Frequency, Neighborhood Density, and Phonological Similarity Effects in Picture Naming: An Artificial Lexicon Study

Austin F. Frank (afrank@bcs.rochester.edu), Michael K. Tanenhaus, Richard N. Aslin, Anne Pier Salverda
Department of Brain and Cognitive Sciences, University of Rochester

Abstract

Subjects learned to name novel shapes using words from an artificial lexicon. Use of an artificial lexicon allowed for tight control over properties of the words in the lexicon, including lexical frequency, neighborhood density, and phonological similarity. After training, subjects' naming latencies and error rates displayed some of the same performance patterns as would be expected in natural language studies of picture naming. These encouraging results argue for the consideration of artificial lexicons as research tools in studies of production, and offer exciting possibilities for future work in language processing.
Keywords: *picture naming, word production, word learning, artificial lexicon, lexical representation, phonological neighborhoods, frequency effects*

In both comprehension and production, certain properties of words are known to affect the way those words are processed. Some of these properties include lexical frequency, phonological similarity to recently used words, and phonological neighborhood density.

Experimental evidence shows that frequency of use and phonological structure play an important role in how a word is processed, in both comprehension and production. Listeners recognize high-frequency words more quickly than low-frequency words (Marslen-Wilson, 1987). Speakers produce high-frequency words more proficiently, as evidenced by shorter naming latencies and fewer errors (Jescheniak & Levelt, 1994; Levelt et al., 1991; Dell, 1986).

While experience with a word consistently facilitates its processing, the impact of a word's phonology on its use is more varied. Exposure to an auditorily-presented phonologically related word during lexical access leads to faster naming performance (Meyer & van der Meulen, 2000; Levelt et al., 1991); but producing a phonologically-related word prior to a naming a picture slows speakers down (Wheeldon, 2003). Phonologically related words need not even be explicitly presented as primes in order to affect lexical access.

Words that are phonologically similar are referred to as neighbors [1] (Luce & Pisoni, 1998). Words with many neighbors are said to come from *dense neighborhoods*, and words with few neighbors from *sparse neighborhoods*. The density of a word's phonological neighborhood affects its processing, even when its neighbors are not explicitly involved in the language task. In comprehension, for example, words from dense neighborhoods are systematically recognized more slowly than words from sparse neighborhoods (Vitevitch & Luce, 1998).

The effects of neighborhood density on production are inconsistent. Vitevitch and colleagues have shown a facilitatory effect of neighborhood density on the production of English words (Vitevitch, 2002; Vitevitch, Armbruster, & Chu, 2004; Vitevitch & Sommers, 2003). The same researchers, however, have found the opposite effect in Spanish. Spanish speakers produced words from dense neighborhoods more slowly than words from sparse neighborhoods (Vitevitch & Stamer, 2006). Additionally, the results of specific experiments have sometimes been difficult to replicate. Vitevitch & Sommers, 2003 show fewer tip-of-the-tongue states for words from dense neighborhoods, but fail to find any effect of neighborhood density in picture naming. This stands in contrast to Vitevitch, 2002 and Vitevitch et al., 2004, where naming latencies were shorter for words from dense neighborhoods.

In comprehension, the inhibitory influence of neighborhood density on recognition has been attributed to phonologically-based competition during lexical access (Mcclelland & Elman, 1986). A similar process could be involved in production. On the other hand, an interactive activation model of production has been used to demonstrate that shared phonology among neighbors should lead to faster responses and fewer errors for words from dense neighborhoods (Dell & Gordon, 2003). At this point, the results from production studies cannot fully explain the role that phonological similarity plays in lexical access for single word production.

One possible confound in studies of neighborhood effects is that neighborhood density is correlated with other distributional properties in the lexicon (*e.g.* phonotactic probabilities; see Vitevitch, Luce, Pisoni, & Auer, 1999). When controlling for correlated factors and trying to sample the lexicon at different levels of other lexical properties (neighborhood density, lexical frequency), it can be difficult to gather an ideally-balanced set of stimuli for studying the interactions among lexical properties in language processing. Furthermore, while current methods of estimating frequency and neighborhood density are useful, these estimates fail to account for important factors like individual experience with language. Citing these reasons and others, Magnuson et al. conducted a series of experiments based on an *artificial lexicon* (2003). Over the course of several days, subjects learned to recognize novel shapes by their "names". Drawing the names of the shapes from a specially designed artificial lexicon allowed the researchers to achieve precise control over theoretically interesting factors such as lexical frequency and neighborhood

[1] In this study and most of the studies presented below, phonological similarity is heuristically defined as a pair of words that differ by no more than one added, altered, or deleted phoneme. See Luce & Pisoni, 1998 for more on this "shortcut rule".

