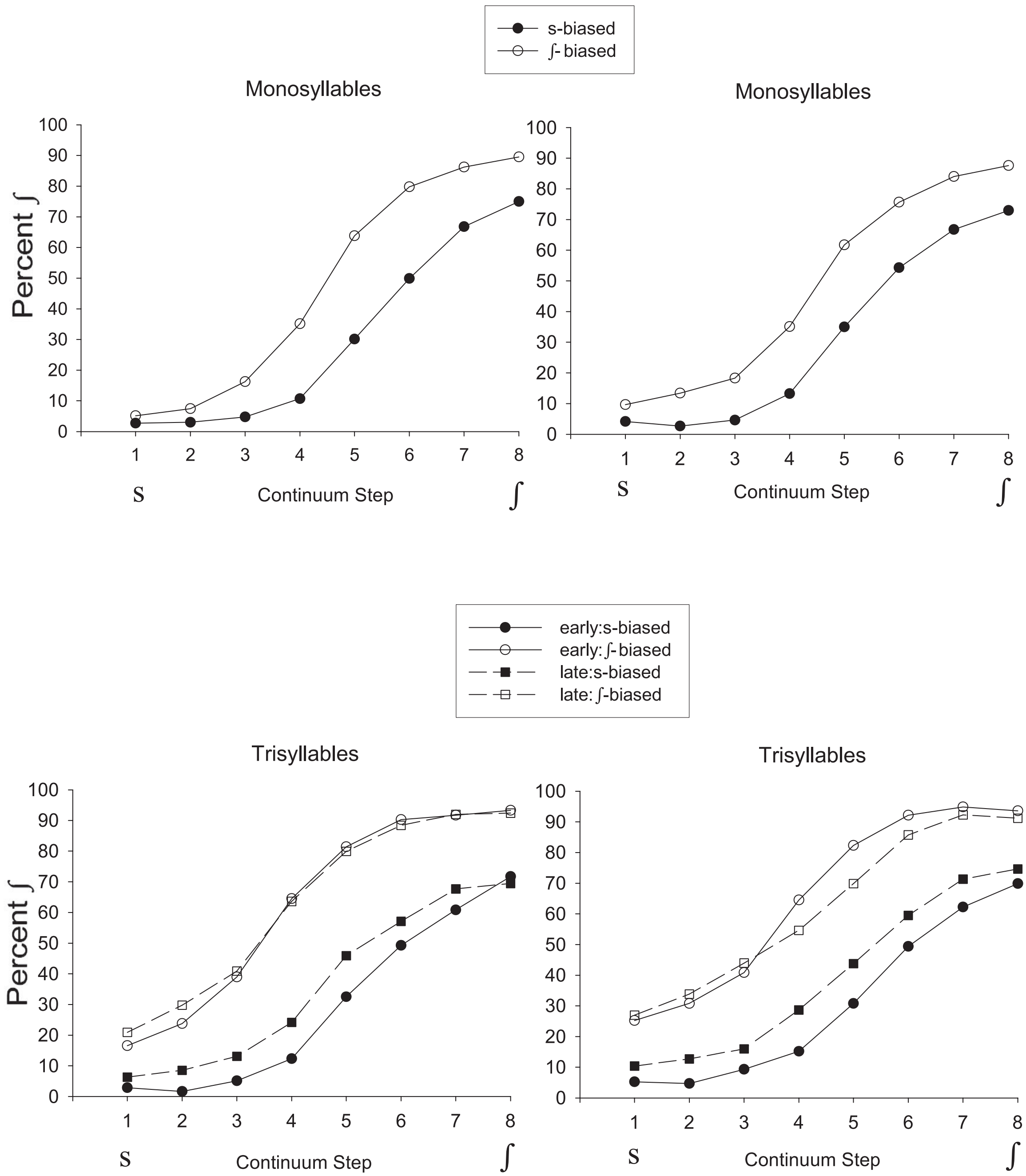

Figure 1. Graphs of the labeling functions in the /s/-biased and /ʃ/-biased contexts for the monosyllabic and trisyllabic stimuli of Experiments 1 (left graphs) and 2 (right graphs).

the stimuli being altered by the least amount and by about the same amount. The monosyllables were time-expanded an average of 19.4%, and the trisyllables were time-compressed an average of 16.5%. The time-alteration was performed with Picola Plus 2.0 (Uchida, 2002), which works by blending adjacent stretches of speech (e.g., pitch periods or voiceless stretches) across the to-be-altered stimulus, creating an entirely new token. For compression, a weighting function combines the two stretches to form new, shorter stretches. For expansion, the same function works in reverse to create longer stretches. The quality of the altered stimuli was very good. Twenty-four participants from the same population used in Experiment 1 were tested in Experiment 2; none had participated in the previous experiment.

Results and Discussion

The data were analyzed following the procedure described in Experiment 1. Mean labeling functions for the three word types are

Table 1
Mean Sizes (in Percentages) of Lexical Shifts for Monosyllables and Trisyllables Across Experiments 1–4

	Experiment				
Word type	1	2: Time-equated	3: Feedback + catch trials	3 (replication): Feedback only	4: Time-compressed
Monosyllables	26	21	39	35	12
Trisyllables	40	35	44	37	33

graphed in the right-hand side of Figure 1. The results closely resemble those on the left and suggest that durational differences between short and long words had negligible influences on lexical activation. A one-way ANOVA across the three word types was reliable, $F(2, 46) = 14.45$. The monosyllables yielded a substantial lexical shift (21%) similar in shape and size to that of Experiment 1 (26%), $F(1, 23) = 27.35$. Trisyllables yielded an even larger shift (35%), $F(1, 23) = 38.07$. The 14% difference between the two shifts was reliable, $F(1, 23) = 13.25$. Clearly visible in the lower graph is a strong effect of uniqueness, with the two late-unique functions lying inside the early-unique functions, most notably in the middle of the continuum. As in Experiment 1, both word types yielded reliable lexical shifts (early-unique: 44%, $F[1, 23] = 74.30$; late-unique: 27%, $F[1, 23] = 14.06$), and the early-unique shift was statistically larger than the late-unique shift, $F(1, 23) = 14.98$. The only difference between experiments is that shifts were slightly smaller in Experiment 2, but they were not reliably so. Table 1 contains mean shift sizes for the monosyllables and trisyllables for all experiments.

Inspection of the lexical shifts for each word pair in Appendix B further reinforces the strength of the replication in that they mirror what was obtained in the averaged data. Shifts obtained with the monosyllables were all smaller than those obtained with the trisyllables, even though the former were quite variable in size. Moreover, the shifts for the two late-unique pairs were smaller than those for the two early-unique pairs.

The results of Experiment 2 demonstrate that durational differences between monosyllables and trisyllables were not responsible for the 60% increase in shift size found with trisyllables. Rather, the amount of phonemic information conveyed by an utterance appears to be the cause. Longer words produce more lexical activation. This is not to say that variation in duration cannot alter lexical activation, only that it is not the cause of the differences in Experiments 1 and 2. As becomes clear in Part 3, the phonemic information available in the longer words (especially those with early uniqueness points) appears to be so strong that manipulation of processing time is relatively unimportant. For monosyllables, lexical activation is more malleable, with particular combinations of processing time and instructional set producing stronger or weaker lexical influences.

