Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 52 (2005) 171–192
Journal of Memory and Language
www.elsevier.com/locate/jml

Does neighborhood density influence repetition latency for nonwords? Separating the effects of density and duration

John Lipinski, Prahlad Gupta*

Department of Psychology, University of Iowa, Iowa City, IA 52242, USA

Received 8 December 2003; revision received 27 October 2004

Abstract

Twelve experiments examined the effect of neighborhood density on repetition latency for nonwords. Previous reports have indicated that nonwords from high density neighborhoods are repeated with shorter latency than nonwords from low density neighborhoods (e.g., Vitevitch & Luce, 1998). Experiment 1 replicated these previously reported results; however, further analysis indicated an interaction of neighborhood density and stimulus duration in determining nonword repetition latency. Experiment 2 employed stimuli with reduced durational differences, finding that there was no effect of neighborhood density on repetition latency when stimulus duration was statistically controlled. Experiments 3 and 4 replicated these results with an alternative presentation regimen. Experiments 5–12 repeated these investigations with different stimulus sets, and obtained consistent effects of stimulus duration on repetition latency, and either no effect of neighborhood density or a latency advantage for low density rather than high density nonwords. The theoretical implications of these results for models of lexical processing are discussed.

Keywords: Neighborhood density; Nonwords; Repetition latency; Lexical processing; Stimulus duration

When human language users are exposed to a word form (i.e., a word or phonotactically legal nonword) of their language, in what way does their existing knowledge of the words of that language influence processing of the word form? This question is of obvious relevance to understanding the processing of known words. The question of how existing knowledge of the words of the language influences processing of *nonwords* is relevant to understanding the learning of new words, because every novel word that will eventually be learned is in effect a nonword to a learner, on first exposure. Thus, the question of how prior knowledge of words of a language impacts processing of other words and of novel words is of considerable importance to the study of lexical processing and lexical learning.

The study of lexical neighborhoods is one approach to addressing this question. The lexical neighborhood of a given word form may be defined as the set of known words that lie within a specified distance of the word form, on some metric of similarity. Phonemic overlap has been the most commonly used metric of similarity, and a one-phoneme difference the most commonly used distance criterion, leading to the lexical neighborhood of a given word form typically being defined as the set of words that can be transformed to that word form through the addition, deletion, or substitution of a single

* Corresponding author. Fax: +1 319 335 0191.
E-mail address: prahlad-gupta@uiowa.edu (P. Gupta).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2004.10.004

phoneme. By this definition, the neighborhood of "big" would include the words "pig," "dig," "bin," and "bug," but not "ban;" and the neighborhood of the nonword "tig" would include "pig," "dig," "big," "tin," and "tug," but not "tan." Following seminal work by Luce and colleagues (Luce, Pisoni, & Goldinger, 1990; Luce & Pisoni, 1998) in establishing the notion of lexical neighborhood, investigation of how various properties of neighborhoods affect the processing of word forms has become a major area of inquiry in the psychology of language.

Neighborhood density, the number of items in the neighborhood, and neighborhood frequency, the mean frequency of the items in the neighborhood, are two properties that have been especially studied. Luce and Pisoni (1998), for example, established that higher neighborhood density led to slowed decision times in an auditory lexical decision task, for both words and nonwords. However, in a simple immediate repetition task, higher neighborhood density appears to have opposite effects on the repetition of auditorily presented words versus nonwords, leading to longer repetition latencies for words (as in lexical decision), but shorter repetition latencies for nonwords (Luce & Pisoni, 1998; Vitevitch & Luce, 1998).

If higher neighborhood density has a facilitative effect on repetition latency for nonwords, this has potentially important implications for the learning of new words, suggesting that it may be easier to repeat, and hence perhaps eventually learn, new words that are from high density neighborhoods rather than low density neighborhoods. In addition to the importance of such a finding for theories of lexical learning, it would have potential practical implications for second language vocabulary learning as well as for programs of rehabilitation/remediation in cases of developmental or acquired language impairment.

The present work was therefore originally motivated by the goal of exploring the effects of neighborhood density in a learning paradigm: we wished to examine the effect of neighborhood density on the learning of nonwords, following repeated exposures to them. As we pursued this goal, however, we had great difficulty in replicating the basic phenomenon described above: a facilitatory effect of high neighborhood density on repetition latency for nonwords following a single exposure (Vitevitch & Luce, 1998). This led us to look carefully for differences between how neighborhood density was manipulated in our stimuli and in the stimuli used in previous studies. Eventually, we were led to examine a wide variety of circumstances under which higher neighborhood density does and does not lead to faster latencies in nonword repetition. The present article describes a series of studies documenting these investigations.

In addition to this empirical motivation for examining the effect of neighborhood density on nonword repetition latency, there is also theoretical motivation for such an investigation. Vitevitch and Luce (1999) proposed an elegant theoretical framework for understanding the processing of words and nonwords, adapted from the ART framework of Grossberg and colleagues (Grossberg, 1986; Grossberg, Boardman, & Cohen, 1997; Grossberg & Stone, 1986). According to the framework outlined by Vitevitch and Luce (1999), auditory presentation of a word form leads to the activation in the lexical system of "list chunks" of various lengths, corresponding to various substrings of the presented stimulus. For example, presentation of the real word "cap" would lead to activation of the list chunks /*kæp*/, which is a list chunk at a lexical level, as well as activation of /*kæ*/, /*æp*/, /*k*/, /*æ*/, and /*p*/, which are list chunks of various unit sizes at a sublexical level. Presentation of the nonword "yush" would lead to the activation of the sublexical level list chunks /*jʌ*/, /*ʌʃ*/, /*j*/, /*ʌ*/, and /*ʃ*/, but no lexical level list chunk exists or would be created or activated that corresponds to the entire nonword /*jʌʃ*/.

One important aspect of this framework is the notion of "masking," whereby longer list chunks mask or inhibit list chunks that correspond to their substrings. Thus, the longest list chunk that is consistent with the input will inhibit the various shorter list chunks that are also activated by the input. A second important aspect of the framework is that list chunks at a particular level compete with each other for activation. Thus for example, the /*kæp*/ list chunk would compete with other activated lexical level list chunks such as /*kæt*/, /*hæt*/, /*h§d*/, etc. This is also true for sublexical list chunks of various unit sizes, so that, for example, the sublexical list chunk /*æp*/ would compete with the sublexical list chunk /*æk*/, both of which are (partially) consistent with and hence activated by the input. A third important aspect of the framework is that the activations of chunks of a particular unit size are a function of their relative frequencies of occurrence. For list chunks such as /*æp*/ and /*æk*/ at the sublexical level, this means that activations are a function of the phonotactic probabilities of the list chunks.

These properties together provided a means of distinguishing the processing of words and nonwords, and thus explained the reported opposite effects of neighborhood density on repetition latencies for words and nonwords. According to this account (Vitevitch & Luce, 1999), higher neighborhood density leads to slower repetition latencies for real words because, when activated, the lexical level list chunks representing words from higher density neighborhoods have to compete with many other lexical level list chunks (representing their neighbors), that are also activated. In contrast, when lexical level list chunks representing words from lower density neighborhoods are activated, they have to compete with only a few neighbors. Thus, repetition latency is faster for words from low density neighborhoods than for words from high density neighborhoods.

For nonwords, however, there are no lexical level list chunks, so processing occurs primarily at the level of sublexical level list chunks. At that level, chunks that incorporate higher phonotactic probabilities will dominate; this leads to a processing advantage for nonwords from high density neighborhoods, because higher density stimuli incorporate higher phonotactic probabilities.

This formulation has provided a useful organizing framework for thinking about neighborhood effects in nonword repetition, and offers an account of neighborhood effects in lexical decision and speeded same–different discrimination performance as well. In a lexical decision task, Luce and Pisoni (1998) found that higher density neighborhoods were associated with slower reaction times for both words and nonwords. According to Vitevitch and Luce's (1999) framework, this makes sense because both words and nonwords must necessarily be engaged at the lexical level if an accurate lexical decision is to be made. Consistent with this, in a same–different discrimination task in which words and nonwords were blocked by lexical status (i.e., all words or all nonwords in a block), same–different discrimination times were faster for nonwords from high density neighborhoods than low density neighborhoods, but faster for words from low density neighborhoods than high density neighborhoods (Vitevitch & Luce, 1999). According to Vitevitch and Luce's framework, this is because the words were processed at the lexical level, with greater competition among high density stimuli, whereas nonwords were processed at the sublexical level, at which higher density stimuli have a processing advantage because of their higher phonotactic probabilities. When words and nonwords were presented together within the same blocks, however, the discrimination latency advantage for words from low density neighborhoods was reduced. In terms of Vitevitch and Luce's framework, this is because the interleaving of word and nonword stimuli leads to processing being focused at the sublexical level, which is the primary shared level of processing for words and nonwords; because phonotactic probability dominates processing at this level, low density is less facilitative.

This framework has thus been useful in thinking not only about immediate repetition tasks, but also in thinking about lexical decision and same–different discrimination tasks. An important aspect of the framework remains, however, the posited difference in the effects of higher neighborhood density on repetition latency for words versus nonwords in an immediate repetition task. To the extent that higher neighborhood density facilitates repetition latency for nonwords in this task, this is consistent with the framework. To the extent that higher neighborhood density does not uniformly facilitate repetition latency for nonwords in an immediate repetition task, this would be less consistent with the framework. The theoretical import of the present investigations is therefore that they offer a means of testing the framework proposed by Vitevitch and Luce (1999), by examining whether, and if so, under what circumstances, higher neighborhood density facilitates immediate repetition latency for nonwords.

Below, we report 12 experiments that examined the effect of neighborhood density on nonword repetition latency in an immediate repetition task. The experiments are presented in three groups of four experiments each. Each group of four experiments has a similar structure; the groups differ primarily in the particular stimulus set they examined. The first two experiments in each group of four examined the effect of neighborhood density on nonword repetition latency in two variants of the underlying stimulus set; the third and fourth experiments in each group replicated the first two experiments in that group using an altered stimulus presentation rate.

Experiments 1–4

Experiments 1–4 employed the original stimulus tokens used by Vitevitch and Luce (1998). Experiment 1 was a straightforward replication of the nonword repetition results from Vitevitch and Luce (1998). It examined the effect of neighborhood density on nonword repetition latency using exactly the same stimulus tokens used in that previous study. For Experiment 2, we digitally altered the duration of these stimulus tokens in an attempt to equate the duration of the high and low neighborhood density stimuli. Experiments 3 and 4 repeated these investigations (with the original and duration-adjusted stimuli, respectively), but employed a different presentation rate.

Subsequent to completion of these four experiments, it was brought to our attention by M. Vitevitch and P. Luce that the digital audio files comprising the duration-adjusted stimuli used in Experiments 2 and 4 contained substantial amounts of leading and trailing silence. File duration and true stimulus duration thus differed for these stimuli. This discovery had two consequences. First, it meant that the stimuli used in Experiments 2 and 4 had in fact been equated for file durations (which included substantial leading and trailing silences surrounding the stimulus), rather than true stimulus durations, as we had intended. The implications of this for Experiments 1–4 will be examined as part of the presentation of those experiments below. The second consequence of this discovery was that it raised the question of where these leading and trailing silences might have originated.