Table 1: Artificial Lexicon

HF-HD	*HF-LD*	*LF-HD*	*LF-LD*
bago	bugo	bagi	baki
dopi	doti	dupi	dupo
gupi	gobi	gopi	gupa
kagi	kigo	kago	kati
pibu	paku	piku	piba
tiku	tibu	toku	tibo

density. Furthermore, the use of novel words meant that, in general, subjects all started and ended with the same amount of experience with each of the words in the lexicon. Use of an artificial lexicon in this study allowed new insight into spoken word recognition.

The present study is an attempt at extending this artificial lexicon methodology to a word production task. Other researchers have used nonsense words to look at issues of frequency and phonological similarity, but it seems likely that lexical processing is importantly different from the processing of arbitrary sequences of phonemes (Lipinski & Gupta, 2005; Cholin, Levelt, & Schiller, 2006; Dell, Reed, Adams, & Meyer, 2000). In this sense, the important distinction is between the use of non-words and novel words. Given that this study requires that subjects learn to use the words in a referential domain, we hope that the tokens that begin as nonsense words will be lexicalized during the course of training. In this experiment we examine the interaction of frequency, neighborhood density, and phonological similarity in the production of novel words. The goal of this study is to determine whether words learned from an artificial lexicon come to be used in similar ways to words from a known language, and to investigate what processes govern the formation and use of these new lexical representations.

Method

Participants

16 subjects participated in five 75 minute sessions on consecutive days. Subjects were all students at the University of Rochester and were native speakers of English. Subjects received payment on each day of participation (in a sequence of $5, $5, $10, $10, $20) for a total of $50.

Materials

Two 24-word artificial lexicons were designed for the purposes of this study. The lexicons included only bi-syllabic words of the form CVCV. All consonants were selected from the set $\{k, p, t, g, b, d\}$, and all vowels from the set $\{a, i, o, u\}$. Neither consonants nor vowels were repeated within a word. A complete lexicon can be found in Table 1.

A lexicon consisted of 12 high-frequency (HF) words and 12 low-frequency (LF) words. High-frequency words appeared approximately four times more often than low-frequency words. Each high-frequency word was played 192 times over the course of training; each low-frequency word was played 48 times. Perfect performance on all testing and training trials would result in a subject producing each high-frequency word 202 times, and each low-frequency word 58 times. If each utterance were also considered an auditory exposure to the word, the ideal subject would have heard each high-frequency word 394 times during the week, and each low-frequency word 106 times. This resulted in 6000 exposures to words in the lexicon, with %78.8 of the exposures being high-frequency words. Since no subject performed perfectly, the exact number of exposures to high- and low-frequency words differed slightly across subjects.

Table 2: Neighborhood types

	Sparse A	*Sparse B*	*Dense A*	*Dense B*
Root	baki	bago	bago	bagi
Neighbor 1	bagi	bugo	bagi	baki
Neighbor 2			bugo	bago
Neighbor 3			kago	kagi

The lexicons were constructed in a manner that gave rise to phonological neighborhoods with varying densities. Words in high-density (HD) neighborhoods had 3 neighbors. Words in low-density (LD) neighborhoods had 1 neighbor. There were 12 high-density words and 12 low-density words. Frequency levels were assigned evenly within each level of neighborhood density, so that there were 6 HF-HD words, 6 HF-LD words, 6 LF-HD words, and 6 LF-LD words (see Table 1).

The precise structure of the phonological similarity found within a neighborhood differed across neighborhoods. Some neighborhoods contained more *onset neighbors*. Onset neighbors are defined as words that are identical through the onset and nucleus of the first syllable. Put differently, onset neighbors are those words which differ by one segment somewhere in the second syllable. A sample of each of the types of neighborhoods found in the lexicon can be seen in Table 2.

The second lexicon was a translation of the first, where each consonant was substituted for some other consonant, and each vowel substituted for another vowel. In executing the translation, voiceless consonants were exclusively swapped with other voiceless consonants, and the same held for voiced consonants.

A female native English speaker was recorded speaking each of the words in the lexicons. Several samples of each word were recorded, and the authors chose the best token for use in the study. Words were spoken with stress on both syllables, in part to make them more distinct from English words. The selected samples were normalized to a standard intensity.

