RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 22 (1998)
Indiana University

All Neighborhoods are not Created Equal: The Phonological *P*-Metric and Spoken Word Recognition[1]

Michael S. Vitevitch

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

---

[1] This research was supported by NIH-NIDCD Training Grant DC00012. I'd like to thank Jesse Young for running participants in this experiment, and Paul A. Luce, Ph.D. for his assistance with calculating the phonotactic probabilities and isolation points of the stimuli.

All Neighborhoods are not Created Equal: The Phonological *P*-Metric and Spoken Word Recognition

Abstract. A great deal of work has shown that the size of the phonological neighborhood affects the speed and accuracy of spoken word recognition. Words with many similar sounding words (i.e., a dense neighborhood) are recognized more slowly and less accurately than words with few similar sounding words (i.e., a sparse neighborhood). However, little work has examined the structural differences that may exist among neighborhoods of equal size and the effects these differences may have on spoken word recognition. The current results suggest that such differences do exist among words with the same lexical density, and that these differences have demonstrable effects on the speed of spoken word recognition. The implications of these results for models of spoken word recognition are discussed.

Introduction

Landauer and Streeter (1973) demonstrated that a single letter substitution results in a difference between common and less common words in the number of confusable words, or neighbors. Common words tended to have many neighbors, and less common words tended to have few neighbors. Words with many neighbors are said to be in a dense neighborhood, and words with few neighbors are said to be in a sparse neighborhood. Work in visual word recognition (Andrews, 1989, 1997; Coltheart, Davelaar, Jonasson, & Besner, 1977) and spoken word recognition (Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990) has shown that the number of competing word forms activated in memory influences the process of word recognition. The results from *spoken* word recognition experiments (Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990) have consistently found that words from dense neighborhoods are recognized more slowly and less accurately than words from sparse neighborhoods.

However, research in *visual* word recognition has produced conflicting results. Coltheart et al. (1977) found that the number of neighbors (*N*) affected the classification of nonwords. Nonwords with many neighbors were classified more slowly than nonwords with few neighbors. However, no difference in performance was found for real words. In contrast, Andrews (1989) found that neighborhood density *did* affect visual word recognition. In visual naming tasks and visual lexical decision tasks, particpants respond faster to words from dense neighborhoods than words from sparse neighborhoods—the opposite of spoken word recognition results.

To muddy the empirical water even further, Grainger, O'Regan, Jacobs, and Segui (1989) failed to find differences in performance (as measured in a visual lexical decision task and an eye-fixation task) between words with few neighbors and words with many neighbors, failing to replicate the findings of Andrews (1989). Grainger et al (1989) did, however, find that repsonses were slower for words with high-frequency neighbors than for words with low-frequency neighbors. This result indirectly *contradicts* the findings of Andrews (1989), because words with high-frequency neighborhoods are also likely to have dense neighborhoods (Andrews, 1997; Bard & Shillcock, 1993). Thus, in Grainger et al's (1989) study, words with high-frequency/"dense" neighborhoods were recognized slower than words with low-frequency/"sparse" neighborhoods—identical to the pattern obtained in spoken word recognition studies (Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990).

A number of studies (see Andrews, 1997; Segui & Grainger, 1993; for reviews) have investigated the conflicting results regarding neighborhood density in visual word recognition. One attempt to address the affects of neighborhood density on visual word recognition by Johnson and Pugh (1994) also examined the influence of what has been referred to as the "spread" of the neighborhood (Andrews, 1997; Pugh, Rexer, Peter, and Katz, 1994), and how this factor may explain the contradicting results. "Spread" refers to the number of letter positions in which a substitution can be made to produce a real word (Johnson & Pugh, 1994; Pugh et al., 1994).

For example, when a single letter is substituted in the word *clue* the words *glue* and *club* are formed, giving *clue* a spread of two. The word *sand* can form the words *hand*, *send*, *said*, and *sank*, giving it a spread of four. Although spread and density are correlated (Andrews, 1997), Pugh et al. (1994) argue that the spread, or *P*-metric, of a word is a better predictor of neighborhood effects than the *N*-count (Coltheart et al., 1977)—the number of similar words formed by the substitution of a single letter in a word.