Part 2: Using RT-Based Partitioning to Probe the Cause of the Word-Length Effect

In the Ganong (1980) paradigm, RT is usually recorded along with the categorization response, which makes it possible to measure the size of the lexical shift at different points in time after onset of the fricative (Fox, 1984). In the current study, such analyses enable us to compare the time course of lexical activation for the short and long words and learn, for example, whether one is delayed relative to the other. We have already shown that longer words produce more activation than shorter words. If this larger "lexical pulse" also builds more quickly, then lexical influences on phoneme classification may be stronger earlier in time with longer words. Such a difference in activation dynamics would show up in the analyses as significantly larger lexical shifts at the shortest RTs. If there are corresponding differences in the dynamics of lexical activation as a function of lexical competition, then differences in the partitioned RT results should also be found as a function of uniqueness point.

As a first step in understanding why short words produce less lexical activation than long words, we measured shift sizes across the full range of RTs. The large number of observations per step (20/participant), combined with four word pairs of each length, provided enough data to construct detailed profiles of lexical activation over time, with bins spaced every 100 ms beginning at 300 ms (300–399, 400–499, etc.) and extending out to 800 ms.[3] Responses shorter than 300 ms were lumped together. Although they were included in graphs of the data to provide a complete picture of performance (see Figure 2), they were not included in the statistical analyses because such short RTs are suspect, possibly being guesses or other types of premature responses. A vertical line was inserted in the upper graphs between the <300-ms and 300-ms bins to remind readers of this.

Procedurally, an RT partition analysis differs from an overall analysis only in that data from a specific RT bin are used in measuring the lexical shift. For a participant's data to be included in calculating shift size, there had to be at least three observations per step. This requirement was imposed to prevent spuriously large or small shifts from distorting the results, which is most likely to occur in the extremes, where the fewest observations are often found. It was rare for listeners not to contribute a value in a bin used in the analyses. When they did (never more than 3% of the time), the mean of the bin was inserted into that cell.

Shown in the upper left graph of Figure 2 are the mean lexical shifts in each RT bin for the monosyllables and trisyllables of Experiments 1 and 2. Lexical shifts are largest in the shortest RT

[3] RT partition analyses are often performed by dividing each participant's data into RT bins (e.g., short, medium, long), which are defined relative to the participant's own RT distribution. Because we were interested in examining the time course of lexical effects across words of different durations, we felt it was most appropriate to take a different approach in this study and fix the boundaries of these RT bins relative to the onset of the final fricative. To do otherwise could distort or obscure effects of word length.

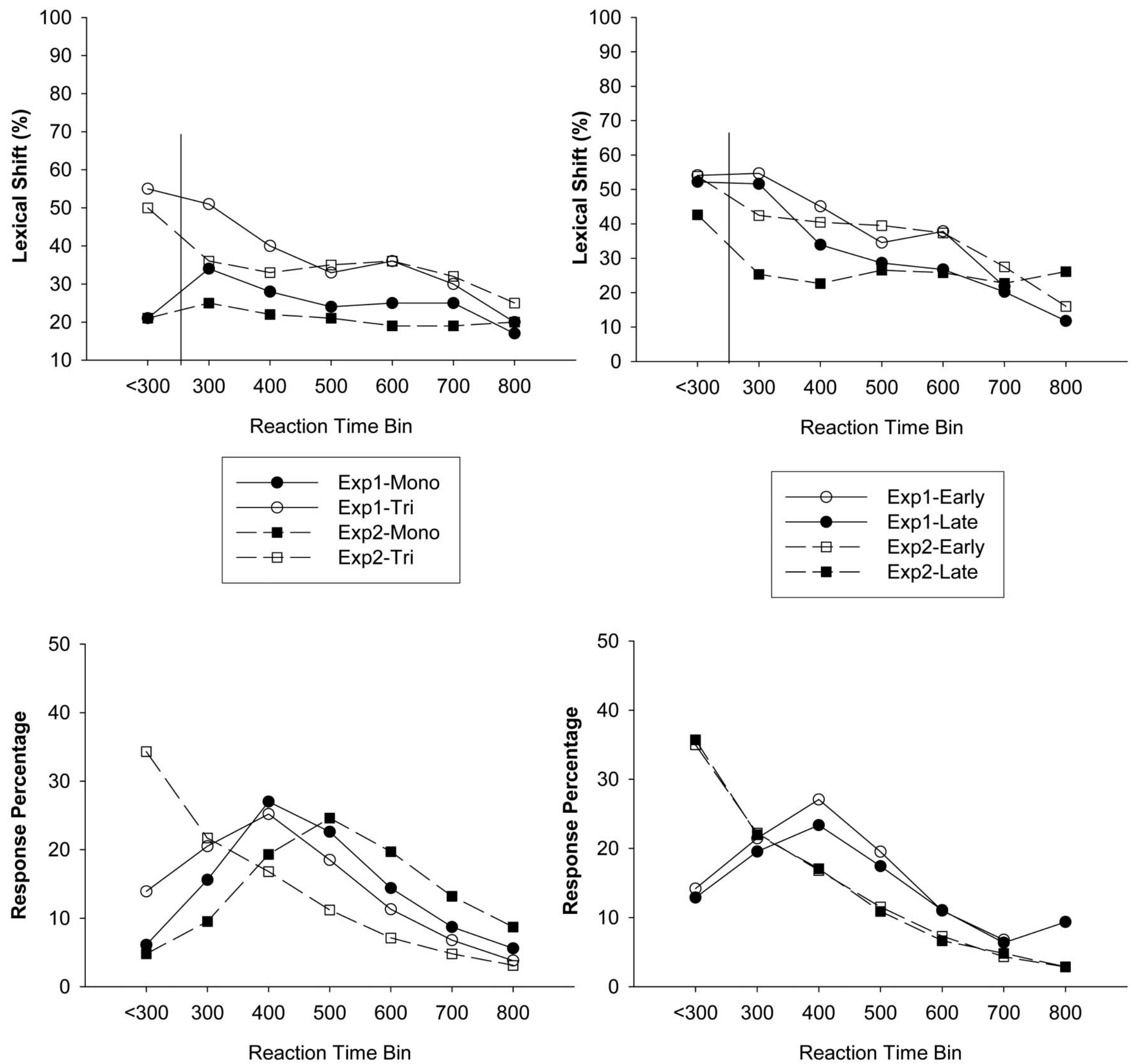

Figure 2. Graphs of the data from the reaction time (in milliseconds) partitioned analyses of Experiments 1 and 2. The vertical lines serve as a reminder that data from the fastest reaction time bin (<300 ms) were not included in the statistical analyses.

bins and decrease in size as RT lengthens. The change in shift size is most apparent for the trisyllables, but even the monosyllables exhibited this trend. This pattern is consistent with previous analyses of final-position Ganong shifts for monosyllables (McQueen, 1991; Pitt & Samuel, 1993).