The images used in the study were the same across both lexicons. A large number of Attneave shapes were randomly generated using a Matlab package (Collin & Mcmullen, 2002). The authors independently evaluated the shapes for name-ability, and then jointly selected a set of 24 equally unnameable shapes. Four sample shapes can be seen in Figure 1.

The experiment was run on the ExBuilder experiment de-

Figure 1: Sample shapes

sign platform on a PC running Windows XP Professional Edition. Subjects wore a Shure WH20 headset microphone which was connected to an M-Audio Fast Track Pro USB sound card with mic pre-amps. Naming latencies were collected using a Cedrus SV-1 voice key.

Procedure

On each day subjects completed a series of five training blocks followed by two testing blocks. Each training block consisted of 120 trials (8 presentations per HF word, 2 per LF word); each testing block consisted of 24 trials (one presentation per word).

Each subject was assigned to one of the two lexicons on the first day. At the beginning of a subject's first session, a random mapping from words to images was generated for that subject. This mapping stayed consistent across days of training for a given subject, but each subject had a unique mapping.

Subjects were instructed to name the picture displayed on the screen as quickly and accurately as possible as soon as they heard a beep. It was further explained that when subjects were uncertain about the name of the shape, they should always name the shape according to their best guess. These instructions were consistent across all types of training and testing trials. On each trial, subjects were presented with a blank white screen for 500 ms, followed by the display of one of the pictures. After 500 ms subjects either heard the name of the shape or were cued to name the shape by a pure-tone beep (depending on the type of trial). If on any trial the subject did not speak within 3 seconds of the beep, the trial ended and the subject was warned to speak more quickly.

Subjects participated in two types of training trials: repetition trials and supervised learning trials. In repetition trials, subjects heard the name of the shape followed by a the beep, at which point they were to repeat the word they had just heard. In supervised naming trials, subjects were cued to speak after 1 second. After subjects had named the picture on the screen, the correct name of the picture was played. During testing trials subjects named the displayed shape and received no feedback on their response.

The schedule of training and testing blocks changed from day to day. There were four repetition trials the first day, three the second, and so on until there was a single repetition trial on day five. Day one includes one supervised naming, and there is one more supervised training block each day. There are two blocks of testing each day.

Results

All responses in test trials were coded for correctness. In addition, correct trials were coded for whether the entire response consisted of a single word. These trials were referred to as "perfect". Naming latencies are measured from the end of the beep that cued participants to speak on every trial.

The analyses described below were carried out in the R statistical programming environment using the lme4 and languageR packages (R Development Core Team, 2006; Bates & Sarkar, 2007; Baayen, 2007). A summary of the results collapsed across all days of the study can be found in Table 3.

Naming Latencies

Looking at the means in Table 3, several basic trends stand out. High-frequency words tend to be named more quickly, as do low-density words. In order to determine the contribution of each of these factors to the data pattern, it is necessary to model the results as a function of our dependent variables and appropriate random effects.

A series of linear mixed effects models were evaluated using naming latency as the dependent variable. Only trials regarded as "perfect" were included in this analysis. Different models represent different hypotheses about the way particular factors interact to produce the observed pattern of results. Models can be evaluated against one another, and individual factors can be shown to have positive or negative effects on the model's ability to account for the data. Subjects, items, lexicon, and pictures are modeled as random factors. Session number, word frequency, neighborhood density, initial syllable frequency, neighborhood density, log phonotactic likelihood, and neighborhood type are included as fixed effects in the model. These factors all reflect the internal structure of the artificial lexicon. Initial syllable frequency is the total number of times the initial syllable of a word appears within the lexicon (regardless of position). Log phonotactic likelihood is a measure of the overall probability that a particular set of phonemes will occur in a specific order, given the patterns established in the rest of the lexicon.

To test the influence of subjects' knowledge of English on their performance in the artificial naming task, several post-hoc covariates were also included in the model. These include summed English syllable frequency, the English syllable frequency for the onset syllable, and the log phonotactic likelihood of the word in English. These measures were calculated in the same manner as those based on the artificial lexicon, but reflect the distributional properties of English. These covariates are further explored below.