In investigating this question, we first reconfirmed that the original stimulus files provided to us by M. Vitevitch and P. Luce had mean file durations of

689.6 ms for the 120 high density stimuli, and 706.1 ms for the 120 low density stimuli, which did not differ significantly, $F(1,238) = 2.55$, $p = .11$. We then verified that these file durations and statistics correspond exactly with those reported as pertaining to stimulus durations for these stimuli (Vitevitch & Luce, 1998, p. 327; Vitevitch & Luce, 1999, p. 382). This exact correspondence indicates that in those previous reports, what was referred to as stimulus duration was in reality file duration. Further analysis revealed that these stimulus files incorporate leading and trailing silences of mean duration 28.1 and 51.7 ms, respectively. This indicated that the original stimuli used in the studies by Vitevitch and Luce (1998, 1999) were in fact the source of the substantial leading and trailing silences, which had simply carried over when we adjusted the durations of these original stimuli for use in our Experiments 2 and 4. It also indicated that what had been reported in Vitevitch and Luce's (1998, 1999) studies as statistically equivalent *stimulus* durations for stimuli from high and low density neighborhoods (690 and 706 ms, respectively; Vitevitch & Luce, 1999, p. 382) were in reality statistically equivalent *file* durations. Our analysis of true stimulus durations (that is, file durations minus leading and trailing silences) indicated a mean true stimulus duration of 589 ms for the 120 high density stimuli and 647 ms for the 120 low density stimuli used in Vitevitch and Luce (1998, 1999). Thus, the high density stimuli used in those studies were 58 ms shorter than the low density stimuli; the difference is significant $F(1,238) = 31.75$, $p < .001$. The report of equivalent stimulus durations for the high and low density stimuli made in Vitevitch and Luce (1998, 1999) was therefore only accurate in terms of file durations; it was inaccurate with respect to true stimulus durations. The large and significant difference in true stimulus durations was therefore confounded with the density status of the stimuli, which in turn confounds Vitevitch and Luce's (1998) finding of faster repetition latencies for the high density nonwords. The further implications of this will be considered following presentation of Experiment 1, as well as in the General discussion, where we consider the broader impact of this stimulus duration confound on other published studies that have used these stimuli.

Method

Several aspects of the method were common across Experiments 1–4, and are discussed below.

Participants

The participants in Experiment 1 and all experiments reported in this article were students from an introductory psychology course at the University of Iowa. All were native speakers of English who reported having normal hearing and normal or corrected-to-normal vision. No participant took part in more than one of the experiments reported here. There were fifteen participants in each of Experiments 1 and 2, and 20 participants in each of Experiments 3 and 4.

Procedure

All stimuli were presented binaurally at a comfortable listening level with a Macintosh G3 PowerPC computer through Aiwa HPX 222 headphones. The PsyScope experiment development system (Cohen, MacWhinney, Provost, & Flatt, 1993) controlled all stimulus presentation and repetition latency recordings. Responses were registered through a Shure SM10A-CN microphone mounted on headphones and individually adjusted to an appropriate sensitivity level.

On each trial of this single-block within-subjects design, participants heard a nonword through the headphones. Participants' instructions were to repeat each nonword as quickly but as accurately as possible. The next nonword was presented 1000 ms after offset of the previous nonword. During the first 500 ms of this 1000 ms ITI, a fixation cross was presented on the display, to signal that the previous stimulus had ended and that the next stimulus would be presented shortly; however, participants were instructed not to link their responding to the appearance of the cross, but to respond as quickly and accurately as possible to each stimulus.

Scoring

Consistent with the original Vitevitch and Luce (1998) procedures, response latency was measured from onset of the stimulus to the onset of the participant's repetition. As is now known, these stimuli incorporate leading silences, so that there is a distinction between file onset and true stimulus onset. In such a situation, measuring from stimulus file onset is incorrect, because it includes the leading silence in the calculated response latency. (The stimulus has not yet begun during this leading silence, therefore this silence should not be included in the participant's response initiation latency). The correct onset to use would be true stimulus onset. The latencies reported in Vitevitch and Luce (1998), however, were measured from stimulus file onset, as the distinction between file and stimulus duration for these stimuli had not been discovered at that time. For the present experiment, we therefore report latency measured both from stimulus file onset (to enable comparison with the results reported in Vitevitch & Luce, 1998), and from true stimulus onset.

For analysis of repetition latencies (whether from stimulus file onset or from true stimulus onset) by subjects, responses more than 2.5 standard deviations from the mean (calculated separately for high and low density stimulus repetitions for each individual participant) were excluded. For analysis of repetition latencies by item (again, whether from stimulus file onset or from true stimulus onset), responses more than 2.5 standard deviations from the mean (calculated separately for the high and low density stimulus repetitions) were excluded. Experimenters scored repetition accuracy on-line, marking as correct only those repetitions that matched the presented nonword on each phoneme.

Experiment 1

Stimuli

Stimuli consisted of the 240 consonant–vowel–consonant (CVC) spoken nonwords used by Vitevitch and Luce (1998), provided to us as digital audio files by M. Vitevitch and P. Luce. As reported by Vitevitch and Luce (1998), all stimuli were low-pass filtered at 4.8 kHz and digitized at a sampling rate of 10 kHz using a 12-bit analog-to-digital converter. The slope of the low-pass filter was not reported. Half of these stimuli were classified as high neighborhood density nonwords. The other half consisted of low neighborhood density nonwords. Six nonwords from each of the high density and low density sets were randomly selected for the practice trials preceding the experiment. These 12 nonwords were presented only during the practice and the recorded repetition latencies were not included in the final analysis. Thus, the number of nonword stimuli included in the final analysis was 228.

Neighborhood measures

Vitevitch and Luce (1998) computed nonword neighborhood similarity measures by comparison of nonword phonemic transcriptions with all entries in an on-line version of Webster's dictionary (Luce & Pisoni, 1998; Vitevitch & Luce, 1998). The neighborhood of a given nonword was defined as all dictionary entries differing from that nonword by a single phoneme addition, deletion, or substitution. The log frequencies of all neighbors (Kucera & Francis, 1967) were subsequently calculated and summed to yield the frequency-weighted neighborhood density. The mean frequency-weighted neighborhood density for the low neighborhood density nonwords was 13.46. For high neighborhood density nonwords, the mean was 44.61 (see Vitevitch & Luce, 1998, for additional details).

Timing and initial segments

Vitevitch and Luce (1998) reported that stimulus duration differences between the high density and low density sets were not statistically significant, $F(1,238) = 2.54$, $p = .11$. (As is now known, the durations that were measured were actually stimulus file durations). Initial segments were balanced for abrupt and nonabrupt onsets (analogous to Jakobson, Fant, & Halle's (1951) continuous–discontinuous dimension) across sets to avoid a possible confound of differential voice key sensitivity as a result of onset phoneme distribution differences (Vitevitch & Luce, 1998).

Results and discussion

We first report results with latencies measured from stimulus file onset (as in Vitevitch & Luce, 1998). Elimination of response latencies that lay more than 2.5 standard deviations from the mean resulted in exclusion of 1% of the data in the subject analysis and 1.3% of the data in the item analysis. For the analysis of repetition latencies by subjects, a significant mean difference of 18.9 ms in repetition latency between the high ($M = 951.98$, $SD = 68.27$) and low ($M = 970.89$, $SD = 72.48$) density stimuli was obtained, $F_1(1,14) = 25.67$, $p = .0002$. For the analysis of repetition latencies by items, there was a significant mean difference of 20.9 ms between high ($M = 950.36$, $SD = 70.02$) and low ($M = 971.27$, $SD = 68.51$) density stimuli, $F_2(1,226) = 5.19$, $p = .02$.

These repetition latency results are consistent with those obtained by Vitevitch and Luce (1998): the present experiment found a significant 18.9 ms repetition latency advantage for high neighborhood density nonwords, comparable to the significant repetition latency difference reported in Vitevitch and Luce (1998), estimated from their Figure 1 as being 20–25 ms. Absolute repetition latencies are also comparable with the latencies of approximately 1000 ms reported by Vitevitch and Luce (1998).

Measuring latency from true stimulus onset yielded the same pattern of results. Elimination of response latencies that lay more than 2.5 standard deviations from the mean resulted in exclusion of 1.2% of the data in the subject analysis and 0.9% of the data in the item analysis. For the analysis of repetition latencies by subjects, a significant mean difference of 37.1 ms in repetition latency between the high ($M = 909.5$, $SD = 69.7$) and low ($M = 946.6$, $SD = 74.8$) density stimuli was obtained, $F_1(1,14) = 88.0$, $p < .0001$. For the analysis of repetition latencies by items, there was a significant mean difference of 36.1 ms between high ($M = 909.4$, $SD = 75.6$) and low ($M = 945.5$, $SD = 71.2$) density stimuli,$F_2(1,226) = 13.8$, $p = .0003$.

The latency results obtained in Experiment 1 were thus closely similar to those of Vitevitch and Luce (1998) when latency was measured in the same way (from stimulus file onset), and the pattern of results did not change when latency was measured from true stimulus onset. Experiment 1 thus replicated the findings reported by Vitevitch and Luce (1998) for repetition latency.

Repetition accuracy was 68% for the high density stimuli and 78.1% for the low density stimuli. The difference was statistically significant, $F(1,14) = 13.02$, $p = .003$. These results were lower than those reported by Vitevitch and Luce (1998) (88 and 87% correct for each condition, respectively), and also differed in that repetition accuracy differed significantly between high and low density stimuli in the present experiment, but not in Vitevitch and Luce's (1998) study. To understand these differences, it is worth considering how accuracy was determined in both studies. A rater compared the participant's spoken response with a written transcription of the stimulus, and scored the repetition response as being overall correct or incorrect (rather than phoneme-by-phoneme). This overall correctness scoring procedure entails that the rater set a subjective threshold for deciding when a stimulus is overall correct or incorrect. Simply adopting a more or less stringent threshold changes the number of stimuli scored as correct versus incorrect, and hence the overall accuracy score for a set of repetition responses. Two different raters with different subjective correctness thresholds would therefore arrive at different repetition accuracy scores. The consequence of this is that it is difficult to compare accuracy results obtained in different experiments.

A likely explanation of the difference in repetition accuracy scores between the present experiment and Vitevitch and Luce's (1998) study therefore is that scoring in the present experiments may simply have incorporated a higher accuracy threshold than did Vitevitch and Luce's (1998) study. This in turn suggests that repetition accuracy may have been significantly greater for low density nonwords than for high density nonwords in the present experiment, but not in Vitevitch and Luce's (1998) study, simply because the lower accuracy levels in the present study may have unmasked an effect not apparent at the higher accuracy levels of the earlier study.

To test this possibility, the recorded responses for the high and low density stimuli were rescored by a rater who was naïve to the previously obtained accuracy scores, and who was instructed to adopt a lenient accuracy criterion. Scoring with this less stringent criterion led to ratings of 87.4% correct and 87.8% correct for the high and low density stimuli respectively, which did not differ from each other $F(1, 14) = .06$, $p = .81$, and mirrored the accuracy results of 88% for the high density stimuli and 87% for the low density stimuli reported by Vitevitch and Luce (1998). Changing the scoring threshold thus changed the accuracy results, as we hypothesized.

