Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Memory and Language 52 (2005) 193–204

Journal of Memory and Language

www.elsevier.com/locate/jml

Increases in phonotactic probability facilitate spoken nonword repetition[☆]

Michael S. Vitevitch[*], Paul A. Luce

Spoken Language Laboratory, Department of Psychology, 1415 Jayhawk Blvd., University of Kansas, Lawrence, KS 66045, USA
Department of Psychology, University at Buffalo, USA

Received 9 September 2004; revision received 28 October 2004

Abstract

Lipinski and Gupta (2005) report the results of 12 experiments and numerous analyses that attempted to examine further the effects of phonotactic probability originally reported in Vitevitch and Luce (1998, & further explored in Vitevitch & Luce 1999). They suggested that Vitevitch and Luce's results were due to differences in the duration of the stimuli rather than to differences in phonotactic probability. The present report describes the results from another nonword naming experiment—employing a new set of duration-matched stimuli—that demonstrate a facilitative effect of phonotactic probability above and beyond that of stimulus duration. The current results provide support for Vitevitch and Luce's original claims. Possible sources of the discrepancies between Lipinski and Gupta's data and those in the present report are discussed. Although many factors may mediate the facilitative effect of phonotactic probability on nonword repetition latency, we believe there is still sufficient evidence to support the claim that increases in phonotactic probability for nonwords are associated with decreases in nonword processing times, even when stimulus duration is controlled.

Keywords: Phonotactic probability; Neighborhood density; Duration; Naming latency

Introduction

Phonotactic probability refers to the frequency with which phonological segments and sequences of phonological segments occur in words in a given language. *Neighborhood density* refers to the number of words, or neighbors, that are phonologically similar to a given word. Although phonotactic probability and neighborhood density are positively correlated (Vitevitch, Luce, Pisoni, & Auer, 1999), such that common segments and sequences of segments tend to occur in words with many similar sounding neighbors, Vitevitch and Luce (1998; hereafter V&L98) observed what appeared to be a counterintuitive result in an auditory naming (or single-word shadowing) task: nonsense words with high phonotactic probability/dense neighborhoods were responded to more quickly than nonsense words with low phonotactic probability/sparse neighborhoods, whereas real words with high phonotactic probability/

[☆] This research was supported in part by Research Grants R03 DC 04259, R01 006472 (University of Kansas), and R01 DC 0265801 (University at Buffalo) from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health.

[*] Corresponding author.
E-mail address: mvitevit@ku.edu (M.S. Vitevitch).

0749-596X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2004.10.003

dense neighborhoods were responded to more slowly than real words with low phonotactic probability/sparse neighborhoods. This apparent contradiction led V&L98 to hypothesize that the influence of phonotactic probability has a sub-lexical locus, whereas the influence of neighborhood density had a lexical locus. This hypothesis was more extensively examined in a number of other experiments reported in this journal in Vitevitch and Luce (1999).

Lipinski and Gupta (2005; hereafter L&G) report a series of 12 nonword naming experiments that attempted to replicate and further examine the findings reported in V&L98. In Experiment 1 of L&G (2005; also published in Lipinski & Gupta, 2003), the authors claimed to have replicated the results of V&L98 using our original stimuli. However, L&G expressed concern "...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" (p. 177). The remaining experiments in L&G attempt to examine further the role that stimulus duration may play in nonword repetition latency. In Experiment 2, L&G used a digital waveform editor to either compress or expand some (but not all) of the stimuli from V&L98 to equate stimulus duration. In Experiment 3, the same stimuli from Experiment 1 were simply presented at a slower rate than in Experiment 1. In Experiment 4, the compressed/expanded stimuli from Experiment 2 were presented at a slower rate than in Experiment 2. The same stimuli from V&L98 produced by a different speaker were used in Experiments 5–8 along with the procedures used in Experiments 1–4, respectively, and a novel set of stimuli were used in Experiments 9–12 with procedures that were similar to those used in Experiments 1, 2, 5, and 6, respectively.

For each of these experiments, L&G report the results of statistical analyses of: (1) response latency from the onset of the stimulus to the onset of the response and (2) response accuracy (with both lenient and strict criteria) using ANOVA and ANCOVA. In the ANCOVA, stimulus duration was the covariate. L&G also report analyses of response latency that were measured from the stimulus offset to the onset of the response as an alternative means of factoring out the possible effects of stimulus duration on response latency.

In answer to the question, "Does neighborhood density influence repetition latency for nonwords?" L&G conclude the following:

(1) "...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." (L&G, p. 189).
(2) "...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." (L&G, p. 189).
(3) "...in the present experiments, the effect of neighborhood density on repetition latency was predominantly in favor of nonwords from low density neighborhoods (L&G, p. 189).

Although we applaud L&G's considerable effort to further understand the effect of phonotactic probability on nonword repetition latencies, we believe questions remain regarding certain of their conclusions. Below, we address the issues both on empirical and statistical grounds and argue that there is sufficient evidence to conclude that under appropriate circumstances, phonotactic probability has the predicted facilitative effect on nonword processing. Before considering our data in detail, however, we first turn to a point of theoretical and terminological clarification.

Neighborhood density and phonotactic probability

L&G refer to the phonotactic probability effect for the nonword stimuli observed in V&L98 as a reversal of the *neighborhood density* effect. Recall that Vitevitch and Luce, 1998, 1999 hypothesized that the influence of phonotactic probability has a sub-lexical locus, whereas the influence of neighborhood density is primarily at a lexical level. Because nonwords lack lexical representations, and the auditory naming task does not demand that lexical representations be invoked to perform the task efficiently, we postulated that the nonwords were processed at the sub-lexical, not lexical, level. Ergo, we observed an effect of phonotactic probability, not a reversal of the neighborhood density effect.