A three-way ANOVA with experiment (1 vs. 2), word length (short vs. long), and RT bin (300–800) as factors confirm what is visible in the figure. There were main effects of RT bin, $F(5, 250) = 11.86$, and word length, $F(1, 50) = 16.59$, and an RT Bin × Experiment interaction, $F(5, 250) = 3.27$, with lexical shifts in the early bins being noticeably smaller in Experiment 2 than in Experiment 1 but more comparable in the later bins. The only other effect to reach significance was that of experiment, $F(1, 50) = 125.86$.

Further analyses probed the reliability of the trends across RT bin within each experiment. For the monosyllables of Experiment 1, shift size dropped from 34% in the 300-ms bin to half of this value (17%) in the 800-ms bin, $F(5, 135) = 3.71$. For the trisyllables, the shift in the 300-ms bin was extremely large (50%), but it decreased to almost the same value as that found with the monosyllables (20%) in the 800-ms bin. This 30% difference was also statistically robust, $F(5, 135) = 9.62$. The changes in shift size as a function of RT were substantially reduced by the neutralization of word-duration differences. In Experiment 2, lexical shifts diminished so little from the 300-ms to the 800-ms bin that the difference was not reliable for the monosyllables (5% [from 25% to 20%]; $F < 1$) or trisyllables (11% [from 36% to 25%]; $F[5, 115] = 1.44$, $p < .22$).

Although the convergence of the monosyllabic and trisyllabic functions across RT bins suggests that these two variables interacted, especially in Experiment 1, in neither experiment did they do so reliably. Instead, the outcomes of these two-way ANOVAs reflect what was reported above. There was a main effect of word length in Experiment 1, $F(1, 27) = 7.58$, and Experiment 2, F(1,

23) = 8.91, and the main effect of RT bin was reliable only in Experiment 1, $F(5, 135) = 13.09$.

A second RT partition analysis examined how these activation profiles across time differed as a function of uniqueness point. The trisyllabic data were broken down by early- and late-unique words in both experiments, and the results are shown in the upper right graph of Figure 2. A three-way ANOVA with experiment, uniqueness point, and RT bin as factors yielded reliable main effects for each (experiment, $F[1, 50] = 109.98$; uniqueness point, $F[1, 50] = 11.96$; RT bin, $F[4, 200] = 13.66$[4]) plus an RT Bin × Experiment interaction, $F(4, 200) = 5.90$, which is a result of the lexical shift in only the early bins being smaller in Experiment 2 than in Experiment 1. Note that this is the same outcome that is visible in the adjacent graph between experiments for both trisyllables and monosyllables.

Analyses of the data within each experiment showed that for Experiment 1, both the early-unique and late-unique words produced equivalent (32%) drops in shift size as RTs lengthened. In a two-way ANOVA with uniqueness point and RT bin as factors, the interaction was not reliable, but the main effect of RT bin was, $F(4, 108) = 15.30$. The main effect of uniqueness point approached significance, $F(1, 27) = 3.57$, $p < .07$. Separate one-way ANOVAs showed that the drop in lexical shift was reliable for both early-unique, $F(4, 108) = 9.89$, and late-unique words, $F(4, 108) = 8.80$. In Experiment 2, the two functions flattened out, but at the same time the effects of uniqueness point were amplified. The functions are wide apart in the fastest bin (17%) and gradually converge across bins. A two-way ANOVA with uniqueness point and RT bin as factors yielded a reliable interaction, $F(5, 115) = 4.83$, indicating that the change across bins was greater for the early-unique words than for the late-unique words.

The RT partition analyses illustrate the dynamics of lexical activation as a function of our manipulations of word length and uniqueness point. Lexical activation was greater for long than for short words across much of the RT range, and even greater still for early-unique than for late-unique words. Such effects dissipate over time, as the convergence of the functions indicates.

One difference that emerged across experiments was that the reduction in shift magnitude from the fastest to the slowest bin was greatly attenuated in Experiment 2. This can be seen by comparing the corresponding functions across the two experiments. For the trisyllables, the larger effect in Experiment 1 was confined to the 300- and 400-ms RT bins, with the two functions overlapping from the 500-ms bin onward even when broken down by uniqueness point. For the monosyllables, the reduction in Experiment 2 was essentially constant across RT partitions, although the difference between them was not reliable. These differences were unexpected, because the only difference between experiments was a slight change in stimulus duration. If duration-equating the stimuli was going to alter the time course of lexical activation, one would have expected the monosyllabic and trisyllabic functions to change differentially. Instead, they were similarly affected, with the size of the lexical shift becoming fairly stable over time.

To understand this finding better, we examined the speed of listeners' responses. The lower left graph in Figure 2 presents the resulting RT distributions, derived by plotting the percentage of participant responses that fell into each RT bin. For the monosyllables, time expansion had the expected effect of slowing responses; the RT distribution of Experiment 2 shifted to the right of that of Experiment 1. Time compression of the trisyllables had a much more extreme effect in the opposite direction, with the peak of the distribution shifting from the 400-ms bin to the <300-ms bin. Fully 34% of the time, listeners responded suspiciously quickly when they heard a time-compressed trisyllable. As the graph to the right shows, this was true whether the word's uniqueness point was early or late in the word.

A possible by-product of this frequent quick-response behavior is that listeners might have compensated by adopting a more conservative response criterion when responding more slowly, one that counteracted what were guesses some of the time. Specifically, listeners might have weighted the speech input more heavily than lexical information on these trials, which would explain why lexical shifts in the faster bins (300 and 400 ms) are smaller in Experiment 2 (note especially the difference between the late-unique functions in the upper right graph of Figure 2), and why shifts for the monosyllables of Experiment 2 were consistently smaller in all but the slowest bin.

Part 3: Using Response Deadlines to Probe the Dynamics of Lexical Activation

How lexical and perceptual information combine to affect phoneme categorization and detection has been researched extensively, with a number of studies identifying conditions that promote or inhibit lexical influences on responding, such as the monotony of the experiment and the inclusion of secondary tasks that bias lexical processing (Cutler, Mehler, Norris, & Segui, 1987; Cutler & Norris, 1979; Eimas, Hornstein, & Payton, 1990; McQueen, Norris, & Cutler, 1999; Pitt & Samuel, 1993). The results of the partitioned RT analyses suggest that word-length effects may be sensitive to such manipulations. Are there conditions in which short words could generate larger lexical shifts, shifts as large as those found with long words? Is lexical activation of long words similarly flexible? Experiments 3 and 4 were undertaken to address these questions. The results in hand suggest that the increase in lexical activation is greater for long words than for short ones, particularly if the long words have an early uniqueness point; the lexical pulse also appears to begin sooner, and it may last longer. The major new manipulation used in Part 3 is the imposition of response deadlines. Such deadlines make participants more aware of the timing of their responses and can oblige them to respond when the sublexical and lexical representations are in various states of activation.