Two measures of significance are available in this analysis. The first measures the contribution of a given factor to the model's ability to account for the data. Individual factors are removed from the model, resulting in a change in the degrees of freedom in the model and in the log likelihood of the data given the model (denoted λ). A χ^2-test is performed on the log likelihood values with the change in degrees of freedom in the model as the d.f. for the test. A significant difference in

Table 3: Mean Naming Latencies (ms) and Percent Correct

	HF	*LF*	*HD*	*LD*
Naming (ms)	533.37	544.20	559.97	516.44
Correct (%)	83.80	68.18	76.56	75.42
	HF-HD	*HF-LD*	*LF-HD*	*LF-LD*
Naming (ms)	553.84	512.62	567.42	521.02
Correct (%)	84.79	82.81	68.33	68.02
	Sparse A	*Sparse B*	*Dense A*	*Dense B*
Naming (ms)	525.61	511.50	563.47	552.80
Correct (%)	80.4	72.89	76.88	75.94

log likelihood scores means that the factor contributes meaningfully to the model's ability to describe the data.

Among the lexical factors included in the model, session, frequency, and neighborhood density, all have significant effects on the model's performance (session: $\chi^2_{\Delta\lambda}(1) = 93.85, p < 0.00001$); frequency: $\chi^2_{\Delta\lambda}(1) = 6.16, p < 0.01$; density: $\chi^2_{\Delta\lambda}(1) = 7.40, p < 0.01$. Importantly, neighborhood type does not affect the model's performance significantly ($\chi^2_{\Delta\lambda}(4) = 5.41, p > 0.2$). The random effects for subject, word, and image were also significant (subject: $\chi^2_{\Delta\lambda}(1) = 286.67, p < 0.0001$; word: $\chi^2_{\Delta\lambda}(1) = 3.89, p < 0.05$; image: $\chi^2_{\Delta\lambda}(1) = 4.51, p < 0.05$). As expected, the lexicon a subject was assigned to did not significantly affect the model's fit to the data. None of the English-based covariates contributed significantly to the model (all $p > 0.15$).

Given this understanding of the role of the candidate factors in explaining the data, it is possible to ask how the significant factors impact subjects' performances. Specifically, the parameter estimates for frequency, neighborhood, and session reflect the change in naming latency between levels of those factors. A Monte Carlo Markov Chain simulation was performed on the full model to test significance and establish confidence intervals for the parameter estimates. Subjects improved their naming speed by an estimated 46.59 ms with each day of training ($p < 0.0001$). Low-frequency words were estimated to be named 41.74 ms slower than high-frequency words ($p < 0.05$). Low-density words were found to be named 63.28 ms faster than high-density words ($p < 0.01$). These results are summarized in Figure 2.

Two different methods of significance testing have converged on the same set of results. Subjects tend to get faster over the course of training, are faster to name high-frequency words than low-frequency words, and are faster to name low-density words than they are to name high-density words. While overall neighborhood density does impact naming speed, the composition of a word's neighborhood does not.

Speech Errors

Means for several factors are reported in Table 3. In general, subjects were more accurate at naming high-frequency words than low-frequency words. Density doesn't appear to affect

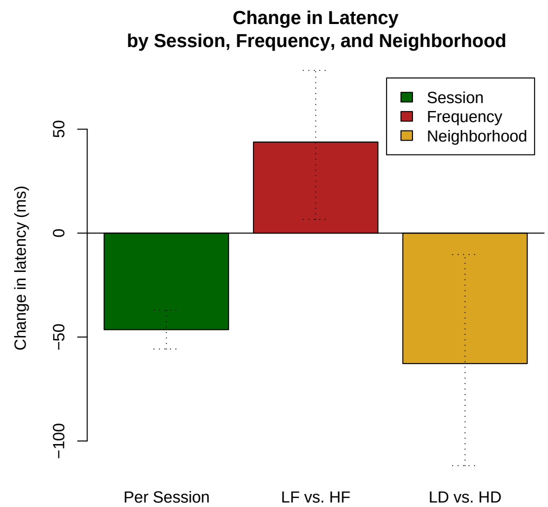

Figure 2: Parameter estimates for naming latencies

error rates. We examine these intuitions using a logit mixed model. This family of model is appropriate to this data set because the response is binary—correct or incorrect. The full model of the error data included the same terms as the full latency model.