According to V&L98, only stimuli—either words or nonwords—that activate and resonate with lexical representations will produce a neighborhood density effect. Consider the results from Experiment 3 of Vitevitch and Luce (1999) in which the same nonsense stimuli used in V&L98 were presented in a lexical decision task, a task that requires access to lexical representations for efficient performance in the task. In that case, the nonsense words were responded to much like the real words: nonwords and words with high phonotactic probability/dense neighborhoods were responded to more slowly than nonwords and words with low phonotactic probability/sparse neighborhoods. (For a true reversal of the neighborhood density effect see Dell & Gordon, 2003; Vitevitch, 1997, 2002; Vitevitch & Rodríguez, in press; Vitevitch & Sommers, 2003).

In short, we have not claimed that neighborhood density per se influences repetition latency for nonwords in an immediate nonword naming task. Thus, in what follows, we will refer to the response latency effect observed in the immediate repetition of nonwords reported in V&L98 as an effect of phonotactic probability. Having clarified our terminology, we now turn to a treatment of the three major conclusions put forward by Lipinski and Gupta (2005).

Stimulus duration and phonotactic probability

We agree that reaction times to spoken stimuli in many tasks, including the immediate repetition task, are influenced by stimulus duration. Years of research in our own laboratories, as well as others, have taught us that stimulus duration is a crucial determinant of reaction times. Years of research have also taught us that the duration of naturally produced stimuli is inextricably related to many linguistically relevant characteristics of those items (e.g., Fowler, 1988; Miller, 1981; Pell, 2001). For example, Wright (1979) found that speakers produce low frequency words approximately 24% slower than high frequency words (as measured by word and segment durations). L&G's observations regarding stimulus duration are corroborated by spoken language researchers' experience in dealing with this variable (and with the relationship of stimulus duration to other variables) either by controlling stimulus duration, explicitly manipulating stimulus duration, statistically removing the influence of stimulus duration, or otherwise "correcting" for the influence of stimulus duration. The important issue is not that stimulus durations (and many other so-called "nuisance" variables) affect reaction times. The crucial issue is whether there is an effect of phonotactic probability *above and beyond* the influence of stimulus duration. The fact that we routinely report the durations of our stimuli is testament to our longstanding sensitivity to the issue of stimulus duration influencing response latency.

We must make an important digression at this point. Although we recognize the importance of controlling *stimulus* duration in experiments on spoken word recognition, we unfortunately erred in V&L98 (and in two of the experiments in Vitevitch & Luce, 1999) by reporting durations for the entire sound files (which included leading and trailing silences as well as the stimulus itself) rather than the durations for just the nonword stimuli—the relevant durational variable in these studies. This error, discovered by the current authors only very recently, indeed complicates interpretation of the previous work. Although we believe that there is sufficient subsequent evidence for our original claims (see, for example, the reversal of the probability effect and the results for the bisyllabic stimuli in Vitevitch & Luce (1999); as well as Luce & Large, 2001; and Vitevitch, 2003), we acknowledge that, taken alone, VL98 is problematic. These problems motivated, in part, the experiment reported here, which does indeed confirm the existence of facilitative effects of probabilistic phonotactics on nonword processing.

Returning to the issue at hand, we do not dispute the importance of the influence of stimulus duration on latencies. However, the crucial issue is whether there is an effect of phonotactic probability above and beyond any variability in stimulus duration (or other nuisance variables) that may exist in naturally produced stimuli. To better examine this important question, raised in such detail by L&G, we sought to replicate the results of V&L98 with a novel set of nonword stimuli that varied in phonotactic probability/neighborhood density in a nonword naming task. More important, the novel stimuli used in the present experiment were controlled in *all* aspects of stimulus and file duration. That is, there was no (statistically significant) difference in the duration of silence that preceded the stimuli, in the duration of the stimuli themselves, in the duration of silence that followed the stimuli, and in the duration of the entire stimulus files. We also conducted additional statistical analyses, including ANCOVA and hierarchical multiple regression, to further examine the influence of phonotactic probability on processing above and beyond the influence of stimulus duration.

Experiment 1

Methods

Participants

Thirty native English speakers from the pool of Introductory Psychology students at the University of Kansas participated in partial fulfillment of a course requirement. None of the participants reported a history of speech or hearing problems, nor participated in any of the other experiments reported here.

Stimuli

The stimuli consisted of 60 nonwords varying in phonotactic probability. Half of the items were high in probability and half were low. Phonotactic probability was computed in the same manner as that in Vitevitch and Luce (1998): the sum of the segments and the sum of the sequences of segments in each nonword were calculated and compared. The sum of the segments for high probability nonwords (mean = .167, $SEM = .004$) was significantly greater than the sum of the segments for low probability nonwords (mean = .086, $SEM = .004$; $F(1,58) = 193.57$, $p < .0001$). The sum of the sequences of segments for high probability nonwords (mean = .008, $SEM = .001$) was significantly greater than the sum of the segments for low probability nonwords (mean = .001, $SEM = .0001$; $F\ (1,58) = 48.28$, $p <$ .0001). In addition, frequency-weighted-neighborhood density was computed. As expected given the correlation between phonotactic probability and neighborhood density, the mean frequency-weighted-neighborhood density value for high probability nonwords ($\text{mean} = 38.42$, $SEM = .78$) was significantly greater than the mean frequency-weighted-neighborhood density value for low probability nonwords ($\text{mean} = 9.17$, $SEM = .69$; $F\ (1,58) = 781.84$, $p < .0001$). These values are comparable to the values reported for the stimuli in Vitevitch and Luce (1998, 1999).