Experiment 3

We began the next phase of this study by asking whether short and long words could generate more lexical activation than that found in the first two experiments. To test this, we biased lexical processing by implementing a few procedural changes that were intended to encourage listeners to process the words more fully. The monotony of the phonetic identification task can discourage

[4] There are 4 degrees of freedom in the numerator instead of 5 because no participant contributed enough data to the slowest (800-ms) RT bin for the early-unique words in Experiment 1 for a lexical shift to be calculated. Only five bins, 300–700 ms, were used in this analysis and others involving comparisons of uniqueness.

such processing and cause listeners to disregard the context and focus solely on the final fricative, which can have the effect of reducing lexical influences. We tried to counteract this tendency by requiring listeners to pay attention to their response speed throughout the experiment, keeping it within a specified range through the use of feedback. If this manipulation reduces the monotony of the task and thereby causes listeners to pay more attention to the word contexts, then both word types should generate larger lexical shifts than those in the preceding experiments.

To prevent overly fast and premature responses, we introduced catch trials into the design. On these trials, stimuli were presented in which the final fricative had been spliced off the end of the word. Their presence forced listeners to wait until they heard the end of the word before responding, because listeners were instructed not to respond on catch trials. In addition, the inclusion of catch trials might further bias lexical processing, because participants have to pay attention to each stimulus to know whether it has a final fricative.

Method

Stimuli were those of Experiment 2, with and without the final fricative, which when absent was spliced off after the final pitch period of the vowel. These catch trials occurred 20% of the time. With the inclusion of so many catch trials (320), we decided to split stimulus presentation in half and present two pairs of words of each length to separate groups of participants. Stimuli were presented in blocks of 80 randomized trials (64 test, 16 catch—two of each context), with each step on each continuum occurring only once in each of the 20 randomizations of the stimuli (1,600 total trials).

Procedure. The testing setup was similar to that of Experiment 1. Listeners sat in front of a computer monitor and used two keys on a computer keyboard (*z*, */*) to categorize the final fricative. Presentation (Version 0.76; Nerobehavioral Systems, Albany, CA [http://nbs.neuro-bs.com/]) software controlled stimulus delivery and response collection. Instructions from Experiment 1 were augmented with information about stimuli with the missing fricative (to which listeners were not to respond) and about feedback on response speed. After each block of 32 trials, one of three phrases appeared on the computer screen: *too fast*, *good pace*, or *faster*. Participants were instructed to adjust their response speed if the feedback was not *good pace*. After a few rounds of pilot testing, *good pace* was defined as a 400-ms window that began a few hundred milliseconds after fricative offset. The lower and upper boundaries of the window were 900 and 1,300 ms, respectively, measured from word onset.

Testing was completed in one session, with rest breaks after every 33% of the trials. There were 20 practice trials followed by a 50-trial warm-up block to allow RTs to stabilize prior to calculating the first mean RT. The first test block followed immediately.

Participants. Listeners were 20 undergraduates enrolled in introductory psychology at Ohio State University. None reported any hearing or language difficulties.

Results and Discussion

The overall labeling data are shown in the left panels of Figure 3. Monosyllables yielded a shift of 39%, $F(1, 19) = 31.70$, and trisyllables yielded a shift of 44%, $F(1, 19) = 46.65$. The 5% difference between them was not statistically reliable ($F < 1$). Compared with Experiment 2, which differed only in instruction and the absence of catch trials, shift size almost doubled for the monosyllables (from 21% to 39%), $F(1, 42) = 5.56$, $p < .02$. For the trisyllables, the increase was 9% (from 35% to 44%). Although relatively large, this difference was not statistically reliable either, $F(1, 42) = 1.12$, $p < .30$.[5]

These data show that lexical shifts for monosyllables are quite malleable; it is possible to design conditions that yield shifts similar to those for trisyllables. In comparison, the pattern for trisyllables is far more rigid, as though lexical activation proceeds much more automatically. Changes in instruction dramatically increased lexical activation of monosyllables but had much less of an effect on trisyllables.[6]

The apparent automaticity of lexical activation for longer words may not be quite as absolute as it seems. Although inclusion of catch trials had minimal effects on the size of the overall labeling shift, the partition analyses graphed on the right side of Figure 3 indicate that the catch trials did in fact alter the activation profile. The data from Experiment 2 are plotted for comparison to show the impact. The monosyllabic and trisyllabic functions of Experiment 3 lie on top of one other, again showing the flexibility in monosyllabic activation. Most striking, however, is the steepness of these functions. Lexical shifts start at a high of 53% in the 300-ms bin and drop steadily to a low of 17% in the 800-ms bin, a threefold decrease. The two-way ANOVA with word length and RT bin as factors yielded only a reliable effect of RT bin, $F(5, 95) = 13.69$.

Comparison of these data with those of Experiment 2 shows that the experimental conditions (feedback plus catch trials) had similar effects across word length but over *different* RT ranges during processing. For the monosyllables, the lexical-shift function went from being flat in Experiment 2 to having a sharp peak at the 300-ms bin in Experiment 3, as though the entire function pivoted clockwise, with the 800-ms bin serving as an anchor point. The boost in activation decreased steadily across bins until it disappeared at the final one. This Experiment × RT Bin interaction was reliable, $F(5, 210) = 4.18$. For the trisyllables, the pivot point was shifted earlier in time relative to the monosyllables, to the 500-ms bin. Because of this, the lexical shift was larger in Experiment 3 in the earlier bins (300 and 400 ms) but smaller in the later bins (600–800 ms). The interaction of these functions was also reliable, $F(5, 210) = 9.57$. In an omnibus ANOVA involving all three variables (word length, RT bin, and experiment), only the main effect of RT bin, $F(5, 210) = 15.59$, and the Experiment × RT Bin interaction, $F(5, 210) = 7.95$, reached significance.

These results suggest that although lexical activation of trisyllables is strong and relatively automatic, procedural changes can affect how listeners tap into the activated representations over time. When the experiment demands that listeners process the words fully, as in Experiment 3, lexical influences are very strong

[5] Because lexical uniqueness was not the focus of Experiment 3, analyses of early- and late-unique words were not performed. Nonetheless, the pattern of larger shifts for early-unique words was replicated. See Appendix B for overall shift sizes for each word pair.

[6] Similar results were obtained in a replication of Experiment 3, which differed only in that there were no catch trials. These data are included in Table 1, Figure 5, and Appendix B to demonstrate the robustness of the word-length differences. Further details about this experiment can be found on M.A.P.'s Web site (http://lpl.psy.ohio-state.edu).