It is also worth noting that these differing accuracy scores are not simply random variation between raters; rather, the "strict" and "lenient" criteria are reliable. When a third naïve rater scored the results using the "strict" criterion, accuracy was 69.8 and 85.4% for the high and low density stimuli, respectively, which did differ significantly from each other, $F(1, 14) = 25.22$, $p = .0002$, and percent agreement on item correctness across the two "strict" scorers was 92.4%. When a fourth naïve rater scored the results using the "lenient" criterion, accuracy was 95.4 and 94.9% for the high and low density stimuli, respectively, which did not differ from each other, $F(1, 14) = .36$, $p = .55$, and percent agreement on item correctness across the two "lenient" scorers was 90.2%. Interrater agreement results for Vitevitch and Luce's (1998) study were not available for comparison.

These points indicate that a straightforward comparison of the present accuracy results with those of Vitevitch and Luce's (1998) study is not possible, given the inherent subjectivity of the scoring procedure employed in both studies. We will therefore focus on the repetition latency results, which in any case constitute the chief phenomenon of interest. We include both correct and incorrect responses in the analysis of latency, in this and all other experiments. Because of the arbitrariness of the accuracy scoring procedure, exclusion of incorrect responses would entail exclusion of arbitrarily different numbers of responses in different experiments, and under different scoring procedures. We therefore consider the inclusion of all responses to be the only nonarbitrary procedure in the present context.[1]

Nevertheless, for consistency with other experimental contexts in which it is appropriate to exclude incorrect responses from further analysis, we also analyzed latencies for correct responses only, with correctness determined by the lenient scoring criterion. This led to 48 additional analyses of latency differences for high versus low density stimuli, consisting of subject and item analyzes for repetition latencies measured from true stimulus onset as well as from stimulus offset (as will be discussed below), for each of twelve experiments. This analysis did not affect the results of the comparison between latencies for high versus low density stimuli except in five out of the 48 cases, where they nevertheless did not alter the conclusions that may be drawn. These five cases are discussed at the appropriate places in the discussion of each experiment below.

Returning to the present discussion of latencies, we note again that the present latency results replicated those of Vitevitch and Luce (1998), using both onset latency measures. However, as noted earlier, the stimuli used in the present experiment (and by Vitevitch & Luce,

[1] It may also be worth noting that the studies of nonword repetition reported by Vitevitch and Luce (1998, 1999) also appear to have included correct as well as incorrect responses in the analysis of latency. In some of the experiments in the 1999 study, the latency analysis was specifically described as being based on correct responses only (Experiments 1: same–different judgments; Experiments 2 and 3: lexical decision). For the task involving nonword repetition, however (Experiment 4), incorrect responses were not reported as having been excluded from the latency analysis; nor was there such an indication in the 1998 study of nonword repetition.

1998) incorporated a mean true stimulus duration of 589 ms for the high density nonwords and a mean of 647 ms for the low density nonwords. This statistically significant 58 ms duration advantage for the high density stimuli is confounded with the latency advantages obtained for high density stimuli in the original study and in the present experiment. It is also considerably larger than the 20–25 ms latency advantage for high density stimuli obtained by Vitevitch and Luce (1998) measuring latency from stimulus file onset, the 19 ms latency advantage for high density stimuli obtained in the present experiment when measuring latency from stimulus file onset, and the 37 ms latency advantage for high density stimuli obtained in the present experiment when measuring latency from true stimulus onset. Because the present experiment (following Vitevitch & Luce, 1998) measured repetition latency from stimulus onset, this stimulus duration difference was effectively added into the repetition latencies. This raised the possibility that the repetition latency advantage for high density stimuli obtained in the present experiment and by Vitevitch and Luce (1998) might have been partly or wholly due to the difference in stimulus duration, rather than the difference in neighborhood density.

To examine this question, we conducted an ANCOVA on the repetition latencies measured from true stimulus onset, with (true) stimulus duration as the covariate and neighborhood density as the categorical variable. The question of interest was whether high neighborhood density would still be associated with faster mean repetition latencies even after controlling for stimulus duration. Consistent with the method outlined by Cohen and Cohen (1983) using multiple regression, we first conducted a test of the complete model incorporating the covariate and categorical variables as well as the interaction. Testing for the significance of the interaction is necessary to insure appropriate interpretation of the ANCOVA results (Keppel, 1991). If the interaction is significant, it indicates that the impact of the categorical variable depends on the value of the covariate, and cannot be interpreted without reference to the value of the covariate; in the present instance, it would indicate that the impact of neighborhood density cannot be assessed without also taking stimulus duration into account. If, on the other hand, the interaction term is not significant, it can be dropped; the covariate and categorical weightings can then be recalculated without the interaction term to reveal the impact of neighborhood density on response latencies with stimulus duration factored out.

The initial regression testing for the significance of the interaction yielded a marginally significant regression coefficient for density ($b_1 = 121.6$, $p = .05$) and significant coefficients for duration ($b_2 = .67$, $p < .0001$) and for the Density × Duration interaction ($b_3 = -.2$, $p = .04$). It is important to note that the marginally significant coefficient for density indicates an overall trend for repetition latency to be *longer* for high density stimuli. However, the significant interaction indicates that the Density × Duration term cannot be justifiably removed from the equation and indicates that the effect of density on repetition depends upon stimulus duration. The obtained adjusted R^2 for this final model was .403.

Another possible means of factoring out the effects of stimulus duration is to measure repetition latencies from stimulus offset rather than onset. Although, given the interaction revealed in the ANCOVA, this would not eliminate all effects of stimulus duration, it would eliminate effects that arose simply because participants waited to respond until stimulus presentation was complete. We therefore re-analyzed the present results with repetition latency measured from stimulus offset (true stimulus offset, rather than stimulus file offset). Elimination of response latencies that lay more than 2.5 standard deviations from the mean resulted in exclusion of 2.3% of the data in the subject analysis and 2.3% of the data in the item analysis. For the analysis of repetition latencies by subject, high ($M = 329.9$, $SD = 66.2$) and low ($M = 305.9$, $SD = 70.4$) density nonwords did differ significantly, with latency being significantly *longer* for the high density nonwords $F_1(1, 14) = 40.1$, $p < .0001$. For the analysis by item, high density nonwords ($M = 330.8$, $SD = 73.0$) were repeated with significantly longer latency than low density nonwords ($M = 305.2$, $SD = 53.9$), $F_2(1, 226) = 9.1$, $p = .003$.

Our analysis of the results of Experiment 1 thus indicates that the shorter true stimulus duration of the high density as compared with the low density stimuli used in this experiment (and in Vitevitch & Luce, 1998) was not merely a logical confound, but in fact accounted for a substantial amount of the variance in an ANCOVA. Moreover, the latency advantage tended marginally in the opposite direction to that reported by Vitevitch and Luce (1998), although this is difficult to interpret because of the interaction. However, the evidence is thus far purely statistical in nature. If stimulus duration is indeed an important factor in determining nonword repetition latency, it should be possible to manipulate the effect of density by controlling stimulus duration experimentally as well. Specifically, reduction of the mean stimulus duration difference between the high and low density stimuli to 0 ms should reduce the latency difference between high and low density stimuli obtained in Experiment 1. Experiment 2 was designed with this goal in mind.

Experiment 2

In Experiment 2, we aimed to digitally alter the durations of the stimuli used in Experiment 1 so that the mean durations of the high and low neighborhood density nonwords would be equal. As we were unaware of the leading and trailing silences in these digital stimuli, we used stimulus file durations as the basis of our calculations and adjustments. As is now known, true stimulus durations differ from stimulus file durations for these stimuli, because of the leading and trailing silences. Our modifications therefore only achieved equalization of mean file durations, rather than of true stimulus durations. The following discussion describes our manipulations of what turned out to be stimulus file duration.

Method

Stimuli

The stimuli were created by expanding or compressing the durations of the stimuli used by Vitevitch and Luce (1998) and in Experiment 1. To match the high density and low density sets for duration, we first calculated the mean stimulus duration separately for the high density and low density sets, obtaining mean durations of 690 ms ($SD =$ 89.9) and 706 ms ($SD =$ 67.99), respectively. The average of these means provided a target mean duration of 698 ms. In each of the high and low density sets, some of the stimuli were expanded in duration and some were compressed, using the "tempo" feature of the SoundEdit 16 Version 2 software (SoundEdit 16, 1995), which alters the duration of a sound stimulus without changing its pitch. In order to equate the mean durations of the stimulus sets while minimizing the number of stimuli to be altered, the initial selection and expansion of the high density stimuli started with those of the shortest duration and progressed until the targeted mean was achieved. Analogously, initial selection and compression of the low density stimuli started with those of the longest duration. In addition, a few stimuli in each set were expanded or compressed to avoid the introduction of extreme skewness.[2] The mean modification in stimulus duration was 5.7% for high density stimuli and 5.2% for low density stimuli. For high density stimuli, 46 high density nonwords were expanded an average of 36 ms and 16 were compressed an average of 41 ms, with duration modifications ranging from 17 to 66 ms, yielding a final mean duration of 698 ms ($SD = 67.75$) for the duration-adjusted high density set. For low density stimuli, 43 nonwords were compressed an average of 38 ms and 18 were expanded for an average of 36 ms, with duration modifications ranging from 15 to 66 ms, yielding a final mean duration of 698 ms ($SD = 48.41$) for the duration-adjusted low density set.

It is important to consider whether the compression and expansion made the stimuli seem unnatural. To test this possibility, a same–different discrimination task was conducted with a different group of participants. Sixty stimulus types were randomly chosen from the set of 62 high density stimulus types whose duration had been modified (46 stretched and 16 compressed, see above), and 60 stimulus types were randomly chosen from the set of 61 low density stimulus types whose duration had been modified (43 compressed and 18 stretched, see above), and the original unmodified tokens of these high and low density stimulus types were then selected. For the same–different discrimination task, a randomly chosen half of these 60 original high density stimulus tokens were paired with themselves for a same–different discrimination, and half were paired with the duration-modified stimulus token of the same type. Similarly, a randomly chosen half of these 60 original low density stimulus tokens were paired with themselves for a same–different discrimination, and half were paired with the duration modified stimulus token of the same type. A different random selection of stimuli was used for each participant.

In the same–different discrimination task, participants listened to pairs of nonword stimuli, presented one after the other with an ISI of 250 ms, and made a same–different judgment about the pair. Thus the pairs of stimuli in the discrimination task consisted of either two original tokens of the same type (i.e., two presentations of identical tokens) or a pairing of the original and modified tokens of the same type. Presentation order was counterbalanced across original–modified pairings. High and low density stimulus pairs were randomly intermixed.

If the modified versions of the stimuli were detectably different from their original versions, then the proportion of "Different" responses should be higher for original–modified pairings relative to the original–original pairings. However, 20 participants showed an equivalent proportion of "Different" responses across the original–original (12%) and original–modified (12.4%) pairings, $t(19) = .381$, $p = .71$. This result indicates that the duration manipulation did not give rise to any perceptible distortion of the duration-adjusted stimuli. Experiment 2 was therefore conducted as planned, employing the duration-adjusted versions of the stimuli used in Experiment 1.