The durations of the overall files as well as the durations of the initial silences, the stimuli, and the trailing silences were also controlled. The overall duration for the sound files containing high probability nonwords ($\text{mean} = 625$ ms, $SEM = 8.34$) did not differ from the overall duration for the sound files containing low probability nonwords ($\text{mean} = 626$ ms, $SEM = 8.41$; $F\ (1,58) < 1$). The duration for the initial silence in the sound files containing high probability nonwords ($\text{mean} = 46$ ms, $SEM = 3.04$) did not differ from the duration for the initial silence for the sound files containing low probability nonwords ($\text{mean} = 42$ ms, $SEM = 3.19$; $F\ (1,58) < 1$). The duration of the actual stimulus in the sound files containing high probability nonwords ($\text{mean} = 505$ ms, $SEM = 12.11$) did not differ from the duration of the actual stimulus for the sound files containing low probability nonwords ($\text{mean} = 515$ ms, $SEM = 9.44$; $F(1,58) < 1$). Finally, the duration for the trailing silence in the sound files containing high probability nonwords ($\text{mean} = 74$ ms, $SEM = 6.94$) did not differ from the duration for the trailing silence for the sound files containing low probability nonwords ($\text{mean} = 68$ ms, $SEM = 4.14$; $F\ (1,58) < 1$).

To control for the differential sensitivity of the voice key to different kinds of phonological segments an equal number of nonwords in each condition contained the same initial consonants (3 nonwords each started with /b/, /d/, /f/, /g/, /ʤ/, /m/, /n/, /p/, /r/, /t/). We are therefore confident that any differences we observe in the present experiment are not due to differences in the initial segments found among the nonwords, nor to any differences in duration. The stimuli were spoken in isolation and recorded by the first author in an IAC sound attenuated booth using a high-quality microphone on to digital-audio-tape at a sampling rate of 44.1 kHz. The digital recordings were then transferred directly to hard-drive via an AudioMedia III card and Pro Tools LE software (both made by Digidesign), and edited into individual digital files (16 bit) that were stored on computer disk for later playback.

Procedure

Participants were tested one at a time. Each participant was seated in a booth equipped with an iMac running PsyScope 1.2.2 (Cohen, MacWhinney, Flatt, & Provost, 1993) that controlled stimulus randomization and presentation, a set of Beyerdynamic DT-109 headphones, and a PsyScope button box with a dedicated timing board. A trial proceeded as follows: the word "READY" appeared in the center of the computer screen for 500 ms to indicate the beginning of a trial. Participants were then presented with one of the randomly selected stimuli at approximately 70 dB SPL over the headphones. Response latency was measured from the onset of the stimulus file to the onset of the participant's response. Recall that there was no significant difference in the amount of silence leading up to the stimulus contained in the sound files (nor in the duration of the stimuli themselves), therefore the addition of (essentially) a constant value to the measurement of response latency will most likely not distort the distribution of response latencies. When a response was made, the word "READY" appeared on the screen, and the next trial began. If a response was not registered, 5 s elapsed before the word "READY" appeared on the screen, and the next trial began. The method of stimulus presentation used in the present experiment (as well as V&L98) contrasts with that employed by L&G in which an inter-trial-interval—defined by L&G (e.g., p. 174) as the time from the offset of the present nonword to the onset of the next nonword—of either 1000 ms (Experiments 1, 2, 5, 6, 9, and 10) or 4500 ms (Experiments 3, 4, 7, 8, 11, and 12) elapsed before the next stimulus was presented. Responses were also recorded on digital-audio-tape for later accuracy analyses.

Prior to the experimental trials, each participant received 10 practice trials. As in V&L98, the stimuli in the practice trials were nonwords from a different and unrelated experiment. None of the items used in the practice session were used in the experiment. The practice trials were used to familiarize the participants with the task, and the data collected from them were not included in the final analysis. The practice trials were also used to familiarize the participants with how loudly they needed to speak to trigger a response via the voice key. In contrast to the methodology employed by L&G, the sensitivity level on the voice key in the present experiment was set at a level that required approximately 70 dB SPL (approximately the level of normal conversational speech) to trigger a response as measured by a Casella CEL-254 sound level meter placed in the same position (approximately 1 cm from the lips) as the headphone mounted microphone used to trigger a response. (Note that L&G adjusted the sensitivity level individually to each participant, p. 174) Participants were instructed to respond as quickly and as accurately as possible.

Results

We conducted ANOVAs on both reaction times and accuracy with both participants and items as random factors. We also computed estimates of effect size, PV, or the proportion of variance explained by the dependent variable (Murphy & Myors, 1998). For reference, $PV = .01$ is considered a small effect, $PV = .10$ is considered a medium effect, and $PV = .25$ is considered a large effect. Finally, to control statistically for stimulus duration, we performed both ANCOVA and hierarchical multiple regression.

A trained speech scientist used linguistic conventions to score the tape-recorded responses of each participant for accuracy. Only accurate responses were included in our analyses of response latency. The same criterion used to determine accuracy that was used in Vitevitch, Luce, Charles-Luce, and Kemmerer (1997), and Vitevitch and Luce (1998, 1999) was used in the present experiment. Responses due to the improper triggering of the voice key (e.g., cough, "uh", etc.) were not included in the present analyses. In addition,

> [a]ccuracy was assessed by listening to the participants' responses and comparing them to a written transcription of the stimuli. A response was scored as correct only if there was a match on all segments of the stimulus. (V&L98, p. 327).

It is important to note that in V&L98, we: (1) made careful use of the term "segment" (see Crystal, 1992), (2) did not include diacritics in the transcriptions of examples of the stimuli (and in the stimuli listed in the appendix of Vitevitch & Luce (1999), with the exception of the rhoticity sign on the vowel /ɝ/), and (3) used angled rather than square brackets in the examples of the stimuli presented in V&L98. The use of such terminology and accepted notation (as per linguistic conventions) was meant to underscore that our criterion for an accurate response was based on broad phonetic (i.e., phonemic) transcriptions, not narrow phonetic transcriptions.