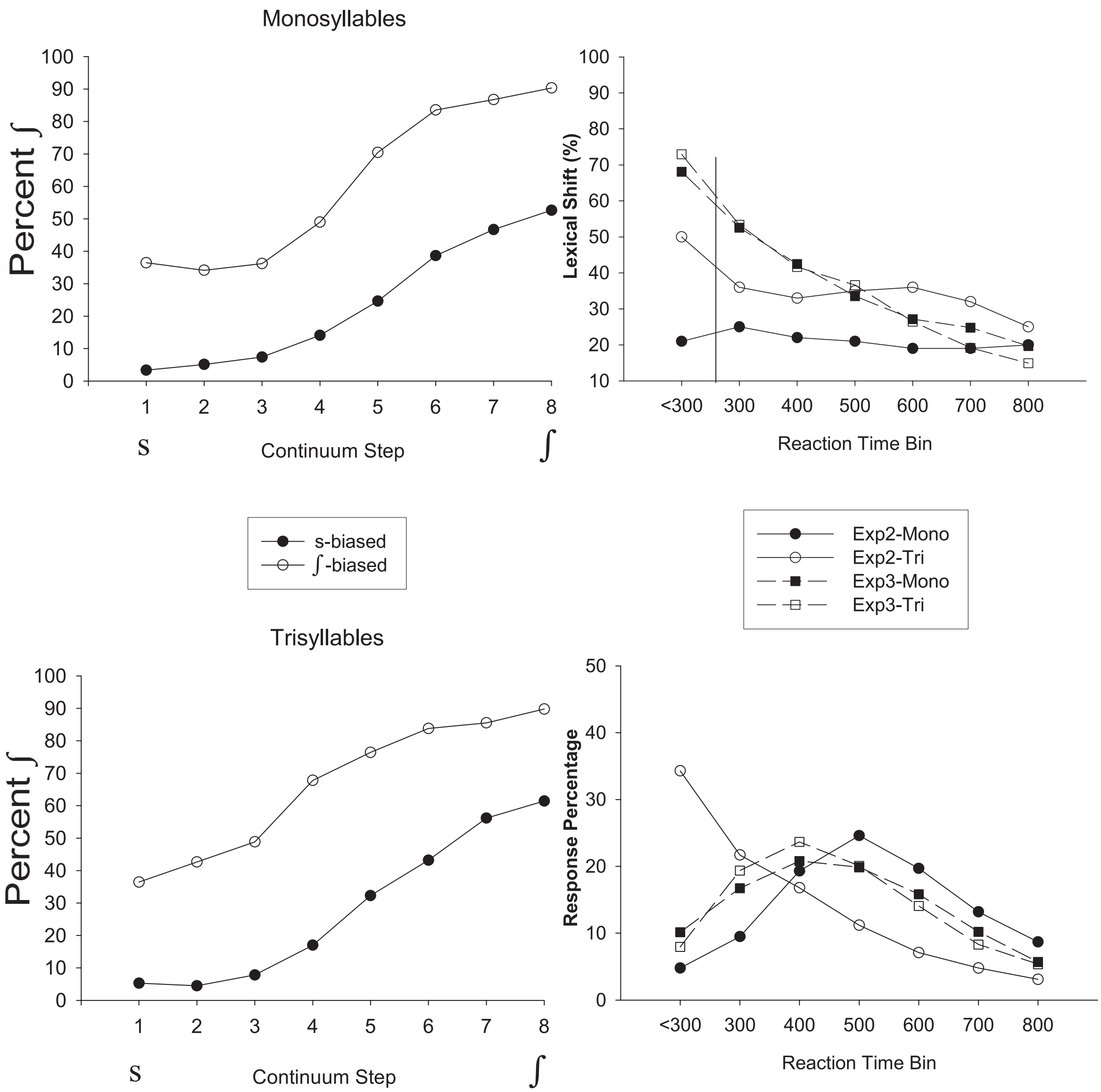

Figure 3. Overall labeling functions (left graphs) and reaction time (in milliseconds) partitioned analyses (right graphs) of the data from Experiment 3. Comparison data from Experiment 2 are included in the graphs of the partitioned results. The vertical line serves as a reminder that data from the fastest reaction time bin (<300 ms) were not included in the statistical analyses.

early on and then decay rapidly thereafter. Activation is a bit more evenly distributed when the experimental setup is less demanding of participants, which means that large lexical effects can still be found late in time. This pattern is consistent with the idea that the lexical pulse is sustained over a relatively long time period for longer words, allowing listeners to base responses on it as needed. For monosyllables, activation was redistributed as well, but it was also accompanied by an overall increase in listeners' reliance on lexical (vs. sublexical) information, which is why lexical shifts late in time in Experiment 3 never fell below those in Experiment 2. If lexical activation is relatively lower and briefer in monosyllables, listeners may base responses on lexical representations only when experimental conditions oblige them to do so.

Experiment 4

Probably the most salient finding across the preceding experiments is that lexical influences can change markedly when listeners process monosyllables but much less so when they process trisyllables. Experiment 4 was a more stringent test of this claim. Words of both lengths were time-compressed by 30%. With shorter words made even shorter than normal, listeners' tendency to respond before lexical representations are well activated should be reinforced, yielding even smaller lexical shifts than those found with the unaltered words of Experiment 1. The amount of compression was almost twice that applied to the trisyllables of Experiment 2 (16%). If lexical activation is minimally affected by durational variation (within reason), lexical shifts for the trisyllables should again be comparable to those in Experiment 1.

Method

The stimuli of Experiment 1 were time-compressed to 70% (greater amounts of compression began to reduce stimulus clarity) of their original duration, reducing the monosyllables by 200 ms (from 666 to 466 ms) and the trisyllables by 285 ms (from 954 to 669 ms), which made them equivalent in duration to the unaltered monosyllables. Otherwise, the experimental conditions and procedure were identical to those of Experiment 1, with the exception that word length was split between participants. Testing was completed in one session consisting of 1,280 trials (8 continua × 8 steps × 20 presentations per step). Rest breaks were provided after every 33% of trials. Participants were 52 new students from the same population that was sampled for Experiment 3. Half responded to the monosyllables, and half responded to the trisyllables.

Results and Discussion

The labeling and RT partition data are graphed in Figure 4. The compressed monosyllables generated a mean lexical shift of 12%, $F(1, 25) = 21.57$, and the compressed trisyllables yielded a lexical shift of 33%, $F(1, 25) = 32.79$. A one-way ANOVA on word length showed that the lexical shift was reliably larger for trisyllables, $F(1, 50) = 10.79$. Compared with the 25% shift for the unaltered monosyllables of Experiment 1, the size of the lexical effect was halved in Experiment 3, a statistically reliable difference, $F(1, 52) = 6.36$, $p < .02$. Although the shift for the trisyllables was 7% smaller than that in Experiment 1 (40%), the drop was not statistically different ($F < 1$).