[2] Ideally, we would have simply increased the duration of each high density stimulus by 8 ms and decreased the duration of each low density stimulus by 8 ms to obtain the target mean duration. However, the "tempo" feature of SoundEdit does not offer an absolute duration modification that can be applied uniformly to every stimulus; rather, it allows a differing range of possible modifications for different stimuli based on their total duration. The alternate strategy we actually adopted therefore changed stimuli so as to achieve the target mean duration without introducing extreme skewness. Skewness was .307 for the original high density set and .25 for the duration-adjusted set. Skewness was $-.318$ for the original low density set and $-.191$ for the duration-adjusted set. None of these values represents a significant departure from normality according to the assessment procedures outlined by Snedecor and Cochran (1980).

Of course, as is now known, our modifications did not result in equal true stimulus durations for high and low density nonwords, because of the leading and trailing silences in the original stimuli. The mean true stimulus durations became 592.7 and 638.8 ms for the 114 high and low density stimuli, respectively, incorporating a duration advantage of 46.1 ms for the high density nonwords. Although this was reduced from the 58 ms duration advantage in Experiment 1, it nevertheless remained significant, $F(1,226) = 30.0$, $p < .001$.

Procedure

The procedure followed for Experiment 2 was identical to that employed for Experiment 1.

Results and discussion

Repetition accuracy was 90.6% for the high density stimuli and 92% for the low density stimuli using the lenient scoring criterion, which we adopt henceforth. The accuracy difference for these scores was not statistically significant, $F(1,14) = 1.65$, $p = .22$. (As a means of highlighting once again the arbitrariness of accuracy scoring, we note that repetition accuracy was 72.3% for the high density stimuli and 82% for the low density stimuli using the strict scoring criterion. The difference was statistically significant, $F(1,14) = 18.49$, $p < .001$).

All discussion of repetition latencies and stimulus durations below refers to latencies measured from true stimulus onset or true stimulus offset, and to true stimulus durations. We first consider response latencies measured from (true) stimulus onset. Elimination of response latencies that lay more than 2.5 standard deviations from the mean resulted in exclusion of 1.4% of the data in the subject analysis and 0.8% of the data in the item analysis. For the analysis of repetition latencies by subjects, the mean difference between the high ($M = 928.6$, $SD = 57.8$) and low ($M = 958.1$, $SD = 51.9$) density stimuli was 29.5 ms and was significant, $F_1(1,14) = 50.9$, $p < .0001$. For analysis of repetitions by item, the mean difference of 28.4 ms between the high ($M = 930.5$, $SD = 70.1$) and low ($M = 958.9$, $SD = 58.8$) density stimuli was significant, $F_2(1,226) = 10.9$, $p = .001$.

Reducing the (true) stimulus duration advantage for high density stimuli from 58 ms in Experiment 1 to 46 ms in Experiment 2 thus reduced the latency advantage for high density stimuli (measured from true stimulus onset) from 37.1 to 29.5 ms, consistent with our hypothesis that the effect of neighborhood density in Experiment 1 was not independent of stimulus duration. However, because our duration modifications had been only partially successful, there remained a significant duration advantage for the high density stimuli, and, as in Experiment 1, this confounded the significant latency advantage for the high density stimuli that was obtained when measuring latency from stimulus onset.

As in Experiment 1, therefore, an ANCOVA was conducted on latency measured from true stimulus onset, with true stimulus duration as the covariate and neighborhood density as the categorical variable. As in Experiment 1, the question of interest was whether high neighborhood density would still be associated with faster mean repetition latencies even after controlling for stimulus duration. The initial regression testing for the significance of the interaction yielded significant regression coefficients for duration ($b_2 = .51$, $p < .0001$) but not density ($b_1 = 81.1$, $p = .3$) or the Density × Duration interaction ($b_3 = -.15$, $p = .25$). The obtained adjusted R^2 for this model was .21. Given that the interaction was nonsignificant, a subsequent regression including only density and duration was then run, thus allowing for a more direct interpretation of the density coefficient after controlling for duration. Results again yielded a significant coefficient for duration ($b_2 = .43$, $p < .0001$) but not density ($b_1 = -8.8$, $p = .29$). The obtained adjusted R^2 for this final model was .21.

As a further means of exploring these results, we analyzed repetition latencies measured from (true) stimulus offset rather than onset. Elimination of response latencies that lay more than 2.5 standard deviations from the mean resulted in exclusion of 2.1% of the data in the subject analysis and 1.9% of the data in the item analysis. For the analysis of repetition latencies by subject, high density nonwords ($M = 341.9$, $SD = 55.3$) were repeated with longer latencies than low density nonwords ($M = 324.4$, $SD = 51.3$), a significant 17.4 ms difference $F_1(1,14) = 16.8$, $p = .001$. For the analysis by item, high density nonwords ($M = 345.4$, $SD = 72.4$) were repeated with significantly longer latency than low density nonwords ($M = 324.8$, $SD = 57.5$), $F_2(1,226) = 5.7$, $p = .02$.

The present results and analysis thus demonstrated the same pattern as in Experiment 1. When measuring repetition latency from stimulus onset, which includes stimulus duration in the latency calculation, a latency advantage was obtained for high density stimuli. However, an ANCOVA indicated that stimulus duration was the variable driving this effect; indeed, in Experiment 2, when this variable was taken into account, neighborhood density did not have a significant effect.

But how robust are these results? One possible factor that might affect repetition latencies is the presentation rate of the stimuli. A slower presentation rate, for instance, might allow more time for processing of stimuli, possibly leading to a different pattern of results. It therefore seemed important to consider whether the results described so far would hold up under a different presentation rate. Experiments 3 and 4 addressed this question by replicating Experiments 1 and 2 with a different regimen of presentation rates. In Experiments 1 and 2, the next nonword stimulus was presented 1000 ms after offset of the previous stimulus. Experiments 3 and 4 employed a slower presentation regimen such that the next nonword stimulus was presented 4500 ms after offset of the previous stimulus. In all cases, however, participants were instructed to repeat each nonword as quickly but as accurately as possible.

Experiment 3

Stimuli consisted of the 240 consonant–vowel–consonant (CVC) spoken nonwords used by Vitevitch and Luce (1998) and in Experiment 1. With the exception of the slower nonword presentation rate, the procedure followed for Experiment 3 was identical to that employed for Experiment 1.

Results and discussion

Repetition accuracy was 89.5% for the high density stimuli and 91.3% for the low density stimuli using the less stringent criterion as discussed in Experiment 1. The difference was not statistically significant, $F(1, 19) = 1.97$, $p = .17$. Elimination of response latencies (measured from true stimulus onset) that lay more than 2.5 standard deviations from the mean resulted in exclusion of 1.2% of the data in the subject analysis and 1.0% of the data in the item analysis. For the analysis of repetition latencies by subjects, the mean difference between the high ($M = 1194.2$, $SD = 151.8$) and low ($M = 1228.7$, $SD = 148.7$) density stimuli was 34.5 ms and was significant, $F_1(1, 19) = 23.5$, $p < .0001$. For analysis of repetition latencies by item, the mean difference of 34 ms between the high ($M = 1193.7$, $SD = 82.7$) and low ($M = 1227.7$, $SD = 77.3$) density stimuli was significant, $F_2(1, 226) = 10.2$, $p = .002$.

In an ANCOVA, the initial regression equation testing for the significance of the interaction component yielded a significant regression coefficient for duration ($b_2 = .62$, $p < .0001$) but not for density ($b_1 = 85.2$, $p = .24$) or for the Density × Duration interaction ($b_3 = -.14$, $p = .23$). The obtained adjusted R^2 for this model was .314. Given the nonsignificant interaction, a subsequent regression including only density and duration was run, yielding a significant coefficient for duration ($b_2 = .54$, $p < .0001$) but not density ($b_1 = -.83$, $p = .93$). The obtained adjusted R^2 for this final model was .31. (It should be noted that with incorrect responses excluded from the latency analysis, the same results were obtained in the initial regression for the duration and density coefficients, but the interaction term was significant, indicating that the effects of density and duration cannot be separated. However, this does not alter the essential conclusion that may be drawn: density did not have a main effect when duration was controlled.)

As in previous analyses, as another means of factoring out the effects of stimulus duration, we analyzed repetition latencies measured from (true) stimulus offset rather than onset. Elimination of response latencies that lay more than 2.5 standard deviations from the mean resulted in exclusion of 2.9% of the data in the subject analysis and 2.6% of the data in the item analysis. For the analysis of repetition latencies by subject, high density nonwords ($M = 617.9$, $SD = 136.0$) were repeated with longer latencies than low density nonwords ($M = 595.3$, $SD = 134.8$), a significant 22.6 ms difference $F_1(1, 19) = 10.9$, $p = .003$. For the analysis by item, high density nonwords ($M = 620.4$, $SD = 78.2$) were repeated with significantly longer latency than low density nonwords ($M = 595.4$, $SD = 65.0$), $F_2(1, 226) = 6.9$, $p = .009$.

The latency and accuracy results of Experiment 3 thus mirror those of Experiment 1 in showing a latency advantage for high density stimuli when latency is measured from stimulus onset. However, the results of the ANCOVA in both cases reveal a strong effect of stimulus duration, showing that the effect of density is either not significant when stimulus duration is taken into account (Experiment 3) or that the effect of density cannot be separated from that of stimulus duration (Experiment 1). This corroborating pattern of results held up despite the variation in presentation rate in Experiment 3 as compared to Experiment 1. Next, in Experiment 4, we examined whether the effects of neighborhood density on repetition latency obtained with the duration-adjusted stimuli in Experiment 2 would also hold up using the presentation regimen of Experiment 3.

Experiment 4

Stimuli consisted of the 240 consonant–vowel–consonant (CVC) spoken nonwords used in Experiment 2, which were adjusted to equate mean stimulus duration across neighborhood densities. The procedure was identical to that of Experiment 3, employing the slower presentation rate.

Results and discussion

Repetition accuracy was 92% for the high density stimuli and 93.6% for the low density stimuli. The difference was statistically significant, $F(1, 19) = 10.79$, $p = .003$. Elimination of response latencies (measured from true stimulus onset) that lay more than 2.5 standard deviations from the mean resulted in exclusion of 1.4% of the data in the subject analysis and 2.1% of the data in the item analysis. In the subject analysis, a significant repetition latency advantage of 19.1 ms was obtained for the high ($M = 1266.2$, $SD = 203.8$) over the low ($M = 1285.3$, $SD = 193.3$) density stimuli, $F_1(1, 19) = 5.9$, $p = .02$. For analysis of repetition latencies by item, the obtained mean advantage of 20.8 ms for the high ($M = 1254.2$, $SD = 74.4$) over the low ($M = 1275.0$, $SD = 76.6$) density stimuli was significant, $F_2(1,226) = 4.3$, $p = .04$ (with incorrect responses excluded, it was marginally significant, $p = .06$).

In an ANCOVA (both with and without incorrect responses included), the initial regression testing for the significance of the interaction component yielded a significant regression coefficient for duration ($b_2 = .60$, $p < .0001$) but not for density ($b_1 = 149.4$, $p = .11$) or the Density × Duration interaction ($b_3 = -.24$, $p = .11$). The obtained adjusted R^2 for this model was .16. Given the nonsignificant interaction, a subsequent regression including only density and duration was conducted, yielding a significant coefficient for duration ($b_2 = .45$, $p < .0001$) but not density ($b_1 = .18$, $p = .98$). The obtained adjusted R^2 for this final model was .16.