In contrast L&G used naïve listeners who employed either a "lenient" or a "strict" criterion to determine response accuracy. However, L&G fail to explicitly define either term. Did all three phonemes in the nonword have to be produced correctly with the "strict" criterion, whereas two phonemes (or perhaps just one phoneme) in the nonword had to be produced correctly with the "lenient" criterion? Did the "strict" criterion rely on a comparison between narrow phonetic transcriptions of the stimuli and responses, whereas the "lenient" criterion relied on a comparison between broad phonetic (i.e., phonemic) transcriptions of the stimuli and responses? Did L&G use one of these methods of assessing the accuracy of a response, which are quite commonly used in spoken language research, or did they develop *an alternative* method of assessing response accuracy? In any case, the observation that changing the scoring criteria (e.g., from one based on broad to narrow transcriptions) will yield different levels of accurate performance (see L&G, p. 176) is not novel, nor particularly surprising. More important, the failure to clearly define these terms makes the interpretation of the accuracy results in Experiments 1–12 in L&G difficult at best.

In the present experiment we obtained a large, significant effect of phonotactic probability found when response latency was measured from the onset of the stimulus file to the onset of the response ($t_1(29) = -6.77, p < .0001, PV = .61$). Participants repeated nonwords with high phonotactic probability more quickly (801 ms; $SEM = 15.09$) than nonwords with low phonotactic probability (815 ms; $SEM = 15.55$). This effect was also significant when items were treated as a random variable: t_2 (58) = −2.48, $p < .05$. There was no difference in overall error rates between the two sets of nonwords (both $t < 1$), suggesting that participants did not sacrifice speed for accuracy in making their responses. Nonwords with high phonotactic probability were correctly repeated 95.2% ($SEM = .013$) of the time and nonwords with low phonotactic probability were correctly repeated 95.0% ($SEM = .013$) of the time.

To examine further the influence of phonotactic probability on processing above and beyond the influence of stimulus duration, we attempted to statistically remove the variability associated with stimulus duration. Recall that the stimulus durations for the two groups of nonwords in the present experiment did not differ statistically. Nonetheless, these stimuli, like all naturally produced stimuli, did exhibit some amount of variability in duration. Thus, we wished to determine if there was an influence of phonotactic probability on processing above and beyond the influence of stimulus duration.

To be consistent with the analyses reported in L&G, we report the results of an ANCOVA on the items. (Note that the ANCOVAs here and in L&G are based on items alone and do not take subject variability into account. Caution should be exercised in drawing strong conclusions based on these analyses alone.) In contrast to L&G, who conducted ANCOVA on response latency for all responses whether the response was correct or incorrect, the present ANCOVA for response latency included only items that were responded to correctly. Although our ANCOVA analyses do not violate the assumption of homogeneity of regression (cf., L&G),[1]

---

[1] Violations of the assumption of homogeneity of regression are evident when there is a statistically significant interaction between the independent variable and the covariate (using a conservative criterion, such as $\alpha = .25$). Violations of homogeneity of regression can also be detected by observing unequal slopes in the regression lines for each of the different conditions (compare the slopes of the regression lines in the two figures in Appendix). Tabachnick and Fidell (1989, p. 325) strongly caution us that: *If any other design* (i.e., other than a between-subjects design) *is used, and interaction between IVs and covariates is suspected, ANCOVA is inappropriate* (emphasis in original). Recall that phonotactic probability was manipulated within-subjects, not between-subjects.

we also report additional analyses, as suggested by Cohen and Cohen (1983), employing hierarchical regression as an alternative means of assessing the influence of phonotactic probability above and beyond the influence of stimulus duration.

In the present ANCOVA analysis, as in the ANCOVA analyses performed by L&G, stimulus duration was the covariate, phonotactic probability was the categorical variable, and "corrected" response latencies (response latencies with the leading silence of each stimulus item subtracted from them) constituted the dependent variable. To evaluate whether the assumption of homogeneity of regression had been violated, we first conducted an ANCOVA incorporating the covariate, the categorical variable, and the interaction of the two variables. If the interaction is not significant (using a conservative criterion, such as $\alpha = .25$), then the assumption of homogeneity of regression has not been violated, and we can remove the interaction term and proceed with the ANCOVA analysis. The interaction term for the ANCOVA was not significant using the conservative $\alpha = .25$ criterion ($F\ (1,56) = .02,\ p = .89$). The subsequent ANCOVA analysis (without the interaction term) showed a significant effect of stimulus duration ($F\ (1,57) = 33.67,\ p < .0001$). Given that these stimuli were produced naturally, it is perhaps not surprising that they vary in duration. Furthermore, given the well-known influence of stimulus duration on response latency, it is also perhaps not surprising that stimulus duration should significantly influence response latency. The crucial issue is whether there is an influence of phonotactic probability above and beyond the influence of stimulus duration on response latency. The present ANCOVA analysis indeed showed a significant effect of phonotactic probability ($F\ (1,57) = 7.83,\ p < .01$), even when stimulus duration was statistically taken into account.

As suggested by Cohen and Cohen (1983), we employed hierarchical regression as an alternative means to assess the influence of phonotactic probability above and beyond the influence of stimulus duration. In this analysis, stimulus duration and phonotactic probability constituted the independent variables, with stimulus duration entered first. The results of the hierarchical regression analysis show that stimulus duration, entered into the regression equation first, produces a semipartial correlation of .601 ($F(1,59) = 32.87, p < .0001$), accounting for approximately 36% of the variance in the dependent variable (corrected reaction times as used in the ANCOVA above). More important, the results of the regression revealed a significant semipartial correlation of .348 for phonotactic probability, accounting for approximately 12% of the variance in the dependent variable (corrected reaction times) above and beyond the influence of stimulus duration ($F\ (2,59) = 22.87$, $p < .0001$). As we (and many others) have acknowledged before, stimulus duration is an important factor that predicts response latency. However, what is more important is that another independent variable, namely phonotactic probability, predicts response latency above and beyond the influence of stimulus duration, suggesting that phonotactic probability is an important factor that influences spoken language processing.