Inspection of Appendix B shows that variability in shift size within Experiment 4 was negligible for the monosyllables but high for the trisyllables. The two late-unique word pairs, *distinguish–*

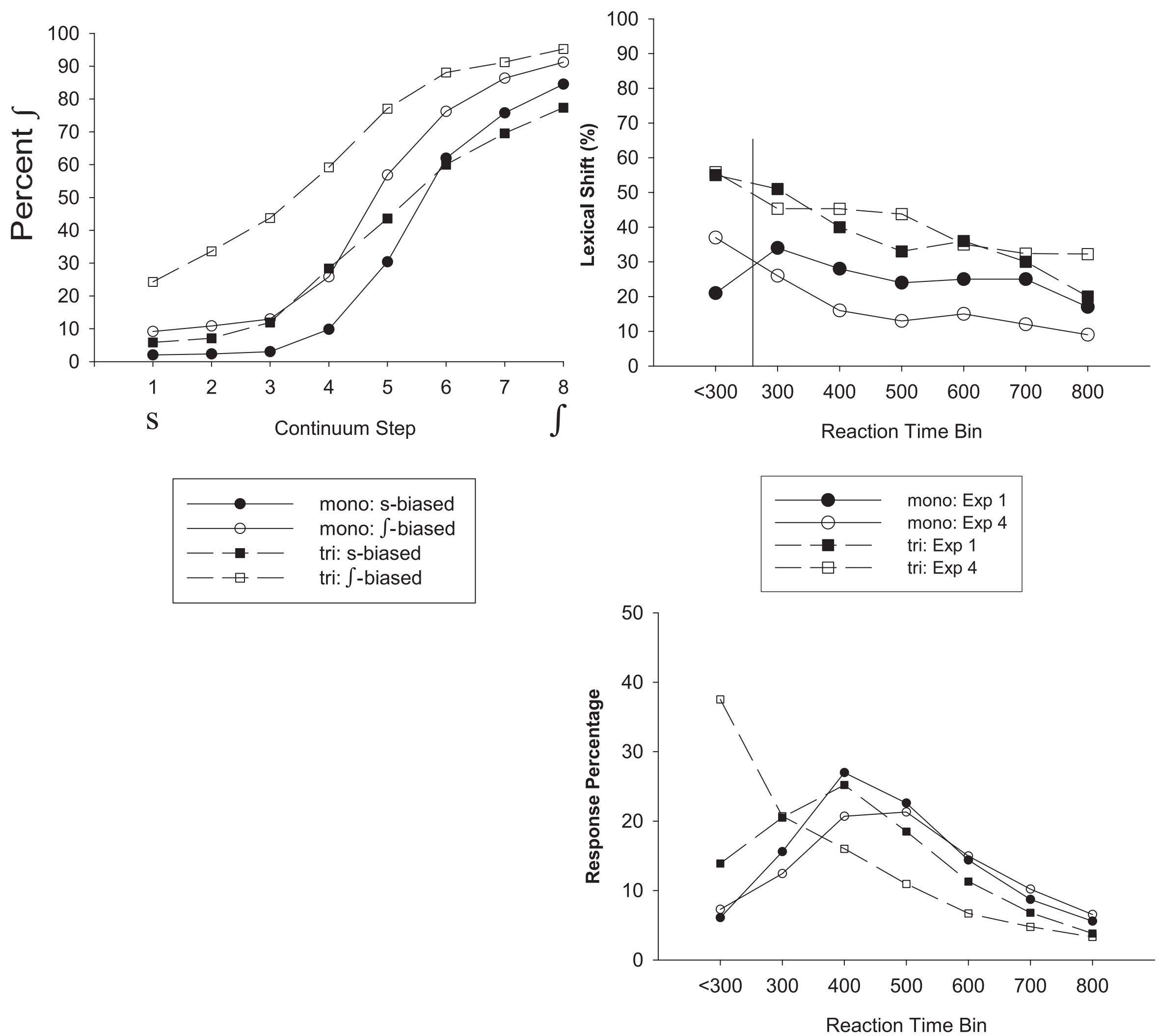

Figure 4. Overall labeling functions (left graph) and reaction time (in milliseconds) partitioned analyses (right graphs) of the data from Experiment 4. Comparison data from Experiment 1 are also included. The vertical line serves as a reminder that data from the fastest reaction time bin (<300 ms) were not included in the statistical analyses.

consensus and *extinguish–contagious*, yielded shifts small enough to give the impression that time compression significantly reduced lexical activation. However, such small shifts were not uniform across participants but restricted to a small group (6) who generated large reverse lexical shifts (−30% or more), virtually only with these two continua. Removal of their data causes shift size to jump to 35% and 27%, respectively, for these two word pairs while changing little for the other two word pairs.[7]

The RT partition data also show that monosyllable processing is much more sensitive to stimulus duration than is trisyllable processing. Time compression of the monosyllables produced a fairly constant reduction in shift magnitude, as shown by the fact that the Experiment 1 and 4 functions parallel each other across bins. At the 300-ms bin, the compressed words produced a shift of 26%, which dropped to 9% at the 800-ms bin, a difference of 17%, $F(5, 125) = 2.43$, $p < .04$. The unaltered monosyllables of Experiment 1 yielded a drop of 17% as well. Cross-experiment comparisons produced a main effect of experiment, $F(1, 52) = 9.56$, and a main effect of RT bin, $F(5, 260) = 5.7$, with no interaction between the two.

No such comparable effect of duration was found with the compressed trisyllables. Instead, the functions from the two experiments are intertwined across RT bins. The Experiment 4 function shows a modest drop of 13% across partitions, $F(5, 125) = 2.96$, $p < .02$. Comparison with the drop in Experiment 1 yielded only a main effect of RT bin, $F(5, 260) = 9.98$. Neither the main effect of experiment nor the Experiment × RT Bin interaction reached significance.

An omnibus ANOVA with experiment, length, and RT bin as factors yielded only three reliable results: In addition to there being main effects of RT bin, $F(5, 520) = 14.22$, and word length, $F(1, 104) = 25.13$, there was a Word Length × Experiment interaction, $F(1, 104) = 5.03$, $p < .03$, which together further highlight the differential effects of time compression on processing monosyllabic and trisyllabic words. Across RT bins, compression had little effect on the processing of trisyllables but a relatively constant effect on processing of monosyllables.

The data in the RT distributions in the lower graph of Figure 4 further reinforce processing differences between long and short words. The two monosyllabic distributions are on top of one another. This is surprising because one might have expected that compression by such a large amount (30%) would have time-shifted the Experiment 4 distribution rightward, producing a delay in peak activation. That it did not suggests that listeners chose to respond on the basis of the sublexical information that was available at the time, before the lexical representations were sufficiently activated to contribute much to task performance. The trisyllabic data replicate what was found in Experiment 2 (see the lower left graph of Figure 2). The RT distribution is highly skewed, with listeners responding under 300 ms a majority of the time—just what would be expected from listeners who relied heavily on lexical representations to respond.

The results of Experiment 4 further accentuate the differences in processing short and long words. Time compression of monosyllables, which reduced lexical-shift size, had the opposite effect of the deadline procedure in Experiment 3, which enlarged the shift. Trisyllable processing, however, was minimally affected by both manipulations. The two manipulations also differed in the pattern of the changes they induced when listeners processed monosyllables. Time compression had a fixed effect on processing, with lexical-shift size changing by an almost constant amount over time relative to the unaltered tokens. In contrast, the effects of instruction significantly heightened lexical activation only early in processing. These different influences may be attributable to the nature of each manipulation. Time compression affects an intrinsic property of the stimulus, whereas variation in instruction is extrinsic to the stimulus. Regardless, lexical activation for short words is extremely fragile, and its appearance is dependent on the details of the experimental procedures.