As an alternative means of factoring out stimulus durations, we analyzed repetition latencies measured from stimulus offset rather than onset. Elimination of response latencies that lay more than 2.5 standard deviations from the mean resulted in exclusion of 2.6% of the data in the subject analysis and 3.0% of the data in the item analysis. For the analysis of repetition latencies by subject, high density nonwords ($M = 677.5$, $SD = 177.1$) were repeated with longer latencies than low density nonwords ($M = 654.6$, $SD = 168.9$), a significant 22.9 ms difference $F_1(1,19) = 10.5$, $p = .004$. For the analysis by item, high density nonwords ($M = 671.4$, $SD = 76.2$) were repeated with significantly longer latency than low density nonwords ($M = 646.7$, $SD = 66.5$), $F_2(1,226) = 6.8$, $p = .009$.

This entire pattern of results was identical to that in Experiment 2. The results of Experiments 3 and 4 thus replicate the main findings of Experiments 1 and 2, suggesting that those results were not an artifact of the particular stimulus presentation rate employed. Results for all four experiments are summarized in Table 1, which shows that in all cases, when repetition latency was measured from stimulus onset, as in Vitevitch and Luce (1998), there was a significant latency advantage for high density stimuli. However, when duration was controlled, there was either no effect of neighborhood density (Experiments 2–4), or a marginally significant effect of density in favor of low density stimuli, together with an interaction of density with duration (Experiment 1). In no case was there a facilitatory effect of high density when duration was controlled. However, before drawing any firm conclusions about the effect of neighborhood density on nonword repetition latency, it seemed appropriate to conduct studies with additional sets of stimuli that differ in neighborhood density. The next several experiments were designed to provide such an investigation. The stimuli used in these remaining experiments were created under our control, and did not incorporate unduly large leading or trailing silences beyond what is unavoidable for such stimuli. Hence there is no distinc-

Table 1
Results summary for subject analyses for Experiments 1–12

	Stimulus durations (ms)		Onset RTs (ms)	Offset RTs (ms)	ANCOVA coefficient (ms)		
	High	Low			Density	Duration	Interaction
Type-V & L, Token-V & L							
Experiment 1	589	647	910(H) < 947(L)	330(H) > 306(L)	121.6[a]	.67**	−.2*
Experiment 2 (duration equated)	593	638	929(H) < 958(L)	342(H) > 324(L)	−8.8,ns	.43**	ns
Experiment 3	589	647	1194(H) < 1229(L)	618(H) > 595(L)	−.83,ns	.54**	ns
Experiment 4 (duration equated)	593	638	1266(H) < 1285(L)	678(H) > 655(L)	.18,ns	.45**	ns
Type-V & L, Token-L & G							
Experiment 5	663	695	923(H) = 926(L)	259(H) > 234(L)	19.39**	.79**	ns
Experiment 6 (filtered)	663	695	939(H) > 931(L)	271(H) > 234(L)	32.19**	.79**	ns
Experiment 7	663	695	1208(H) = 1210(L)	545(H) > 518(L)	16.62*	.73**	ns
Experiment 8 (filtered)	663	695	1207(H) > 1193(L)	564(H) > 521(L)	39.86**	.80**	ns
Type-L & G, Token-L & G							
Experiment 9	511	525	779(H) > 775(L)	269(H) > 253(L)	10.68*	.63**	ns
Experiment 10 (duration equated)	517	517	781(H) > 768(L)	268(H) > 256(L)	12.59**	.62**	ns
Experiment 11	511	525	883(H) = 880(L)	385(H) > 368(L)	14.60**	.72**	ns
Experiment 12 (duration equated)	517	517	953(H) > 931(L)	443(H) > 407(L)	29.98**	.72**	ns

(H) and (L) indicate high and low density stimuli, respectively. An equality indicates a nonsignificant difference whereas an inequality indicates a significant difference.
[a] p = .05.
* p < .05.
** p < .01.

tion between stimulus file duration and true stimulus duration for the stimuli used in Experiments 5–12.

Experiments 5–8 employed the same stimulus types used by Vitevitch and Luce (1998), but used tokens created by a different speaker.

Experiments 5–8

Method

Twenty participants took part in each of Experiments 5 and 6, and 30 participants in each of Experiments 7 and 8.

Stimuli consisted of the 240 consonant–vowel–consonant (CVC) spoken nonword *types* used in original Vitevitch and Luce (1998) work and in the present Experiments 1–4. However, all these nonword stimuli were now re-recorded by a female native English speaker using SoundEdit 16 Version 2 software (SoundEdit 16, 1995) at a sampling rate of 44.1 kHz and 16-bit resolution. Each nonword was recorded directly onto a computer hard drive and stored as an individual file. The mean durations for the high density and low density stimuli respectively were 663 ms ($SD = 74.5$) and 695 ms ($SD = 78.26$), a difference of 32 ms that was statistically significant, $F(1, 238) = 10.47$, $p = .001$. This duration difference was in the same direction as that in Vitevich and Luce's tokens used in Experiment 1, with low density stimuli being longer than high density stimuli.

All procedures followed for Experiments 5–8 were identical to those employed for Experiments 1–4, respectively.

Experiment 5

Experiment 5 examined repetition latencies for high and low neighborhood density nonwords using the newly recorded tokens of Vitevitch and Luce's (1998) stimulus types.

Results and discussion

Repetition accuracy was 97.1% for the high density stimuli and 94.6% for the low density stimuli. This difference was statistically significant, $F(1, 19) = 6.32$, $p = .02$. Results for analysis of latencies are shown in Table 2. There was no significant difference between repetition latencies for high and low density stimuli in either the subject or item ANOVAs, when measuring latencies from stimulus onset. Given that the measurement of latency from stimulus onset includes stimulus duration in the measured latency, and given that stimulus durations were significantly shorter for the high density stimuli, there was an inherent bias toward finding shorter latencies for high density stimuli when using this onset measure. Despite this bias, no latency advantage was obtained for the high density stimuli. This suggests that controlling for stimulus duration might reveal faster latencies for the low density stimuli.

An ANCOVA indicated significant regression coefficients for duration but not density or the Density × Duration interaction. The regression was therefore rerun without the interaction term, yielding significant coefficients for both duration and density. Given that low density was coded as a "zero," and high density as a "one," the density coefficient indicates that repetition latency was *longer* for high density nonwords than for low density nonwords. That is, there was a latency advantage for *low density* nonwords, which were repeated 19 ms more quickly than high density nonwords, once duration was controlled. Consistent with this, repetition latencies measured from stimulus offset were significantly faster for low density than high density nonwords in both subject and item analyses.

These results are opposite to those previously reported in the literature. Taken together with the results of Experiments 1–4, they suggest that the effect of neighborhood density on nonword repetition latency may be rather variable, being affected by stimulus duration (as indicated by Experiments 1–4), and by differences in stimulus tokens, as suggested by the present results. The present stimuli consisted of exactly the same nonword *types* used in Experiment 1 (and in Vitevitch & Luce's, 1998 study), and hence incorporated exactly the same neighborhood characteristics; they differed only in that the *tokens* had been recorded by a different speaker. However, rather than a repetition latency advantage for high density stimuli, we obtained a latency advantage for low density stimuli, once duration was controlled.

The present stimuli also differed, however, in that the stimuli used in Experiment 1 (i.e., the tokens used by Vitevitch & Luce (1998)) had been low-pass filtered, whereas the stimuli used in Experiment 5 had not. Experiment 6 examined whether the reversed effect of neighborhood density obtained in Experiment 5 might have been due to this difference.

Experiment 6

Experiment 6 was a replication of Experiment 5 in all respects, except that the stimuli used in Experiment 5 were low-pass filtered. The new stimulus set was produced by low-pass filtering each stimulus from Experiment 5 at 4.8 kHz (consistent with the low-pass filtering that had been applied to the stimuli used in Experiments 1–4) using Sound Studio 2.0 (Sound Studio 2.0, 2002). This software does not allow direct manipulation of the slope of the filter, but rather provides a window width parameter that indirectly controls the slope. The value of this parameter was set to 63 samples, which

Table 2
Results for Experiments 5–8

	Experiment 5	Experiment 6	Experiment 7	Experiment 8
Onset latency subject analysis (ms)				
% outliers eliminated	0.6	0.8	1	1.2
Hi latency mean	922.7	938.58	1208.26	1206.52
Hi latency *SD*	59.41	88.83	195.51	169.58
Lo latency mean	926.3	930.95	1209.5	1193.07
Lo latency *SD*	64.27	90.57	201.19	159.76
Lo–Hi mean difference	3.6	−7.63	1.24	−13.45
$F1\,(df)$	0.94(1, 19)	4.65(1, 18)	0.05(1, 29)	6.59(1, 29)
p value	.34	.03	.83	.02
Onset latency item analysis (ms)				
% outliers eliminated	1.8	3.4	3.7	2.5
Hi latency mean	920	949.5	1199.41	1202.87
Hi latency *SD*	78.51	79.29	75.91	86.39
Lo latency mean	925	942	1205.47	1187.51
Lo latency *SD*	78.98	77.43	76.94	81.94
Lo–Hi mean difference	5	−7.5	6.06	−15.36
$F2\,(df)$	0.23(1, 226)	0.52(1, 226)	0.36(1, 226)	1.92(1, 226)
p value	.63	.47	.55	.16
Initial ANCOVA coefficients				
Duration and p	.77, $p < .0001$	.76, $p < .0001$	.7, $p < .0001$	.77, $p < .0001$
Density and p	−7.25, $p = .9$	−8.14, $p = .89$	−19.99, $p = .75$	8.6, $p = .9$
Interaction and p	0.04, $p = .66$	0.06, $p = .66$	0.05, $p = .56$	0.05, $p = .65$
Coefficients without interaction term				
Duration and p	.79, $p < .0001$	.79, $p < .0001$	.73, $p < .0001$	.8, $p < .0001$
Density and p	19.39, $p = .005$	32.19, $p < .0001$	16.62, $p = .02$	39.86, $p < .0001$
R^2	0.59	0.59	0.52	0.53
Offset latency subject analysis (ms)				
% outliers eliminated	3.5	5.4	3.5	2.7
Hi latency mean	259.35	271.45	544.97	541.9
Hi latency *SD*	59.1	87.46	193.32	168.85
Lo latency mean	233.5	234.1	518	498
Lo latency *SD*	63.18	89.72	201.81	158.15
Lo–Hi mean difference	−25.85	−37.35	−26.97	−43.9
$F1\,(df)$	48.94(1, 19)	131.98(1, 18)	22.21(1, 29)	62.32(1, 29)
p value	< .0001	< .0001	< .0001	< .0001
Offset latency item analysis (ms)				
% outliers eliminated	2.2	3.7	3.7	2.4
Hi latency mean	257.49	281.81	536.19	537.69
Hi latency *SD*	50.15	50.07	51.93	61.14
Lo latency mean	230.47	244.25	509.33	498.04
Lo latency *SD*	53.35	52.52	55.58	56.43
Lo–Hi mean difference	−27.02	−37.56	−26.86	−39.65
$F2\,(df)$	15.52(1, 226)	30.55(1, 226)	14.21(1, 226)	25.89(1, 226)
p value	< .0001	< .0001	0.0002	< .0001

roughly corresponds to a filter slope of 24 dB/kHz. All other stimulus characteristics were identical to those described in Experiment 5, as was the procedure.