Why, then, are there such striking discrepancies between our results and those of L&G? Although there are a number of issues that bear on this question, one reason for the discrepancies in results may concern methodology. In the current experiment, participants were granted a comfortable amount of time to respond. Although the reaction times clearly show that our participants responded quickly, they were nonetheless able to control, to a certain extent, the pace of the experiment, thus being able to respond with high levels of accuracy. In Experiments 1, 2, 5, 6, 9, and 10 of L&G, the pace of the experiment was much more rapid: subjects had only 1 s to respond from the offset of the stimulus to the onset of the next stimulus. Our own phenomenal experience of a similarly designed experiment using the original V&L98 stimuli was one of frustration in attempting to respond to sometimes quite difficult nonwords. Indeed, we found ourselves often responding to a given stimulus while the next one was being presented. Our sense is that, if the conditions of our trial experiments conducted in both Kansas and New York approximate those of L&G, participants in certain of the L&G studies may have experienced considerable difficulty in responding, affecting accuracy and adding considerable noise to the data, thereby making it potentially difficult to obtain the original (admittedly small) facilitative effect. Evidence for these speculations can be seen in the accuracies of L&G's experiments employing the stimuli from V&L98 and a "fast" presentation rate (e.g., Experiments 1 and 2), which were low for experiments in which participants were required to simply repeat back stimuli presented in the clear. (We leave aside the issue of scoring "leniency" for the moment.)

To garner some empirical support for this speculation, we reran Experiment 1 using a modified version of the L&G script. In L&G's Experiments 1, 2, 5, 6, 9, and 10 subjects were allowed 1 s to respond after the offset of the sound file and the presentation of the next stimulus. In the current experiment, we also allowed subjects 1 s to respond before the next stimulus was presented. Does the pace of the experiment contribute to our ability to detect the small facilitative effect of phonotactic probability?

Experiment 2

Methods

Participants

Thirty-four native English speakers from the pool of Introductory Psychology students at the University of Kansas participated in partial fulfillment of a course requirement. None of the participants reported a history of speech or hearing problems, nor participated in any of the other experiments reported here.

Stimuli

The same stimuli used in Experiment 1 were used in the present experiment. Recall that these items varied in phonotactic probability/neighborhood density, but were controlled with regard to the duration of the initial silence, the stimulus, the trailing silence, and the overall file.

Procedure

The same equipment and procedure used in Experiment 1 were used in the present experiment, with the following exception. The PsyScope script used in the present experiment was modified to approximate the method of presentation employed by L&G (see p. 174). In the PsyScope script in the present experiment, a stimulus item was presented to a participant. Immediately after the offset of the nonword stimulus file, a fixation cross was presented for 500 ms signaling the end of the previous sound file and the presentation of the next sound file. As in L&G, the inter-trial-interval, or the time between the offset of the present stimulus and the onset of the next stimulus, was 1000 ms. This method of presentation afforded participants 1 s to respond before the presentation of the next stimulus item, similar to the method employed by L&G. Response latency was measured from the onset of the sound file to the onset of the response. The responses in the present experiment were also recorded on digital-audio-tape for later accuracy analyses.

Results

A trained speech scientist scored the tape-recorded responses of each participant for accuracy. Only correct responses were included in our analyses of response latency. The same criterion used to determine accuracy in Experiment 1 was used in the present experiment.

Although a significant effect of phonotactic probability was found with these stimuli in Experiment 1, we failed to find a significant difference in response latency in the present experiment in which the pace of the experiment was much more rapid than the presentation pace of Experiment 1 (t (33) = −.045, $p = .96$). Note that the present experiment had a slightly larger sample size ($n = 34$) than that used in Experiment 1 ($n = 30$), which should have made the present analysis more powerful than the analysis in Experiment 1. Participants repeated nonwords with high phonotactic probability with a mean of 867 ms ($SEM = 15.61$), and nonwords with low phonotactic probability with a mean of 868 ms ($SEM = 17.39$). There was also no difference in overall error rates between the two sets of nonwords (t (33) = .922, $p = .36$). Nonwords with high phonotactic probability were correctly repeated 96% ($SEM = .007$) of the time and nonwords with low phonotactic probability were correctly repeated 93% ($SEM = .029$) of the time.

As in Experiment 1 we further examined the influence of phonotactic probability on processing above and beyond the influence of stimulus duration, using ANCOVA and hierarchical regression. In the present ANCOVA analysis, stimulus duration was the covariate and phonotactic probability was the categorical variable. To evaluate whether the assumption of homogeneity of regression has been violated, we first conducted an ANCOVA incorporating the covariate, the categorical variable, and the interaction of the two variables. If the interaction is not significant (using a conservative criterion, such as $\alpha = .25$), then the assumption of homogeneity of regression has not been violated, and we can remove the interaction term and proceed with the ANCOVA analysis. The interaction term for the ANCOVA was $F\,(1, 56) =\ 2.45$, $p = .13$. Although this interaction is not considered to be significant using the traditional criterion of $\alpha = .05$, the interaction is considered significant using the conservative criterion of $\alpha = .25$, suggesting that the assumption of homogeneity of regression may have been violated. Visual inspection of the bottom figure provided in the appendix verifies that the assumption of homogeneity of regression has been violated, meaning we cannot properly interpret the results of an analysis employing ANCOVA.