The results obtained with monosyllables contrast sharply with the generally robust lexical effects for longer words that were found across a range of experimental conditions. As noted in Experiment 2, there is some evidence that the immutability of lexical activation for longer words is not absolute—those with late uniqueness points showed some sensitivity to duration. With the moderate reduction in duration in Experiment 2, the 27% lexical shift for such words was somewhat lower than the 34% shift found for the uncompressed versions of those words in Experiment 1; for early-unique words, in contrast, there was a negligible change across experiments (46% in Experiment 1 and 44% in Experiment 2). Under the more extreme compression used in the current experiment, it was again the late-unique words that showed some loss of lexical activation, although again the loss was not huge (and was largely driven by a small number of participants). Collectively, these results suggest that the extremely constraining nature of the early-unique words made processing time largely irrelevant, with late-unique long words showing a similar but not as extreme indifference to variation in duration.

General Discussion

For an understanding of how spoken words are recognized, the variables that affect how words map onto their representations in memory must be delineated. The results of the current study show that word length is an important variable: Long words can generate more lexical activation than short words. In Experiments 1 and 2, trisyllables produced lexical shifts that were 60% larger than those produced by monosyllables. The two follow-up experiments revealed that activation strength and its time course are not fixed but, instead, are subject to the task demands of the experiment, with word length playing an important moderating role. Instructions to monitor response speed were intended to counteract experiment monotony and to bias lexical processing. This change in experimental setup was far more effective with monosyllables than with trisyllables, as can be seen by scanning across the first four columns of Table 1. Lexical-shift size increased with the manipulation when participants responded to monosyllables but showed much smaller changes when they responded to trisyllables. The reverse effect, reducing the size of the lexical shift, was also achieved with the monosyllables via time compression. Differences in shift size as a function of these manipulations were always

[7] Additional analyses of the data provided some evidence to suggest that the reversals for these 6 participants were attributable to a significant change in response speed (either a speedup or a slowdown) in the latter half of the experiment. The change was largely confined to the ambiguous region of the continuum, suggesting a change in response strategy when these participants heard ambiguous fricatives.

statistically reliable for the monosyllables, but they were never so for the trisyllables.

A clearer understanding of the word-length differences found in this study can be seen in Figure 5, which contains graphs of the lexical shifts across experiments, plotted separately for monosyllables and trisyllables. It is readily apparent that lexical-shift size differs much more widely for monosyllables than for trisyllables. When lexical shifts for monosyllables increase as a result of experimental changes, the increase is confined to the early bins (300–500 ms), with the functions shifting upward to resemble the

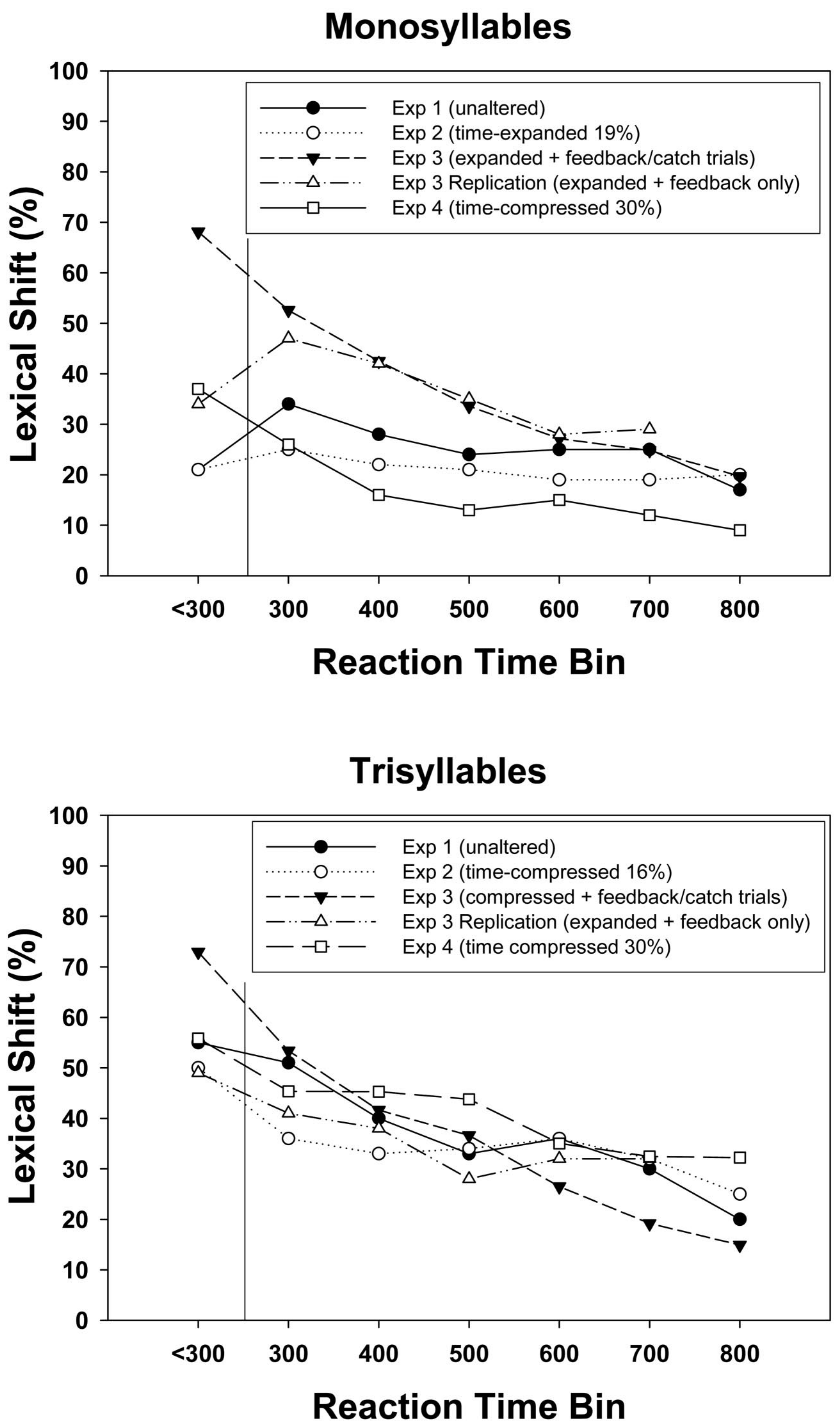

Figure 5. Mean lexical shift size as a function of reaction time (in milliseconds) bin across all experiments. Vertical lines indicate that the <300-ms reaction time bins—graphed here to provide a complete picture of performance—were not included in the statistical analyses.

continuously downward-sloping function of the trisyllables in the lower graph. This increment was short-lived, not long-lasting, suggesting that it was based on a relatively brief period of increased lexical activation.

The trisyllabic functions in the lower graph are bunched together. To the extent that instruction and catch trials affected activation, the effect was qualitatively different than that found with the monosyllables. For longer words, large lexical effects can be long-lasting, though they appear to have peaked early on, leading listeners to report the final phoneme in a lexically consistent manner.

Our RT partitioned analyses in Parts 2 and 3 provide details of how lexical and sublexical activation lead to phoneme decision making. For trisyllables, lexical activation is so strong by the time the word-final fricative is heard that lexical influences dominate and are virtually immutable. The weighting of these two sources of information is far more variable when listeners responded to monosyllables. Such short words do not automatically achieve the high level of activation by fricative onset that is necessary to overwhelm perceptual input and yield a large, consistent lexical bias. Rather, there is flexibility in the use of lexical activation, which can be boosted by instructing listeners to process the word more fully.