Results and discussion

One subject was eliminated due to a data collection error, resulting in 19 total subjects. Repetition accuracy was 90.2% for the high density stimuli and 93.1% for the low density stimuli, a statistically significant difference of 2.9%, $F(1, 18)$, $p = .01$, although in the opposite direction of that for the unfiltered stimuli of Experiment 5. Results for analysis of latencies are shown in Table 2. Repetition latency measured from stimulus onset was significantly faster for the low density than the high density stimuli in the subject but not the item ANOVA.

(With incorrect responses excluded, the latency advantage for low density over high density stimuli was only marginally significant: 919.3 ms (low) versus 927.2 ms (high), $p = .07$. Nevertheless, the pattern of results remained the same as with incorrect responses included, and in any case did not yield a high density advantage). The ANCOVA (both with and without incorrect responses) yielded the same pattern of results as in Experiment 5: significant regression coefficients for duration but not density or the Density × Duration interaction. Rerunning the regression without the interaction term yielded significant coefficients for both duration and density, as in Experiment 5. The density coefficient indicates that low density nonwords were repeated roughly 32 ms more quickly than high density nonwords, after controlling for stimulus duration. Consistent with this, repetition latencies measured from stimulus offset (both with and without incorrect responses included) were significantly faster for low density than high density nonwords in both subject and item analyses.

The present results thus replicate the main finding of Experiment 5: a repetition latency advantage for low density over high density nonwords, when duration is controlled. The results of Experiment 5 thus do not appear to have arisen merely from the fact that the stimuli were not low-pass filtered. Next, in Experiments 7 and 8, we examined the effect of presentation rate, as in Experiments 3 and 4.

Experiment 7

The goal of Experiment 7 was to examine the effect of nonword density on repetition latency using the stimuli introduced in Experiment 5, which consisted of re-recorded tokens of the stimulus types used by Vitevitch and Luce (1998). Experiment 7 differed from Experiment 5 only in using the slower presentation rate employed in Experiments 3 and 4.

Results and discussion

Repetition accuracy was 97.1% for the high density stimuli and 95.4% for the low density stimuli. This difference was statistically significant, $F(1,29) = 5.78$, $p = .02$. Results for analysis of latencies are shown in Table 2. As can be seen from the table, the pattern of results in Experiment 7 was identical to that in Experiment 5, suggesting that the findings of Experiment 5 were not an artifact of presentation rate. Next, in Experiment 8, we examined the effect of low-pass filtering the stimuli, as we did in Experiment 6.

Experiment 8

The stimuli in Experiment 8 were the low-pass filtered stimuli used in Experiment 6, and the procedure was that used in Experiments 3, 4, and 7, employing the slower presentation rate.

Results and discussion

One subject's accuracy data was not included due to experimenter error, thus resulting in 29 total subjects. Repetition accuracy was 87.9% for the high density stimuli and 93.4% for the low density stimuli, again a statistically significant difference, $F(1,28) = 17.03$, $p = .0003$, although in the opposite direction to that obtained with the unfiltered stimuli used in Experiment 7. Results for analysis of latencies are shown in Table 2. As can be seen from the table, this pattern of results was identical to that obtained in Experiment 6, indicating that the results of Experiment 6 were not simply a function of the particular presentation rates used in that experiment. (As in Experiment 6, with incorrect responses excluded, the latency advantage for low density over high density stimuli was reduced, and did not reach significance: 1179.2 ms (low) versus 1186.7 ms (high), $p = .14$. Nevertheless, the pattern of results remained the same as in Experiment 6, and in any case did not yield a high density advantage. The ANCOVA and analysis of latencies from stimulus offset was unaffected by exclusion of incorrect responses and was identical to that in Experiment 6.) In addition to Experiment 8 replicating Experiment 6, the finding of largely identical patterns of results in Experiments 6 and 8 as compared with Experiments 5 and 7 indicates that the presence or absence of low-pass filtering did not play a role in determining these results (see Table 1 for an abbreviated summary of results).

Overall, the results of Experiments 1–8 indicate two things. First, stimulus duration is an important factor to take into account when examining the effect of nonword neighborhood density on repetition latency. Second, repetition latency is not necessarily always faster for high neighborhood density than for low neighborhood density stimuli, when stimulus duration is controlled. Indeed, it was only when measuring latency from stimulus onset (and hence without controlling for stimulus duration) in Experiments 1–4 that a repetition latency advantage was obtained for high density stimuli. In all the other experiments and analyzes described thus far, there was either no significant repetition latency difference for high versus low density nonwords, an interaction between duration and density, or a latency advantage for low density nonwords over high density nonwords.

However, the results described in Experiments 1–8 were all obtained with variants of the particular stimulus types employed in the original study by Vitevitch and Luce (1998). It seemed important to examine what the impact of neighborhood density differences would be in a different set of nonword stimuli. Experiments 9–12 addressed this question.

Experiments 9–12

The goal of Experiments 9–12 was to extend the investigation of neighborhood density/duration effects to a different set of stimuli. To this end, we devised a novel set of nonwords that could be partitioned into a high neighborhood density set and a low neighborhood density set.

Method

All nonword stimuli in this new set were single syllables of consonant–vowel–consonant (CVC) construction and were recorded by a female native speaker of English (different from the speaker who recorded the stimuli used in Experiments 5–8). The stimuli were recorded using Cool Edit (1996) at a sampling rate of 44.1 kHz and 16-bit resolution. Each nonword was recorded directly onto a computer hard drive and stored as an individual file. Neighborhood density measures were computed in the manner previously described in Experiment 1. The stimuli were categorized as members of the high or low density group by the resulting frequency-weighted density measure. A total of 200 nonwords (100 high density and 100 low density) were generated for purposes of the experiment with an additional 10 created for practice. The average frequency-weighted neighborhood density measure for the high density set was 37.01 ($SD = 8.9$); it was 5.25 ($SD = 3.05$) for the low density set.

The initial phonemes for both the high density and low density stimuli were limited to 10 possibilities (/b/, /d/, /f/, /g/, /k/, /p/, /z/, /ʃ/,/tʃ/, and /d£) each of which appeared 10 times in both the high and low density stimulus sets. Unlike the stimulus sets used in Experiments 1–8, the mean durations for the high ($M = 510.93$, $SD = 69.54$) and low density ($M = 524.68$, $SD = 73.74$) sets did not differ significantly, $F(1, 198) = 1.84$, $p = .17$.

To ensure the power to detect differences in repetition latency with the new stimulus set, Experiment 9 was conducted with a larger number of participants. Thus, 48 participants took part in Experiment 9, 20 in Experiment 10, 40 in Experiment 11, and 20 in Experiment 12.

Experiment 9

The procedure followed was identical to that in Experiments 1, 2, 5, and 6.

Results and discussion

Repetition accuracy was 96.8% for the high density stimuli and 95.8% for the low density stimuli. The difference was statistically significant, $F(1, 47) = 9.32$, $p = .004$. Results for analysis of latencies are shown in Table 3. Repetition latency measured from stimulus onset was significantly faster for the low density than the high density stimuli in the subject but not the item analysis. (With incorrect responses excluded, the latency advantage for low density over high density stimuli in the subject analysis was reduced from 3.6 to 2.3 ms, and did not reach significance: 767.7 ms (low) versus 770.0 ms (high), $p = .21$. Nevertheless, the pattern of results remained the same as with incorrect responses included, and in any case did not yield a high density advantage). The ANCOVA (both with and without incorrect responses included) yielded significant regression coefficients for duration but not density or the Density $\times$ Duration interaction. Rerunning the regression without the interaction term yielded significant coefficients for both duration and density. As in Experiments 5, 6, 7, and 8, the density coefficient indicates a repetition latency advantage for low density stimuli once duration is controlled; low density stimuli were repeated approximately 11 ms faster than high density stimuli. Consistent with this, repetition latencies measured from stimulus offset (both with and without incorrect responses included) were significantly faster for low density than high density nonwords in both subject and item analyses.

Experiment 9 thus extends the investigation of Experiments 1–8 to a different set of nonwords. The results with the new stimulus set are consistent with the those of Experiments 1–8 in indicating that the repetition latency advantage for high density over low density nonwords is not robust; and that in fact the more common effect of neighborhood density on repetition latency appears to be in the reverse direction.

However, the control of stimulus duration in Experiment 9 was largely statistical and analytical in nature. As in the rationale for Experiment 2, we reasoned that it should be possible to manipulate the effect of density by controlling stimulus duration experimentally as well. In Experiment 1, low density stimuli were longer in duration than high density stimuli; the finding was that of a repetition latency advantage for high density stimuli (when measuring latency from stimulus onset, and without controlling for stimulus duration). Because stimulus durations are built into repetition latencies that are measured from stimulus onset, we reasoned that eliminating the longer mean stimulus duration for low density stimuli would reduce the repetition latency disadvantage for these stimuli; and that was the rationale of Experiment 2. In Experiment 9 also, low density stimuli were longer in duration than high density stimuli; when measuring latency from stimulus onset without controlling for stimulus duration, the finding was that of a small repetition latency advantage for low density stimuli. Experiment 10 aimed to examine the effect of eliminating this difference in mean stimulus duration between low and high density stimuli.

Table 3
Results for Experiments 9–12

	Experiment 9	Experiment 10	Experiment 11	Experiment 12
Onset latency subject analysis (ms)				
% outliers eliminated	0.9	1	1	0.9
Hi latency mean	779	780.65	883.13	953.35
Hi latency *SD*	61.9	65.93	148.46	125.87
Lo latency mean	775.4	767.85	880.13	930.55
Lo latency *SD*	63.23	64.8	151.79	114.39
Lo–Hi mean difference	−3.6	−12.8	−3	−22.8
$F1\,(df)$	4.83(1,47)	42.09(1,19)	.28(1,39)	19.26(1,19)
p value	.03	.0001	.6	.0003
Onset latency item analysis (ms)				
% outliers eliminated	2.1	2.1	3.6	3
Hi latency mean	782.59	780.46	910.33	965.16
Hi latency *SD*	53.54	52.65	56.08	61.08
Lo latency mean	780.6	767.84	905.6	935.15
Lo latency *SD*	57.06	53.91	65.09	66.65
Lo–Hi mean difference	−1.99	−12.62	−4.73	−30.01
$F2\,(df)$	0.07(1,198)	2.81(1,198)	0.3(1,198)	11.02(1,198)
p value	.79	.1	.58	.001
ANCOVA coefficients				
Duration and *p*	.65, $p < .0001$	.62, $p < .0001$	.77, $p < .0001$	.75, $p < .0001$
Density and *p*	31.48, $p = .34$	11.59, $p = .71$	70.65, $p = .04$	62.69, $p = .11$
Interaction and *p*	−0.04, $p = .53$	.002, $p = .97$	−0.11, $p = .09$	−.06, $p = .4$
Coefficients without interaction term				
Duration and *p*	.63, $p < .0001$	.62, $p < .0001$	.72, $p < .0001$	.72, $p < .0001$
Density and *p*	10.68, $p = .02$	12.59, $p = .003$	14.6, $p = .002$	29.98, $p < .0001$
R^2	0.67	0.69	0.71	0.67
Offset latency subject analysis (ms)				
% outliers eliminated	5	2.6	3	4.1
Hi latency mean	268.73	268.15	385.48	442.55
Hi latency *SD*	61.98	69.46	133.25	124.9
Lo latency mean	252.56	256.05	368.45	406.9
Lo latency *SD*	64.08	67.85	132.95	114.76
Lo–Hi mean difference	−16.17	−12.1	−17.03	−35.65
$F1\,(df)$	106.19(1,47)	46.29(1,19)	9.45(1,39)	47.98(1,19)
p value	< .0001	< .0001	0.004	< .0001
Offset latency item analysis (ms)				
% outliers eliminated	2.1	2.2	1.7	3
Hi latency mean	271.53	263.52	400.32	452.87
Hi latency *SD*	40.11	38.13	35.94	45.38
Lo latency mean	257.25	250.98	381.04	411.71
Lo latency *SD*	38.22	35.57	35.59	40.09
Lo–Hi mean difference	−14.28	−12.54	−19.28	−41.16
$F2(df)$	6.64(1,198)	5.78(1,198)	14.53(1,198)	46.22(1,198)
p value	.01	.02	.0002	< .0001

Experiment 10

The stimuli used in Experiment 10 were duration-modified versions of the stimuli used in Experiment 9. As file duration and true stimulus duration did not differ for the stimuli used in Experiment 9, the adjustment of stimulus durations did achieve the desired result of equating mean durations.