As suggested by Cohen and Cohen (1983), when the assumption of homogeneity of regression has been violated, hierarchical regression may be employed as an alternative means to assess the influence of phonotactic probability above and beyond the influence of stimulus duration. In this analysis, stimulus duration and phonotactic probability constituted the independent variables, with stimulus duration entered first. The results of the hierarchical regression analysis show that stimulus duration, entered into the regression equation first, produces a semipartial correlation of. 669 ($F\,(1, 59) = 47.06$, $p < .0001$), accounting for approximately 45% of the variance in the dependent variable (corrected reaction times). However, in the present experiment in which the pace of the experiment was much more rapid than the presentation pace of Experiment 1, phonotactic probability had a semipartial correlation of .056, accounting for less than 1% of the variance in the dependent variable (corrected reaction times) above and beyond the influence of stimulus duration. As we hypothesized, the pace of the experiment may indeed contribute to our ability to detect the small facilitative effect of phonotactic probability.[2]

Discussion

Clearly, modifying the pace of the experiment eliminated the effect observed in Experiment 1. Whereas it may well be that the effect of phonotactic probability on naming times is a fragile one that is dependent on subtle variations in presentation timing, it may also be the case that requirements to respond quickly provoke a different analysis of the input (see, e.g., McLennan & Luce, in press; McLennan, Luce, & Charles-Luce, 2003), cause less accurate responding, and add considerable noise that complicates detection of a small effect.

What, then, do we make of L&G's experiments that used a slower pace allowing participants 4.5 s to respond before the next stimulus was presented? Unfortunately, we now encounter another incommensurate set of data in which reaction times to the stimuli are considerably longer than those obtained in V&L98 or in the present experiments. Given research by Newman, Sawusch, and Luce (1997) demonstrating fairly well-defined and short time windows for effects of neighborhood density on phoneme processing, it is not unreasonable to conclude that responses as long as those reported by L&G in Experiments 3, 4, 7, 8, 11, and 12 reflect processes that are well downstream from what are probably fast-acting effects of sub-lexical probabilities. The same logic has also been used to partially account for attenuated influences of other variables, such as word-frequency, in the delayed naming task (see Goldinger, Azuma, Abramson, & Jain, 1997). In short, it appears that the details of timing—in both presentation and their concomitant effects on responses—may matter in detecting subtle and small effects of phonotactic probabilities.

General discussion

L&G (p. 188) state: "...there was no effect of neighborhood density on repetition latency for the nonwords when stimulus duration was statistically controlled." Despite the claims of L&G, the present results suggest that there is indeed valid and demonstrable evidence that phonotactic probability has an affect on processing when stimulus duration is controlled, either statistically or methodologically.

In Experiment 1 a novel set of carefully controlled nonwords was presented in an immediate nonword repetition task. Recall that the high and low probability nonwords had durations of the silence leading to the stimulus, the stimulus itself, the silence trailing from the stimulus, and the entire sound file that were not statistically different. A simple analysis of the response latencies from the immediate nonword repetition task showed a significant effect of phonotactic probability, such that high probability nonwords were responded to more quickly than low probability nonwords. Additional analyses employing ANCOVA and hierarchical regression took stimulus duration into account and further demonstrated that phonotactic probability affects response latency in an immediate nonword repetition task *above and beyond* the influence of stimulus duration.

Why, then, did L&G fail to observe a facilitative influence of phonotactic probability in the majority of experiments they conducted employing an immediate nonword repetition task? Although there are many possible answers to this question, the results from Experiment 2 in the present report suggest that timing might be an issue. When a rapid rate of stimulus presentation was used in the present Experiment 2—similar to the rate of presentation used in Experiments 1, 2, 5, 6, 9, and 10 of L&G—responding may have been relatively difficult, especially before the next stimulus was presented. The difficulty associated with simply responding at such a rapid rate of presentation in the immediate nonword repetition task may have introduced increased variability in the response latencies, thereby obscuring the (small) influence of phonotactic probability on response latency. In contrast, the presentation rate in Experiments 3, 4, 7, 8, 11, and 12 of L&G may have been too slow. The response latencies observed with a slower presentation rate may reflect processes that are well downstream from what are probably fast-acting effects of sub-lexical probabilities, thereby making it difficult to observe effects of phonotactic probability on response latency.

Another factor that may have made it difficult for L&G to observe the effect of phonotactic probability on response latency was their practice of including correct and incorrect responses in their ANCOVA analyses of response latency (see p. 176). It is generally assumed that the response latency associated with the production

---

[2] Four participants spontaneously commented on the rapid pace of stimulus presentation, corroborating our phenomenal experience regarding the pace of stimulus presentation. Stimulus presentation was so rapid that seven participants in the present experiment were not able to respond to all of the stimuli that they were presented with before the next stimulus was presented (high probability stimuli: mean = 4, range 1–9; low probability stimuli: mean = 2, range 0–6). The rapid pace of stimulus presentation could also result in the middle of the response to trial X being recorded as the onset of the response to trial $X+1$. An analysis of the data from the slowest participant in this experiment shows that significantly shorter response latencies (mean = 269 ms; range: 26–938 ms) tended to follow long response latencies (of 1250 ms or longer; $t(4) = 5.57$, $p < .01$). Thus, it is quite possible that the rapid rate of stimulus presentation could have obviated detection of the effects of phonotactic probability that were obtained in Experiment 1.

of the response/m^b/tells us something about the processes that were involved in the comprehension of the stimulus item /m^b/. It is unclear, however, what the response latency associated with the production of the response /b^b/ (an example of an erroneous response) tells us about the processes that were involved in the comprehension of the stimulus item /m^b/ (see also Levelt, Roelofs, & Meyer, 1999 for another discussion of what errors tell us about normal language processing). Indeed, it has long been suggested that errors and correct responses come from different distributions (e.g., Hale, 1968; Luce, 1986; Ratcliff & Rouder, 1998; see also Gehring, Goss, Coles, Meyer, & Donchin, 1993 for electrophysiological differences between errors and correct responses). Therefore, the ANCOVA analyses of response latency that include both correct and incorrect responses may have increased the variability in response latency and obscured the admittedly small influence of phonotactic probability on response latency, making it difficult to clearly interpret the results of such analyses.