Testing the contribution of individual factors to the fit of the model reveals significant decreases in model performance when frequency, session, or neighborhood composition is excluded (session: $\chi^2_{\Delta\lambda}(1) = 1069.50, p < 0.0001$; frequency: $\chi^2_{\Delta\lambda}(1) = 52.58, p < 0.0001$; neighborhood composition: $\chi^2_{\Delta\lambda}(5) = 12.498, p < .05$). Again, the random effects all significantly improve the performance of the model (all $p < 0.0001$), with the exception of which lexicon the subject was assigned to ($p > 0.50$). Neither phonotactic likelihood nor overall syllable frequency significantly improves the performance of the model (both $p > 0.50$); there was a marginal effect of onset syllable frequency ($\chi^2_{\Delta\lambda}(1) = 3.48, p < 0.10$).

We can also examine contrasts within the factors that contribute to explaining the error data. Parameter values are tested against the null hypothesis that the factor does not contribute to the model (a parameter value of 0). Using the model discussed above, we find that the parameters corresponding to frequency and session are significant (both $p < 0.001$). Planned contrasts were carried out to investigate the role of neighborhood density and phonological similarity. The parameter representing neighborhood density is not significantly different than 0, meaning that a word's neighborhood density does not contribute to its the likelihood of producing it correctly. The parameter estimates for session, frequency, and neighborhood density are shown in Figure 3.

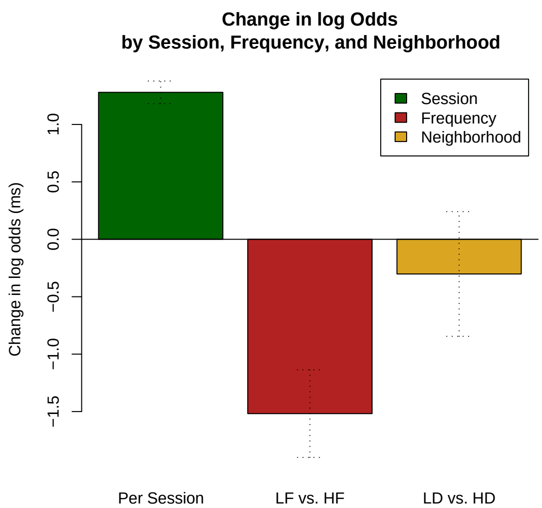

Figure 3: Parameter estimates for error rates

Both estimates of initial syllable frequency are marginally significant (both $p < 0.10$). Words with a high initial syllable frequency within the lexicon are slightly less likely to be produced correctly, while words with high initial syllable frequency in English are slightly more likely to be produced correctly.

Influence of English

Many studies of nonsense word production have shown "bleed in" effects from a speaker's native language, including effects of phonotactics and of syllable frequency (Cholin et al., 2006). For both dependent measures (reaction time and error rates) we failed to find any impact of the distributional properties the words would have if they were in the English lexicon (with the exception of the marginally significant advantage in error rates for words with a high English onset syllable frequency). To further examine the relationship between English and the artificial lexicon, a correlational analysis was performed comparing the distribution of summed syllable frequency and log phonotactic likelihood in the two languages. The results are displayed in Figure 4.

The figure shows the distribution of each measurement in a histogram on the diagonal. The right hand side of the figure displays scatterplots that graphically represent the correlation

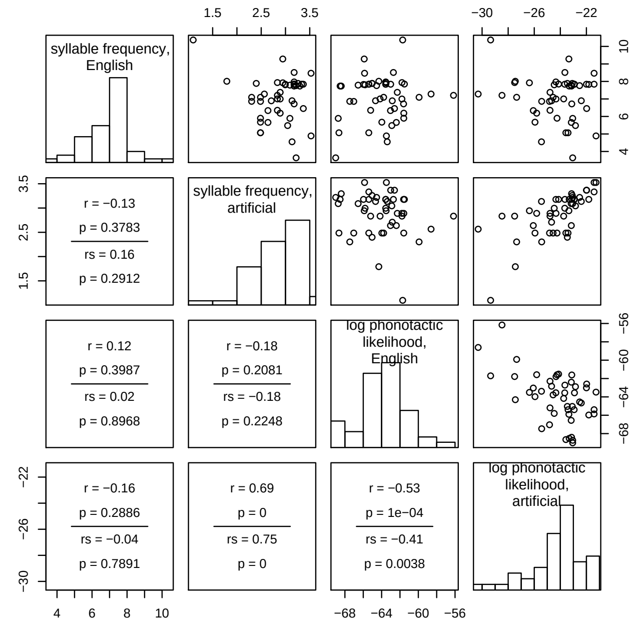

Figure 4: Correlations between English and the artificial lexicon

between pairs of measurements. The left hand side of the figure reports two measures of correlation (r and Pearson's r) along with significance levels.