Lexical competition also contributes to the word-length effect. In Experiments 1 and 2, we found that lexical uniqueness affected activation level, with early-unique words generating larger lexical shifts than those found with late-unique words (a pattern also seen in the later experiments). The even smaller lexical shifts found with monosyllables could represent an exaggerated uniqueness effect, because three- and four-phoneme words rarely become unique before the final phoneme (this lack of uniqueness by word end was the case for all eight monosyllables used in this study). This problem boils down to one of redundancy. Activation of monosyllables is weak relative to that of trisyllables because there is minimal divergence from competitors. This creates ambiguity for the recognition system, slowing processing.

To account for these results, a model must be sensitive to word length and lexical uniqueness, and it must have some means of biasing (sub)lexical processing. Localist connectionist models, such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994), are well equipped to do so. Because words are represented as individual nodes in a network, activation level can be a direct function of word length. Nodes that match a longer stretch of speech will be activated more highly than those that match a smaller stretch (Frauenfelder & Peeters, 1990). Inhibitory connections between lexical nodes will produce effects of uniqueness (Frauenfelder & Peeters, 1998). One means of simulating the effects of instruction is with the use of a gain-control mechanism that modulates lexical influences on phoneme processing. A parameter has been recently added to TRACE that produces this behavior in the model (Mirman, McClelland, & Holt, 2004). The parameter dampens activation of lexical nodes, effectively attenuating the degree to which a node can be activated. One consequence of reduced lexical activation is that lexical feedback (i.e., word-to-phoneme excitation) will also be reduced, leading to weaker lexical biases in phoneme processing. This mechanism may well be adequate to capture the results found in the present study if the dampening effect on lexical activation can be conditioned by word length, being more variable for short than for long words.

It is less clear whether a distributed connectionist model like DCM (Gaskell & Marslen-Wilson, 1997) can accommodate the present findings as easily. The DCM would have no trouble producing the lexical-uniqueness effect. Indeed, in an informative investigation that explored the advantages of distributed representations for spoken word processing, Gaskell and Marslen-Wilson (1999) showed that this is one of the architecture's strengths. Recognition in DCM is driven largely by lexical divergence (i.e., uniqueness). The smaller lexical shifts found with monosyllables could be a result of late divergence as well, as suggested above.

Such a parsimonious explanation has considerable appeal, but divergence alone is probably not sufficient, at least as currently implemented in DCM. Divergence will not always occur, especially for short words, which makes it difficult to model length differences. A word's representation is shared by other phonetically overlapping words in the model's space of hidden units. Each word can be thought of as a path through this space, with short words (e.g., *car*) traversing a subset of the space of longer words in which they are embedded (e.g., *carpenter*). When the input to the model is *car*, all *car*-initial words (*carpet*, *carbon*, etc.) will match the input equally well, not just *car*. Even when short words do become lexically unique, there is nothing to differentiate the quality of this match from that of a longer word. A longer word will generate more lexical activity, but it is not specific to the word, nor will it lead to a stronger match (i.e., better fit). Both will match their respective signals equally well. Modifications to the model that overcome these limitations should do a more complete job of capturing the word-length effect. A word's semantic representation, which is stored in the model, might assist in this regard. As for the task effects, an additional mechanism would probably need to be introduced as well.

However the current results are modeled, this study clearly establishes the existence of word-length effects and delineates some of the dynamics of activation. As we noted in the introduction, most modeling of word recognition to date has been based on strictly monosyllabic lexicons. Our results and analyses suggest that a fuller understanding of lexical processing must by grounded in studies using words of varying lengths.

References

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.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have special status in auditory word recognition? *Journal of Memory and Language, 32,* 193–210.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19,* 141–177.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Erlbaum.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 218–244.

Eimas, P. D., Hornstein, S. M., & Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language, 29,* 160–180.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 526–540.

Frauenfelder, U. H., & Peeters, G. (1990). Lexical segmentation in TRACE: An exercise in simulation. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 50–86). Cambridge, MA: MIT Press.

Frauenfelder, U. H., & Peeters, G. (1998). Simulating the time course of spoken word recognition: An analysis of lexical competition in TRACE. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 101–146). Mawah, NJ: Erlbaum.

Ganong, W. F. (1980). Phonetic categorization in auditory perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science, 23,* 439–462.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology, 45,* 220–266.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28,* 501–518.

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

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading, and dyslexia: Insights from connectionist models. *Psychological Review, 106,* 491–528.

Kessler, B., & Treiman, R. (2002). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language, 24,* 592–617.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1–36.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101,* 653–675.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29–63.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 433–443.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1363–1389.

Mirman, D., McClelland, J., & Holt, L. (2004, August). *Attentional modulation of lexical effects in an interactive model of speech perception.* Paper presented at the 26th Annual Meeting of the Cognitive Science Society, Chicago.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences 23,* 299–370.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39,* 347–370.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 699–725.

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.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474–494.

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology, 18,* 452–499.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General, 125,* 28–51.

Samuel, A. G. (1997). Phoneme restoration. In F. Grosjean & U. H. Frauenfelder (Eds.), *A guide to spoken word recognition paradigms* (pp. 647–654). Hove, England: Psychology Press.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language, 48,* 416–434.

Uchida, T. (2002). Effects of the speech rate on speakers' personality-trait impressions. In A. Alippi (Ed.), *Proceedings of the 17th International Congress on Acoustics: VIII. Music, psychoacoustics, speech* (pp. 286–287). Rome: International Commission for Acoustics.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374–408.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 98–108.

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics, 41,* 262–275.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]–[š] boundary. *Journal of the Acoustical Society of America, 69,* 275–282.

Appendix A

Context Words Used in Experiment 1 and Their Durations (in Milliseconds)

Monosyllables		Trisyllables			
		Early-unique		Late-unique	
Word	Duration	Word	Duration	Word	Duration
fish	692	abolish	891	distinguish	944
kiss	636	arthritis	933	consensus	967
wish	664	establish	947	extinguish	955
miss	637	malpractice	1,006	contagious	989
dash	660				
pass	692				
fresh	724				
press	637				

Note. Word pairs are listed together, with the /ʃ/-biased (e.g., *fish*) word immediately before the /s/-biased word (e.g., *kiss*).

Appendix B

Mean Size (in Percentage) of Lexical Shift for Each Word Pair Used in Each Experiment

Word pair	Experiment				
	1	2	3	3 (replication)	4
fish–kiss	25*	28*	35*	34*	11*
wish–miss	32*	21*	69*	47*	15*
dash–pass	25*	19*	12*	34*	11*
fresh–press	23*	14*	48*	26*	11*
abolish–arthritis	51*	52*	44*	52*	58*
establish–malpractice	44*	39*	50*	39*	39*
distinguish–consensus	45*	31*	45*	17	20*
extinguish–contagious	29*	31*	23*	42*	15*

* $p < .05$.

Received January 31, 2005
Revision received December 4, 2005
Accepted December 7, 2005 ■