In Experiment 6, Johnson and Pugh (1994) held neighborhood density constant while manipulating the spread of the neighborhoods, and held the spread of the neighborhoods constant while manipulating neighborhood density. In a visual lexical decision task holding *N* constant, words with many positions producing words in the neighborhood (large-*P*) were responded to more slowly and less accurately than words with few positions producing words in the neighborhood (small-*P*). When *P* (spread) was held constant, words with large *N* (dense neighborhoods) were responded to faster than words with small *N* (sparse neighborhoods). They argued that the process of winnowing down the neighborhood to the target word is not influenced by the number of items that need to be eliminated (i.e., the size of the neighborhood). Rather, resolving lexical activation is influenced by the number of letter positions that yield words (i.e., spread).

Furthermore, when resolving time is held constant (*P* held constant), the size of the neighborhood (*N*) in a visual lexical decision task acts as a response bias. Words in dense neighborhoods have a stronger WORD response bias than words in sparse neighborhoods, resulting in words in dense neighborhoods being responded to faster than words in a sparse neighborhood. Nonwords that held *P* constant and varied *N* had the opposite effect. In the visual lexical decision task, nonwords with many neighbors were responded to more slowly than nonwords with few neighbors, further suggesting that *N* may act as a response bias. The failure to adequately control *P* in past studies (e.g., Andrews, 1989; Coltheart et al., 1977; Grainger et al, 1989) may explain why previous findings in the visual word recognition literature have not been consistent.

Although Johnson and Pugh (1994) have demonstrated the importance of spread (*P*) in visual word recognition, no analogous analysis has been performed on spoken words. Are there similar differences in the *phonological* structure among words with equivalently sized neighborhoods? Do words differ in the number of words formed when a *phoneme*, instead of a letter, is substituted into a word? If such differences do exist, what affects do they have on the process of *spoken* word recognition?

The present experiment was an attempt to examine these questions. Words with equivalent densities (phonologically-based *N*-count), but different spreads (phonologically-based *P*-count) were selected and presented to participants in an auditory naming task. Based on Johnson and Pugh's (1994) results, we predicted that words with a small "spread" would be processed faster than words with a large "spread." That is, words that had few positions result in real words by a single phoneme substitution would be processed faster than words that had many positions result in real words by a single phoneme substitution. Such a difference would suggest that structural differences within equivalently sized neighborhoods do affect spoken word recognition.

Method

Participants

Fifteen native English speakers from the Indiana University pool of Introductory Psychology students participated in partial fulfillment of a course requirement. None of the participants had a history of speech or hearing problems at the time of testing.

Stimuli

Two hundred consonant-vowel-consonant (CVC) words were used in the experiment, and are listed in the appendix. The stimuli were divided into two groups of 100 words each. One group contained words that formed a word in the lexicon when a single phoneme was substituted into any of the phoneme positions of the word. For example, when a phoneme was substituted into any position of the word /dov/, a real word (e.g., /kov/, /daiv/, /dos/) was formed. These words had a phonological spread (phonological-*P*) of three. The other group contained words that formed a word in the lexicon when a single phoneme substitution was made into only two phoneme positions of the word. For example, a real word was formed when one phoneme was substituted into two positions of the word /kIN/ (e.g., forming /rIN/, /k*N/, and /kIn/), or /lof/ (e.g., forming /*of/, /lif/, and /lod/). These words had a phonological spread (phonological-*P*) of two. It should be noted that no distinction was made as to which two positions formed words. Thus, a word that formed a lexical item when a phoneme was substituted into the initial and medial positions was grouped with words that formed a lexical item when a phoneme was substituted into the initial and final positions, or medial and final positions.

Although the two groups of words differed in the number of words that could be formed by a single phoneme substitution (i.e. phonological-*P*), both groups were equivalent in word familiarity (F (1, 198) < 1;) as measured by a seven point scale (Nusbaum, Pisoni, and Davis, 1984) ranging from "1-Don't know the word," to "7-Know the word." Words having a phonological-*P* of two had a mean familiarity of 6.86, and words having a phonological-*P* of three had a mean familiarity of 6.84,

The two groups were also equivalent in word frequency (F (1, 198) < 1), as measured by log-frequency estimates derived from the Kucera and Francis (1967) word counts. Words having a phonological-*P* of two had a mean log-frequency of 1.04, and words having a phonological-*P* of three had a mean log-frequency of 1.03.