The mean duration of the stimuli used in Experiment 9 was 511 ms for the high density set and 525 ms for the low density. To control for this mean stimulus duration difference, the mean stimulus duration of both sets was made to equal 517 ms, the overall mean of all the stimuli combined. To achieve this, each stimulus in the high density set was stretched by 7 ms and each stimulus in the low density set was shortened by 7 ms using the Praat sound editing software (Boersma & Weenink, 2003). Changing the mean stimulus durations through the addition or subtraction of a constant was made possible by the use of a different sound editing software package, and provides a more intuitive approach to equating mean stimulus duration than that used in Experiment 2. By simply shifting each distribution by a 7 ms constant, the overall shapes of the respective duration distributions were maintained, and the comparatively small change in stimulus duration ($\pm$7 ms) minimized changes in stimulus integrity. Apart from these changes in stimulus tokens, the stimulus types were identical to those in Experiment 9. All other procedures were also identical to those implemented in Experiment 9.

Results and discussion

Repetition accuracy was 96.4% for the high density stimuli and 94.5% for the low density stimuli. This difference was significant, $F(1,19) = 7.38$, $p = .01$. Results for analysis of latencies are shown in Table 3. How do these results relate to those of Experiment 9? In Experiment 9, latencies measured from onset revealed a small but significant advantage for low density stimuli. However, these latencies included the stimulus durations, which were longer for the low density stimuli. Measuring repetition latency from stimulus offset revealed a larger significant advantage for low density stimuli, suggesting that this advantage was masked by stimulus duration differences that were included when measuring latency from stimulus onset. Experiment 10 eliminated mean stimulus duration differences between the high and low density stimuli experimentally. With stimulus duration experimentally controlled, the latency advantage for low density stimuli was greater than in Experiment 9 when latency was measured from stimulus onset (12.8 ms in Experiment 10 versus 3.65 ms in Experiment 9), consistent with the hypothesis that the stimulus duration differences in Experiment 9 had masked a larger underlying latency advantage for low density stimuli, when measuring latency from stimulus onset. For latencies measured from stimulus offset, however, stimulus duration differences would not be expected to mask underlying latency differences. Consistent with this, for latencies measured from stimulus offset, the latency advantage for low density stimuli was not higher in Experiment 10, even though stimulus duration had been controlled (12.1 ms in Experiment 10 versus 16.17 ms in Experiment 9).

Overall, the results of Experiments 9 and 10 extend the findings of Experiments 1–8 to a completely different set of stimuli, indicating, as in Experiments 1-8, that the inclusion of stimulus duration in latencies that are measured from onset does impact the results that are obtained. The results are also in agreement with the suggestion from Experiments 1–8 that high density nonwords do not always enjoy a repetition latency advantage over low density nonwords, and, indeed, that the latency advantage is frequently in the opposite direction.

Experiment 11

Experiment 11 aimed to replicate Experiment 9 but using the slower presentation rate of Experiments 3, 4, 7, and 8.

Results and discussion

Repetition accuracy was 97.2% for the high density stimuli and 96.8% for the low density stimuli. The difference was not statistically significant, $F(1,39) = 1.5$, $p = .23$. Results for analysis of latencies are shown in Table 3. As can be seen from the table, the pattern of results generally mirrored that of Experiment 9. Although, unlike Experiment 9, the latency advantage for low density over high density stimuli was not significant when latency was measured from stimulus onset, the ANCOVA, as in Experiment 9, indicated a significant effect of both stimulus duration and neighborhood density, and a nonsignificant interaction. The effect of density was in favor of low density stimuli, which were repeated approximately 14 ms more quickly than high density stimuli (as compared with 11 ms more quickly, in Experiment 9). Thus the key results of Experiment 9 held up in Experiment 11 despite the altered presentation regimen. Next, in Experiment 12, we re-examined the results of Experiment 10, using the slower stimulus presentation rate.

Experiment 12

The stimuli were the duration-equated stimuli of Experiment 10; the presentation rate was the slower one used in Experiment 3, 4, 7, 8, and 11.

Results and discussion

Repetition accuracy was 96.6% for the high density stimuli and 94.5% for the low density stimuli. The difference was statistically significant, $F(1,19) = 9.75$, $p = .006$. Results for analysis of latencies are shown in Table 3. As can be seen from the table, the pattern of results generally mirrored that of Experiment 10. Although, unlike Experiment 10, the latency advantage for low density over high density stimuli was significant when latency was measured from stimulus onset, in both the subject and item analyses, the ANCOVA, as in Experiment 10, indicated a significant effect of both stimulus duration and neighborhood density, and a nonsignificant interaction. The effect of density was again in favor of low density stimuli, which were repeated approximately 30 ms more quickly than high density stimuli (as compared with 13 ms more quickly, in Experiment 10). Thus the key results of Experiment 10 held up in Experiment 12 despite the altered presentation regimen.

Overall, the results of Experiments 9–12 (summarized more briefly in Table 1) indicate that the results of Experiments 1–8 were not dependent on the particular stimulus types employed in those experiments. They provide additional evidence that stimulus duration is an important variable in examining the effect of neighborhood density on repetition latency, that high density nonwords do not always enjoy a repetition latency advantage over low density nonwords, and that the latency advantage is frequently in the opposite direction.

General discussion

In this article, we presented the results of 12 experiments examining the effect of neighborhood density on repetition latency for nonwords in an immediate repetition task (see Table 1). Experiment 1 replicated the widely cited finding (Vitevitch & Luce, 1998) that repetition latency is faster in an immediate repetition task for nonwords from high density neighborhoods than for nonwords from low density neighborhoods, but also presented statistical analyses showing that the shorter true stimulus duration of the high density as compared with the low density stimuli was not merely a logical confound, but in fact accounted for a substantial amount of the variance. Experiment 2 further examined this possibility with stimuli whose durational differences had been reduced, replicating the results of Experiment 1 and of Vitevitch and Luce (1998), and finding that there was no effect of neighborhood density on repetition latency for the nonwords when stimulus duration was statistically controlled. Experiments 3 and 4 replicated Experiments 1 and 2 employing an altered presentation regimen. The results mirrored those of Experiments 1 and 2, suggesting that the findings of Experiments 1 and 2 are fairly robust across presentation regimen. Experiment 5 repeated the investigation with our own recordings of the same stimulus types used in Experiments 1 and 2, and found that repetition latency was actually faster for the low density nonwords than for the high density nonwords, once stimulus duration was controlled. Experiment 6 examined whether this result could simply have been due to the fact that the stimuli of Experiment 5, unlike those of Experiment 1, had not been low-pass filtered, and employed low-pass filtered versions of the Experiment 5 stimuli. The results indicated a repetition latency advantage for low density over high density nonwords, when duration was controlled, thus replicating the main finding of Experiment 5 even when the stimuli were low-pass filtered. Experiments 7 and 8 replicated Experiments 5 and 6 employing an altered presentation regimen, again confirming the robustness of the results. Experiment 9 introduced a novel set of nonwords that were partitioned into a high neighborhood density set and a low neighborhood density set, with stimulus onsets controlled across the high and low density sets. In Experiment 9, repetition latency was faster for the low density nonwords once stimulus duration was statistically controlled. In Experiment 10, we adjusted the durations of the stimuli so that they were equated across the high and low neighborhood density groups, and found once again that repetition latency was faster for the low density than for the high density nonwords. Experiments 11 and 12 repeated Experiments 9 and 10 employing the altered presentation regimen, again mirroring those results.

In a number of the experiments we have reported, the pattern of results for high versus low density stimuli was opposite for repetition accuracy versus repetition latency. Thus in Experiment 1, when measuring latency from stimulus onset, and without controlling for stimulus duration, repetition latency was slower for low density than high density stimuli; however, repetition accuracy was greater for low density than high density stimuli (when using the "strict" scoring criterion, although not when using the "lenient" scoring criterion). In Experiment 5, repetition latency was faster for the low density stimuli (when stimulus duration was controlled), whereas repetition accuracy was greater for the high density than the low density stimuli. Could such effects reflect a speed-accuracy tradeoff? This does not appear to be the case, for two reasons.

First, as discussed in detail in the context of Experiment 1, the accuracy results are essentially a function of a subjective rating threshold; given that they can be altered simply by altering the rating threshold, it does not seem appropriate to infer speed-accuracy tradeoffs based on these ratings. As a case in point, in Experiment 1, when repetition accuracy was rescored using a "lenient" scoring criterion, the accuracy difference between high and low density stimuli disappeared, eliminating the apparent speed-accuracy tradeoff.

Second, if speed-accuracy tradeoffs were operative, then we would expect a consistent opposite pattern of results for latency and accuracy. This is not the case. For instance, in Experiment 5, repetition latency was faster for the low density stimuli (when stimulus duration was controlled), whereas repetition accuracy was greater for the high density than for the low density stimuli. However, in Experiment 6, using the low-pass filtered versions of the same stimuli as in Experiment 5, although repetition latency remained faster for the low density stimuli (when stimulus duration was controlled), repetition accuracy was greater for the low density than for the high density stimuli. This inconsistent patterning of results does not appear explicable in terms of a speed-accuracy tradeoff. For these reasons, it does not appear tenable that the results in the present study reflected systematic speed-accuracy tradeoffs.

What, then, are the implications of the present study? In our view, there are three main empirical conclusions to be drawn from the results of these 12 experiments. First, stimulus duration is an important factor to take into account when examining the effect of neighborhood density on nonword repetition latency in an immediate repetition task. In ten of the 12 experiments reported here, the effect of neighborhood density on repetition latency was qualitatively or quantitatively different when stimulus duration was factored out (either through an ANCOVA, or by measuring repetition latency from stimulus offset rather than onset) than when stimulus duration was not factored out (see Table 1). The exceptions were Experiments 10 and 12, in which durational differences between the high and low density stimuli had been controlled by adjusting the stimulus durations themselves; under these circumstances, the effect of density on repetition latency did not change qualitatively when stimulus duration was factored out. These results seem to indicate quite clearly that stimulus duration plays an important role in determining the effect of neighborhood density on nonword repetition latency in an immediate repetition task.