It should also be noted that the results of statistical analyses obtained when the basic assumptions of the statistic have been violated are difficult to clearly interpret as well. Note that in Experiment 2 in the present report, a violation in the assumption of homogeneity of regression was observed, making an ANCOVA analysis of the data inappropriate. Had we proceeded with the ANCOVA analysis despite the violation of the assumption of homogeneity of regression, the obtained result would have made it difficult to interpret properly the influence of phonotactic probability above and beyond the influence of stimulus duration. As suggested by Cohen and Cohen (1983), hierarchical regression can be used in such cases as an alternative means of assessing the influence of phonotactic probability *after* the effect of stimulus duration has been removed. Our analysis in Experiment 1 employing hierarchical regression (as well as ANCOVA) clearly demonstrates that phonotactic probability does influence response latency in an immediate nonword repetition task above and beyond the influence of stimulus duration. Because L&G did not include regression plots for any of their experiments to allow for visual inspection of the regression lines, it is unclear whether the assumption of homogeneity of regression was violated in the ANCOVA analyses performed by L&G. Analyses in which the statistical assumptions have been violated make the results of those analyses difficult to interpret. Interpretation of the results of L&G is also made difficult by the inclusion of incorrect as well as correct responses in their (ANCOVA) analyses. Even the (ANOVA) analyses by L&G that examined only correct responses are difficult to interpret because the criterion used to score a response as accurate or incorrect (i.e., "lenient" and "strict") were never defined.

The influence of phonotactic probability has been observed in many other studies using different tasks, different stimulus materials, different stimulus tokens, different dependent measures, different populations, and different measures of phonotactic probability (e.g., Bailey & Hahn, 2001; Fritch, Large, & Pisoni, 2000; Gathercole, Frankish, Pickering, & Peaker, 1999; Gaygen, 1998; Jusczyk, Luce, & Charles-Luce, 1994; Luce & Large, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999; McLennan, Luce, & La Vigne, 2004; Pitt & McQueen, 1998; Pylkkänen, Stringfellow, & Marantz, 2002; Storkel, 2001; Storkel & Rogers, 2000; Vitevitch, 2003; Vitevitch, Armbrüster, & Chu, 2004; Vitevitch et al., 1997; Vitevitch, Pisoni, Kirk, Hay-McCutcheon, & Yount, 2002).[3] Recently, McLennan et al. (2004) reported a facilitative effect of probabilistic phonotactics on nonword repetition in a learning paradigm, in which phonotactic probability was manipulated directly during training by selective repetition of segments and patterns. Because all nonword items served as their own controls, effects of phonotactics could not be attributed to stimulus duration. McLennan et al. demonstrated that increases in phonotactic probability during training speeded nonword repetition at test (relative to the exact same items that had not been trained). In short, we believe that the effect of phonotactic probability originally reported by V&L98 is subtle, dependent on crucial aspects of the timing of the presentation of the stimuli as well as the speed of the response itself. This may point to an effect whose demonstration may require careful experimental and stimulus control, however, it nonetheless does not undermine its theoretical importance, nor the theoretical account described in Vitevitch and Luce

---

[3] Note that the nonwords that were used in the nonword naming task of Vitevitch and Luce (1998) were also used in a same-different matching task (Experiments 1 and 2) and a lexical decision task (Experiment 3) in Vitevitch and Luce (1999). A subset of these items was also used in a same-different matching task (Experiment 1) and a lexical decision task (Experiment 2) in Vitevitch et al. (2002). The pattern of results from these two tasks obtained in Vitevitch et al. in a group of cochlear implant users was similar to the pattern of results obtained in the same tasks in normal hearing adults (Vitevitch & Luce, 1999). In the same-different tasks (as in the nonword naming task), high probability nonwords were responded to more quickly than low probability nonwords. However, in the lexical decision tasks, high probability nonwords were responded to more slowly than low probability nonwords. It is unclear how a difference in stimulus duration alone, as hypothesized by L&G, could explain why the pattern of results obtained in the nonword naming and same-different matching tasks was the opposite of the pattern of results obtained for the nonwords in the lexical decision tasks. In contrast, the hypothesis originally put forward in Vitevitch and Luce (1999)—the amount of activation in lexical and sub-lexical representations is a function of task demand—readily accounts for the pattern of results observed across tasks.

(1998, 1999). We grant that L&G have helped to define some of the circumstances under which the facilitative effect of probabilistic phonotactics is not obtained.

Appendix

Fig. 1.