Most importantly, the number of similar sounding words, or neighborhood density, did not differ between the groups (F (1, 198) < 1). Both groups had a mean density of 8.53 words.

Finally, the two groups did not differ in neighborhood frequency (F (1, 198) < 1), as measured by log-frequency estimates derived from the Kucera and Francis (1967) word counts. The mean neighborhood log-frequency for words having a phonological-*P* of two was 2.20, and the mean neighborhood log-frequency for words having a phonological-*P* of three was 2.18.

When using a voice-activated response, it is important that the initial segments of the words found in each condition are comparable in their acoustic-phonetic properties. An analysis of the 18 initial segments from the stimulus words showed no difference (^{2} (1 7) = 13.49, $p > .10$) between the distribution of initial segments in words having a phonological-*P* of two and the distribution of initial segments in words having a phonological-*P* of three.

The stimuli were spoken in isolation and recorded by the author in an IAC sound attenuated booth. The stimuli were filtered at 10.4 kHz and digitized at a sampling rate of 20 kHz using a 16-bit analog-to-digital converter. All words were edited into individual files, leveled at 60 db, and stored on computer disk. Stimulus durations for both groups were also equivalent (F (1, 198) < 1). Words having a phonological-*P* of two had a mean stimulus duration of 845 *msec*, and words having a phonological-*P* of three had a mean stimulus duration of 837 *msec*.

Procedure

Participants were tested individually. Each participant was seated in a booth equipped with a computer terminal, a pair of Beyerdynamic DT-100 headphones, and a Shure Unidyne III dynamic microphone model 545 interfaced with a voice-activated response key. Presentation of stimuli and response collection was controlled by a 200MHz Gateway 2000 Pentium computer.

A trial proceeded as follows: A prompt appeared on the computer screen, and one of the spoken stimulus items was presented at 70 db over the headphones. The participant then repeated the item as quickly and as accurately as possible into the microphone, which was placed approximately 2-3 inches from the mouth of the participant. Each participant received all two hundred words.

Reaction time was measured by the computer from the onset of the stimulus to the onset of the participant's verbal response. All responses were recorded on audiotape for accuracy analysis. Accuracy was assessed by listening to the participants' responses and comparing them with a written transcription of the words. A stimulus was scored as correct if there was an identical match on all segments of the word.

Prior to the experimental trials, each participant received ten practice trials. These trials were used to familiarize the participants with the task and were not included in the final data analysis.

Results

The mean reaction time for each condition of phonological "spread" is shown in Figure 1. Accuracy scores for each condition are superimposed on the appropriate bars. Separate ANOVAs were performed on reaction times and accuracy scores by participants (F_1) and by items (F_2).

For the reaction times a significant difference between the two conditions was found (F_1 (1, 14) = 49.04, $p < .001$; F_2 (1, 198) = 5.95, $p < .05$). Words with a phonological-*P* of two were repeated faster (869 msec) than words with a phonological-*P* of three (899 msec). No differences were found for the accuracy scores (both Fs < 1).

Discussion

The present results demonstrate that structural differences do exist among words of equivalent phonological neighborhood densities. More importantly, these results demonstrate that structural differences among equivalently sized neighborhoods affect the speed of processing spoken words. Specifically, a word that has few positions that form a lexical item when a phoneme is substituted (small phonological *P*) is repeated faster than a word that has many positions that form a lexical item when a phoneme is substituted (large phonological *P*).

The ability to account for this data varies among current models of spoken word recognition. Cohort theory (Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978) does not associate processing costs with cohorts of varying sizes. According to cohort theory, a word with a large cohort will be processed at the same speed as a word with a small cohort. Based on this assumption, words with equivalently sized neighborhoods should also be processed at the same speed. This prediction was not confirmed, making cohort theory unable to account for the current results.