The second conclusion to be drawn from these results is that repetition latency in an immediate repetition task is not always faster for high neighborhood density than for low neighborhood density nonwords, when stimulus duration is controlled. Indeed, a latency advantage for high density nonwords was obtained in only four of the 12 experiments. These (Experiments 1–4) were the experiments employing stimulus tokens that were identical to or derived from those used by Vitevitch and Luce (1998). However, even in these four experiments, the latency advantage for high density stimuli was obtained only when measuring latency from stimulus onset without taking stimulus duration into account. These results indicate that the finding of faster repetition latencies for high density nonwords (Vitevitch & Luce, 1998; the present Experiments 1–4) is not a very robust effect, and is certainly not uniform or exceptionless.

The third empirical conclusion is that, at least in the present experiments, the effect of neighborhood density on repetition latency was predominantly in favor of nonwords from low density neighborhoods. In eight of the twelve experiments reported here, there was a repetition latency advantage for low density nonwords over high density nonwords, when stimulus duration was controlled in an ANCOVA. In the remaining four experiments, there was either no effect of neighborhood density in the ANCOVA (Experiments 2, 3, and 4), or a marginally significant latency advantage for low density stimuli that interacted with stimulus duration. Thus repetition latency appears frequently to be faster for low density than for high density nonwords, in an immediate repetition task.

Turning beyond empirical conclusions, at least two questions come to mind. First, why were the recorded durations of low density nonwords consistently longer than those of high density nonwords? Second, what are the theoretical implications of the finding of faster latencies for low density nonwords? Let us consider each of these questions in turn.

With regard to the durations of stimuli, it can be seen from Table 1 that the low density nonwords had longer mean durations than the high density nonwords for all three sets of stimuli: the original tokens of Vitevitch and Luce (1998) used in the present Experiments 1 and 2; the different speaker recordings of those stimulus types used in the present Experiments 5–8; and the novel stimuli (recorded by a still different speaker) used in the present Experiments 9 and 11. This suggests that the durational difference is not merely an idiosyncratic effect of particular stimulus types, nor of particular stimulus tokens, nor an artifact of a particular speaker. What, then, are we to make of this difference?

We suggest that this finding is a reflection of the fact that nonword stimuli from low density neighborhoods are atypical. There are several aspects to this atypicality. First, low density nonwords are (by definition) similar to fewer real words than are high density nonwords. This could be thought of as a "global" atypicality. Second, neighborhood density is highly correlated with phonotactic probability (Vitevitch & Luce, 1998, 1999), so that higher density nonwords incorporate segments and biphones that are more likely to appear in real words than do low density nonwords; thus low density stimuli are also "locally" atypical. Third, low density nonwords may also be "structurally" atypical. In an analysis of approximately 2000 CVC words of English, Kesler and Treiman (1997) found that certain patterns of relationship between onsets and codas occurred more frequently than would be expected by chance. This was also true of the relationship between vowels and codas, but not of the relationship between onsets and vowels. In other words, of the various possible C_C and _VC (i.e., rime) combinations, there were some that were preferred in the vocabulary. To the extent that low density nonwords incorporate low probability biphones, they may be violating typical within-rime and onset-coda co-occurrence patterns, or in other words, structural relations.

Thus, low density nonwords can be characterized as being atypical at a number of levels. The general notion of such atypicality is corroborated by Vitevitch, Luce, Charles-Luce, and Kemmerer's (1997) finding that low phonotactic density/probability nonwords received poorer ratings as examples of possible English words—in other words, were rated as atypical. But if such stimuli are atypical at multiple levels, then they violate expectations at various levels, and incorporate less familiar transitions at multiple levels; they are therefore likely to take longer to articulate. This would account for the consistently longer durations we observed for the low density nonword stimuli.

We turn now to the second question: What are the theoretical implications of the present results? As noted in the introduction, the theoretical framework proposed by Vitevitch and Luce (1999) was motivated in important measure by the goal of explaining previous findings suggesting that repetition latency in an immediate repetition task is faster for nonwords from high density neighborhoods than from low density neighborhoods. To the extent that repetition latency is not in fact faster for high density than low density nonwords in an immediate repetition task, this poses a puzzle for the original account, and thus has theoretical consequences.

How then might we account for the finding that repetition latencies in the present experiments were for the most part faster for low density stimuli than for high density stimuli? To answer this question, let us return to Vitevitch and Luce's (1999) account and examine whether it could be modified to accommodate the present results. As noted in the introduction, this account proposes that higher neighborhood density leads to slower repetition latencies for real words because lexical level list chunks representing words from higher density neighborhoods have to compete with many other neighboring lexical level list chunks whereas lexical level list chunks representing words from lower density neighborhoods have to compete with only a few neighbors, so that repetition latency is faster for words from lower density neighborhoods. For nonwords, however, there are no lexical level list chunks, so processing occurs primarily at the level of sublexical level list chunks, where phonotactic probability dominates competition, leading to a processing advantage for high density/high phonotactic probability nonwords.

One particular assumption underlying this account is critical to this explanation. This is the assumption that competitive effects for higher density at the sublexical level are relatively weak as compared with the facilitative effect of higher phonotactic probability at that level. This is a crucial assumption: if competition were as strong relative to the effects of phonotactic probability at the sublexical level as it is at the lexical level, then competitive effects and hence slower processing times would be predicted by the framework for high density nonwords, just as they are for high density words, and it would be difficult to account for findings of faster repetition latencies for high density nonwords. However, the present results suggest that higher density nonwords are *not* in fact repeated with shorter latencies, when stimulus duration is controlled. In fact, to the extent that it is lower density nonwords that tend to be repeated with shorter latencies, nonwords are behaving like words: higher density appears to be detrimental. We therefore propose that the balance at the sublexical level between inhibitory competitive effects and facilitative phonotactic probability effects may be more similar to that at the lexical level than was assumed in the original formulation of this account. This would provide a simple explanation of the present findings (because in the revised account we propose, competitive effects are fairly strong at both the lexical and sublexical levels).

The present results thus bear on theoretical accounts of lexical processing, and on how such accounts explain the relationship between lexical and sublexical information. In particular, the present results suggest that a sharp distinction between the strength of competition at lexical and sublexical levels does not hold up, at least not as formulated in Vitevitch and Luce's (1999) account. That account in its original form does not appear able to explain the present results; however, in the revised form we propose, it does.[3] The present results thus serve to test and extend Vitevitch and Luce's theoretical account, and provide information relevant to theories of lexical processing more generally.

Finally, a methodological point highlighted by the present study is that different measures of response latency may yield rather different results. As has been implicit throughout the present analyses, measuring participants' response latency from stimulus onset (rather than from stimulus offset) runs the risk of confounding response latency differences with stimulus duration differences. What, then, is the appropriate bound from which to measure response latency in the investigation of neighborhood density effects: stimulus onset or offset? There does not appear to be any uniformly accepted theoretical basis for preferring one over the other in studies of word/nonword repetition latency in general: some studies have measured latency from stimulus offset (Onishi, Chambers, & Fisher, 2002; Radeau, Morais, & Segui, 1995) while others have measured latency from stimulus onset (Norris, McQueen, & Cutler, 2002; Slowiaczek, McQueen, Soltano, & Lynch, 2000). The present analysis highlights the importance of identifying the effects of stimulus duration. While measuring from stimulus onset runs the risk of confounding response latency differences with stimulus duration differences, measuring from stimulus offset may factor out duration inappropriately, where there is an interaction of density and duration. The safest

[3] Note that an alternative possible revision would also serve the same purpose. This revision relates to the assumption that, when a nonword is presented, there are transient activations of lexical level list chunks that partially overlap with the nonword, but that these activations are very weak, and thus engender little lexical level competition (Vitevitch & Luce, 1999, p. 379). Revising this assumption to allow for such transient activations to be stronger would provide for competition effects for nonwords.

course would appear to be to measure repetition latency from stimulus onset, but additionally conduct an ANCOVA to assess the effect of stimulus duration, as in the present experiments.

Although these considerations are methodological, they do raise empirical (and ultimately, theoretical) questions about the extent to which other reported differences between response times for high versus low density stimuli may have been contaminated by stimulus durations, if they were measured from stimulus onset. Even if stimulus durations were not significantly different in such studies, it is important to analyze the results with stimulus duration controlled, to determine whether the density variable then remains significant, and whether it interacts with duration. Although a meta-analysis of the literature along such lines is beyond the purview of the present article, any studies that employed the same stimuli as in Vitevitch and Luce (1998) potentially incorporated a significant difference between the true stimulus duration of high and low density nonwords. For example, Vitevitch and Luce (1999) Experiment 1 included a same–different judgement task for pairs of auditorily presented nonwords, which employed the same 240 nonword stimuli as in Vitevitch and Luce (1998), and compared reaction time for the high versus low density nonwords in this set. Caution may be appropriate in interpreting the results of any such studies that used these stimuli, and ANCOVAs would be valuable in all these cases.

At a more general level, the methodological considerations raised in this article highlight the difficulty of identifying and controlling all relevant variables when examining the effects of neighborhood properties: after all, studies that employ a neighborhood property as the independent variable are essentially correlational studies. In this vein, recent work by Luce and Large (2001) explicitly addresses phonotactic probability as a variable that is correlated with neighborhood density and attempts to separate the effect of neighborhood density from that of phonotactic probability. The present results suggest that stimulus duration should be added to the list of correlated variables; and they highlight the importance of separating the effects of stimulus duration from those of neighborhood density.

Acknowledgments

We thank Rochelle Newman for recording the stimuli employed in Experiments 9–12, Rebecca Reese for assistance in creation of other stimuli, Steve Luck for many helpful comments, Ryan Bankson, Matt Brown, Dan Kresowik, Sarah Oakley, Emily Stiefel, and Jane Wu for assistance in conducting the present experiments, and Naveen Khetarpal and James Malicki for assistance in scoring and analysis. Thanks are also due to Paul Luce and Mike Vitevitch for providing the stimulus tokens used in Vitevitch and Luce (1998).

References

Boersma, P., & Weenink, D. (2003). Praat: Doing Phonetics by computer [Computer software]. Retrieved from: http://www.fon.hum.uva.nl/praat/.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. D., MacWhinney, B., Provost, J., & Flatt, M. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

Cool Edit [Computer software]. (1996). Phoenix, AZ: Syntrillium Software Corporation.

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior; Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines; Vol. 1, Speech perception* (pp. 187–294). New York: Academic Press.

Grossberg, S., & Stone, G. O. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93*, 46–74.

Grossberg, S., Boardman, L., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 483–503.

Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kesler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language, 37*, 311–395.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altman (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press.

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

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes, 16*, 565–581.

Norris, D., McQueen, J. M., & Cutler, A. (2002). Bias effects in facilitatory phonological priming. *Memory & Cognition, 30*, 399–411.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*, B13–B23.

Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic words. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1297–1311.

Slowiaczek, L. M., McQueen, J. M., Soltano, E. G., & Lynch, M. (2000). Phonological representations in prelexical speech processing: evidence from form-based priming. *Journal of Memory and Language, 43*, 530–560.

Snedecor, G. W., & Cochran, W. G. (1980). *Statistical methods* (7th ed.). Ames, IA: The Iowa State University Press.

SoundEdit 16 [Computer software]. (1995). San Francisco, CA: Macromedia, Inc.

Sound Studio 2.0 [Computer software]. (2002). Wayne, PA: Felt Tip Software.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*, 47–62.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*, 325–329.

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