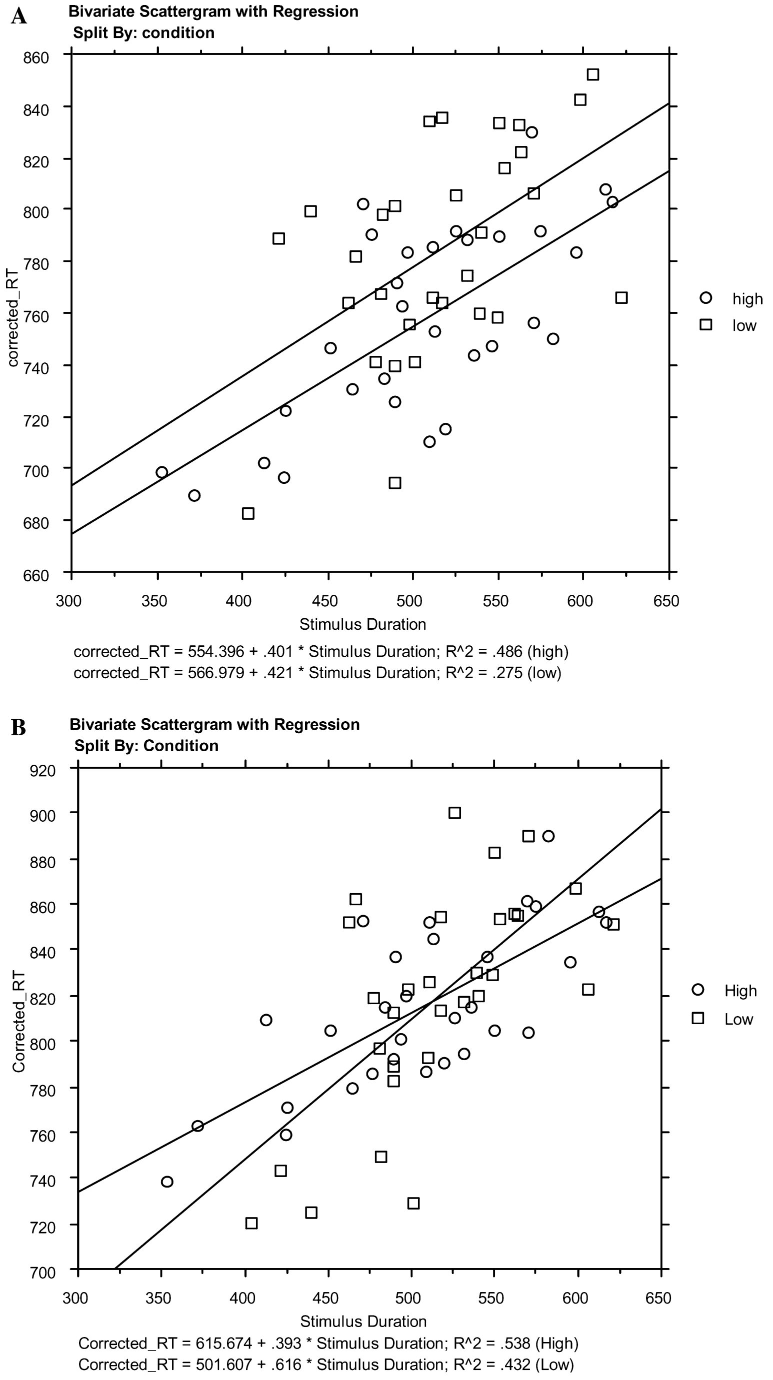

Fig. 1. (A) Slopes of the regression lines from Experiment 1; these data do not violate the assumption of homogeneity of regression. (B) Slopes of the regression lines from Experiment 2; these data do violate the assumption of homogeneity of regression.

References

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods?. *Journal of Memory and Language, 44*, 568–591.

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

Crystal, D. (1992). *An encyclopedic dictionary of language and languages*. London: Blackwell.

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities*. Berlin: Mouton de Gruyter.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 31*, 307–319.

Fritch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42*, 481–496.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 84–95.

Gaygen, D. E. (1998). *Effects of phonotactic probability on the recognition of words in continuous speech*. Unpublished doctoral dissertation, Univ. at Buffalo, Buffalo, NY.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error-detection and compensation. *Psychological Science, 4*, 385–390.

Goldinger, S. D., Azuma, T., Abramson, M., & Jain, P. (1997). Open wide and say "Blah!" Attentional dynamics of delayed naming. *Journal of Memory and Language, 37*, 190–216.

Hale, D. (1968). The relation of correct and error responses in a serial choice reaction task. *Psychonomic Science, 13*, 299–300.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*, 630–645.

Lipinski, J., & Gupta, P. (2003). Separating the effects of duration and neighborhood density in nonword repetition latency. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual conference of the cognitive science society* (p. 718). Hillsdale, NJ: Lawrence Erlbaum.

Lipinski, J., & Gupta, P. (2005). Does neighborhood density influence repetition latency for nonwords? Separating the effects of density and duration. *Journal of Memory and Language, 52*, 171–192.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Brain and Behavioral Science, 22*, 1–75.

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

Luce, R. D. (1986). *Response times*. New York: Oxford University Press.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465–494.

McLennan, C., & Luce, P.A. (in press). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 539–553.

McLennan, C. T., Luce, P. A., & La Vigne, R. (2004). *The role of non-adjacent phonotactic dependencies in the perception of spoken language*. Presentation at the 45th Annual Meeting of the Psychonomic Society, Minneapolis, MN, November, 2004.

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. NJ: Erlbaum Associates.

Murphy, K. R., & Myors, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Mahwah, NJ: Erlbaum.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Effects of lexical neighborhood density on phoneme perception. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 873–889.

Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *Journal of the Acoustical Society of America, 109*, 1668–1680.

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

Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language, 81*, 666–678.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*, 347–362.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321–1337.

Storkel, H. L., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics and Phonetics, 14*, 407–425.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*. Harper Collins Publishers.

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech, 40*, 211–228.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 735–747.

Vitevitch, M. S. (2003). The influence of sublexical and lexical representations on the processing of spoken words in English. *Clinical Linguistics and Phonetics, 17*, 487–499.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sub-lexical and lexical representations in speech production: Effects of phonotactic probability and onset-density. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 514–529.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *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.

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., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*, 306–311.

Vitevitch, M. S., Pisoni, D. B., Kirk, K. I., Hay-McCutcheon, M., & Yount, S. L. (2002). Effects of phonotactic probabilities on the processing of spoken words and nonwords by postlingually deafened adults with cochlear implants. *Volta Review, 102*, 283–302.

Vitevitch, M. S., & Rodríguez, E. (in press). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders*.

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition, 31*, 491–504.

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition, 7*, 411–419.