Another assumption of cohort theory, however, states that the isolation point of a word determines how quickly activation in the neighborhood/cohort is resolved. The isolation point is that point in a word which distinguishes it from all other words in the lexicon. Words that have an early isolation point are recognized faster than words that have a late isolation point. Cohort theory could, therefore, account for the current findings if one assumes that the words with a phonological spread (*P*) of two have earlier isolation points than the words with a *P* of three. This prediction, however, is also not supported. A computational analysis by Luce (1986) suggests that isolation points do not reliably distinguish monosyllabic words from other words in the lexicon (N.B., the 200 stimuli used in this study were monosyllabic CVC's). Furthermore, an analysis of the isolation points for the 200 stimuli used in this study shows no difference ($F(1, 198) = 1.00, p > .10$) between words with a *P* of two (a mean isolation point of 2.99 phonemes) and words with a *P* of three (a mean isolation point of 3.00 phonemes). Thus, processing or information about the entire word—for words with a *P* of two and three—was required to distinguish it from any other word in the lexicon.

Insert Figure 1 about here

Additional analyses of the distribution of phoneme positions that formed real words (i.e., initial and medial, medial and final, etc.) for the words with a *P* of two further suggest that most if not all of the word was required in order to disabiguate the lexical item from competing *neighbors*. Thirty percent (30%) of the words with a *P* of two formed lexical items when phonemes were substituted into the initial and medial positions. Fifty-one percent (51%) of the words with a *P* of two formed lexical items when phonemes were substituted into the initial and final positions. Nineteen percent (19%) of the words with a *P* of two formed lexical items when phonemes were substituted into the medial and final positions. Thus, in seventy percent (51% + 19% = 70%) of the stimulus items with a *P* of two, lexical competition within the neighborhood could not be resolved until the final phoneme had been heard. This further demonstrates that cohort theory can not account for the current findings.

The current instantiation of the neighborhood activation model (Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990) also appears unable to account for the present results. The neighborhood activation model (NAM) predicts that words with dense neighborhoods will be processed slower and less accurately than words with sparse neighborhoods (Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990). No predictions regarding differences in processing are made for words of equivalent neighborhood density. Presumably, words of equivalent neighborhood density would be processed with the same speed and accuracy when frequency is held constant. Similarly, the NAM fails to discuss differences that may—and indeed *do*—exist among words of equivalent neighborhood density. Thus, the NAM also appears unable to account for the present findings.

The fact that the spoken word recognition system is sensitive to differences among words that have approximately equal amounts of activation at the lexical level (i.e., the same number of neighbors) suggests that processing for these words may be influenced by structural information from other levels of representation and processing. Recent work investigating the influences of phonotactic information on spoken word recognition (Vitevitch, 1997; Vitevitch & Luce, 1998a; 1998b) suggests that a sub-lexical level of representation an processing is also involved in the process of spoken word recognition. The