We find that the phonotactics of the artificial lexicon are strongly negatively correlated with those of English. This reflects the fact that the lexicon was constructed without regard for English phonotactic distributions—while the words were chosen to be pronounceable, they were not selected to resemble existing English words. The end result is that the words that are most likely to be English words are the least representative of the internal phonotactic distributions of the artificial lexicon. There is also a significant positive correlation between phonotactic likelihood and syllable frequency within the lexicon. This can be explained by observing that many syllables are repeated across the lexicon, and a common syllable will also include high transitional probabilities between its onset consonant and nucleus.

The behavioral results showed no influence of English knowledge on naming performance in the artificial lexicon naming tasks. This analysis reveals that the units that make up the words in the lexicon (phonemes and syllables) are distributed quite differently in the lexicon and in English. These differences may have cued learners to track the distributional properties of the new words separately from those of known English words.

Discussion

This study offers a promising look at the potential usefulness of artificial lexicons in studying language production. We replicated classic frequency effects in both naming latencies and error rates after just five days of exposure to a relatively subtle frequency distinction (*c.f.* Levelt et al., 1991; Dell, 1986).

We found an inhibitory effect of neighborhood density on naming latencies, consistent with the results of some studies of neighborhood effects in production (Vitevitch & Stamer, 2006). While this finding fails to replicate the facilitatory effects of neighborhood density observed in English, it is possible that the artificial lexicon technique will allow us to design stimuli that will reveal the cause for these disparate findings.

Finally, while we found an effect of neighborhood density on naming, we found an effect of neighborhood structure on error rates. This indicates that phonology may play different roles at different points in lexical processing. Lexical selection may be sensitive to similarity in the onset of a word, leading to poorer performance in neighborhoods where the similarity tends to be loaded at the front end of the word. Articulatory planning or phoneme selection may be influenced more globally by phonological similarity, leading to effects of neighborhood density on naming latencies.

Acknowledgments

This work was supported by NIH grant DC005071. Thanks to Natalie Klein for her assistance and patience generating stimuli, and to T. Florian Jaeger for his generous help with the statistical analyses presented here.

References

Baayen, R. H. (2007). *languager: Data sets and functions with analyzing linguistic data.* (R package version 0.1)

Bates, D., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using s4 classes.*

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006, March). Effects of syllable frequency in speech production. *Cognition*, *99*(2), 205–235.

Collin, C. A., & Mcmullen, P. A. (2002, February). Using matlab to generate families of similar attneave shapes. *Behavior Research Methods, Instruments, & Computers*, *34*(1), 55–68.

Dell, G. S. (1986, July). A spreading-activation theory of retrieval in sentence production. *Psychol Rev*, *93*(3), 283–321.

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.* New York: Mouton.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000, November). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *J Exp Psychol Learn Mem Cogn*, *26*(6), 1355–1367.

Jescheniak, J. D., & Levelt, W. J. (1994, July). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological review*, *98*(1), 122–142.

Lipinski, J., & Gupta, P. (2005, February). Does neighborhood density influence repetition latency for nonwords? separating the effects of density and duration. *Journal of Memory and Language*, *52*(2), 171–192.

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

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003, June). The time course of spoken word learning and recognition: studies with artificial lexicons. *J Exp Psychol Gen*, *132*(2), 202–227.

Marslen-Wilson, W. D. (1987, March). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102.

Mcclelland, J. L., & Elman, J. L. (1986, January). The trace model of speech perception. *Cognit Psychol*, *18*(1), 1–86.

Meyer, A. S., & van der Meulen, F. F. (2000, June). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychon Bull Rev*, *7*(2), 314–319.

R Development Core Team. (2006). *R: A language and environment for statistical computing.* Vienna, Austria. (ISBN 3-900051-07-0)

Vitevitch, M. S. (2002, July). The influence of phonological similarity neighborhoods on speech production. *J Exp Psychol Learn Mem Cogn*, *28*(4), 735–747.

Vitevitch, M. S., Armbruster, J., & Chu, S. (2004, March). Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. *J Exp Psychol Learn Mem Cogn*, *30*(2), 514–529.

Vitevitch, M. S., & Luce, P. A. (1998, July). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4), 325–329.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999, June). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, *68*(1-2), 306–311.

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

Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in spanish speech production. *Language & Cognitive Processes*, *21*(6), 760–770.

Wheeldon, L. (2003, February). Inhibitory form priming of spoken word production. *Language and Cognitive Processes*, *18*(1), 81–109.