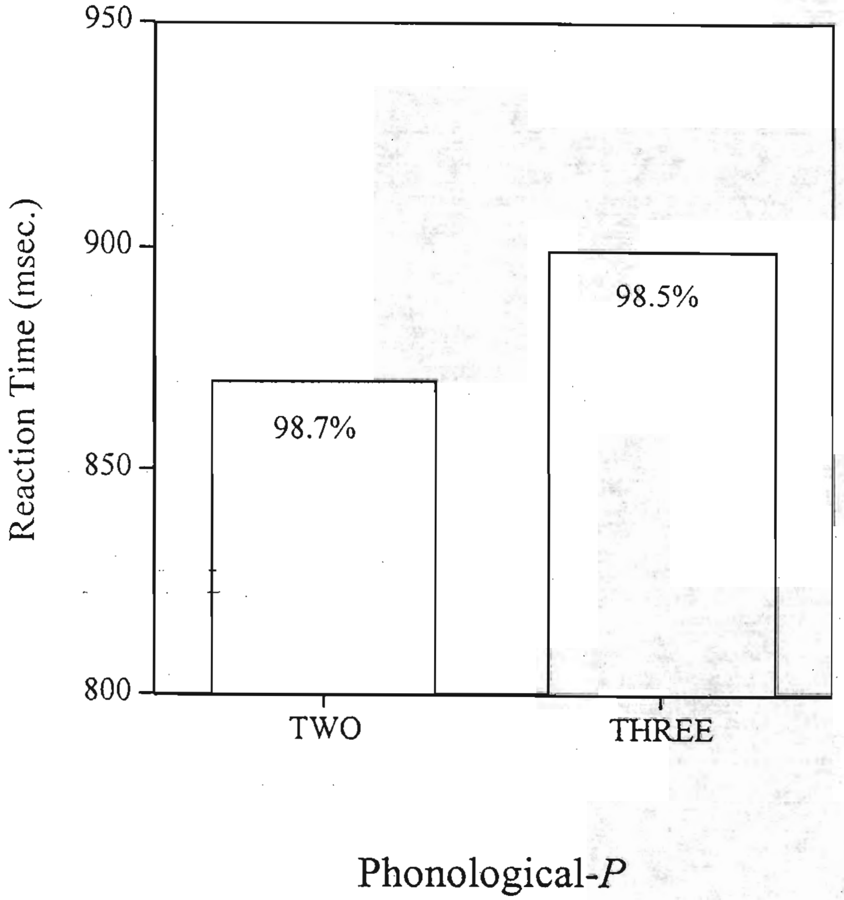

Figure 1. Mean reaction time in milliseconds for words with a spread (*P*) of two or three phonemes. Percent correct is superimposed on the appropriate bars.

present result—a processing difference among words with equivalent neighborhood density—adds support to the hypothesis of a sub-lexical level of representation and processing influencing spoken word recognition. Additional information regarding the number of phoneme positions that form real words was required to resolve the activation and competition among lexical items, and to recognize the spoken word.

The comparison of the present results with the work of Vitevitch (1997) and Vitevitch and Luce (1998a; 1998b) is not meant to imply that the present findings are an inadvertent or serendipitous effect of probabilistic phonotactics. Indeed, a post-hoc analysis of the stimuli shows no difference in phonotactic probabilities between words with a *P* of two and words with a *P* of three.[2] Rather, the comparison was designed to demonstrate another way that sub-lexical representations and processes may be used in the processing of spoken words.

As in Johnson and Pugh (1994), we hypothesized that the sub-lexical representations activated in words with a spread of two afford greater "resolving power" to the word recognition system than words with a spread of three. Precious time and resources may be wasted in attempting to resolve lexical competition in a word with a spread of three at a phoneme position that received unambiguous input. Words with a spread of two may focus the recognition process on fewer phoneme positions within a word, allowing for the rapid discrimination of phonemic contrasts which are *crucial* for identifying the critical word among competing lexical neighbors. Therefore, only models which postulate a sub-lexical and a lexical level of representation and processing, such as SHORTLIST (Norris, 1994), TRACE (McClelland & Elman, 1986), and PARSYN (Auer & Luce, 1998), would be able to account for the present findings.

Further investigation of this phenomenon using other tasks—including lexical decision, AX matching, perceptual identification, priming, and gating—are currently underway. Computer simulations using PARSYN (Auer & Luce, 1998) should also provide valuable new information for understanding the present results and in generating predictions for future experiments. In addition, other manipulations, such as holding *P* constant and varying neighborhood density, as in Johnson and Pugh (1994), or orthogonally manipulating both *N* and *P*, are also planned.

References

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition,* 15, 802-814.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4, 439-461.

---

[2] The mean log-frequency-weighted phone probabilities for the initial consonant of words with a *P* of two was .038, and for words with a *P* of three was .040 ($F(1, 198) < 1$). The mean log-frequency-weighted phone probabilities for the vowel of words with a *P* of two was .037, and for words with a *P* of three was .035 ($F(1, 198) < 1$). The mean log-frequency-weighted phone probabilities for the final consonant of words with a *P* of two was .039, and for words with a *P* of three was .035 ($F(1, 198) < 1$). The mean log-frequency-weighted biphone probabilities for CV's of words with a *P* of two was .002, and for words with a *P* of three was .002 ($F(1, 198) < 1$). The mean log-frequency-weighted biphone probabilities for VC's of words with a *P* of two was .002, and for words with a *P* of three was .002 ($F(1, 198) < 1$). The mean sum of the log-frequency-weighted segments of words with a *P* of two was .114, and for words with a *P* of three was .110 ($F(1, 198) < 1$). Finally, the mean sum of the log-frequency-weighted sequences of words with a *P* of two was .004, and for words with a *P* of three was .004 ($F(1, 198) < 1$).

Auer, E.T. & Luce, P.A. (1998). PARSYN: A processing model of neighborhood activation and phonotactics in spoken word recognition. Unpublished manuscript. University at Buffalo, Buffalo, NY.

Bard, E., & Shillcock, R. (1993). Competitor effects during lexical access: Chasing Zipf's tail. In G.T.M. Altmann & R. Shillcock, (Eds.) *Cognitive Models of Speech Processing: The second Sperlonga meeting*. (pp. 235-275) Hillsdale: Lawrence Erlbaum Associates.

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic, (Ed.), *Attention and Performance VI* (pp. 535-556). New York: Academic Press.

Grainger, J., O'Regan, J.K., Jaccobs, A.M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics*, 45, 189-195.

Johnson, N.F. & Pugh, K.R. (1994). A cohort model of visual word recognition. *Cognitive Psychology*, 26, 240-346.

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

Landauer, T.K. & Streeter, L.A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 12, 119-131.

Luce, P.A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics*, 39, 155-158.

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

Luce, P.A., Pisoni, D.B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In G.T.M. Altmann (ed) *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. (pp.142-147) Cambridge: MIT Press.

Marslen-Wilson, W.D. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1-71.

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

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

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

Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In *Research on Speech Perception Progress Report No. 10.* Speech Research Laboratory. Psychology Department, Indiana University, Bloomington, Indiana.

Pugh, K.R., Rexer, K., Peter, M., & Katz, L. (1994). Neighborhood effects in visual word recognition: Effects of letter delay and nonword context difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 639-648.

Segui, J. & Grainger, J. (1993). An overview of neighbourhood effects in word recognition. In G.T.M. Altmann & R. Shillcock, (Eds.) *Cognitive Models of Speech Processing: The second Sperlonga meeting*. (pp. 497-516) Hillsdale: Lawrence Erlbaum Associates.

Vitevitch, M.S. (1997). Phonotactics and spoken word recognition. Unpublished doctoral dissertation. University at Buffalo, Buffalo, NY.

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

Vitevitch, M.S. & Luce, P.A. (1998b). Probabilistic phonotactics and neighborhood activation in spoken word recognition. Manuscript submitted for publication.

Appendix

Phonological-*P* of two

royal
gouge
nerve
choice
noise
loyal
jewel
wash
verge
fetch
poise
vague
verb
chef
sour
towel
youth
jazz
move
watch
judge
chive
join
dodge
thief
church
mesh
good
use
cough
moth
thought
loud
deaf
choose
dose
juice
tube
hedge
geese
sheath
noun
yearn
yawn
dual
soil
nudge
tease
palm
five

womb
wrong
jade
toss
soot
fish
shirt
doll
yoke
loaf
joke
mob
lull
gene
boil
path
siege
foul
cage
lose
veal
yell
thug
worse
gab
chalk
duke
badge
shawl
leash
thin
king
shine
thick
far
page
chute
yacht
maize
shag
hen
pub
league
wing
chill
sash
kite
sag
mug
toil

Phonological-*P* of three

vogue
bush
push
merge
vowel
shove
hoof
serf
lure
knife
turf
lodge
couch
notch
mouth
thong
dog
serge
serve
pause
gem
sauce
gown
fuss
thud
germ
jerk
tour
fuzz
wedge
lurch
search
teeth
retch
roof
theme
dirt
chess
gauze
firm
hog
guess
ledge
guise
pouch
lobe
fog
psalm
balm
nurse

foot
tooth
teach
herb
neck
wool
peg
safe
cheese
shout
term
chief
van
wipe
mouse
doubt
niece
dove
song
worm
goose
toad
vote
booth
verse
coil
worth
gym
sham
hive
jab
guide
gush
foam
mop
curb
news
hoop
chop
loin
shave
shop
wise
tong
howl
weave
chose
niche
dish
booze