# For Preschoolers, Lexical Access is Purely Lexical: Neighborhood Density Effects in Child Speech Perception and the Emergent Phoneme Hypothesis

Melinda Woodley
*University of California, Berkeley*

Both infants and adults are sensitive to phonotactic probability, the statistical distribution of sequences of sound below the level of the whole word. A now classic study by Jusczyk, Luce, and Charles-Luce (1994) demonstrated that 9-month-olds (but not 6-month-olds) prefer to listen to sounds patterns that occur more frequently in their native language, suggesting that sensitivity to sublexical sound patterns develops rather rapidly during the second half of the first year of life. At the other end of the spectrum, work with adults has shown that adults process speech at both the sublexical and whole-word levels. Vitevitch and Luce (1998, 1999) presented adult speakers of American English with both words and nonwords of either high or low phonotactic probability/neighborhood density (these factors were always correlated, such that no stimuli had high phonotactic probability but low neighborhood density, or vice versa). In several different perceptually-oriented tasks, adults were found to process words from dense lexical neighborhoods more *slowly* than words from sparse lexical neighborhoods, while nonwords with high phonotactic probability were processed more *quickly* than nonwords with low phonotactic probability.

Vitevitch and Luce argued that the frequency of sound patterns in the stimuli was either inhibitive or facilitatory depending on whether adults were processing at the lexical or sublexical level. Real words with frequent sound patterns were responded to more slowly due to a lexical competition effect, while nonwords with frequent sound patterns were responded to more quickly due to a phonotactic facilitation effect. Lexical competition and phonotactic facilitation were argued to be operative forces at all times, with the "winning" level of processing driving the response time; in essence, if stimuli were processed more quickly as words, a lexical competition effect was observed, whereas if they were processed more quickly via their component sounds, a phonotactic facilitation effect emerged.

The picture is less clear regarding the effects of phonotactic probability on speech processing in young children. Given that both infants and adults are sensitive to phonotactic probability, one might expect phonotactic facilitation to be obvious in children's speech processing as well. However, Munson, Swenson, and Manthei (2005), for example, found no reliable effects of lexicality, phonotactic probability, or neighborhood density on young children's response latencies in a repetition task similar to the one in Vitevitch and Luce (1998) when they used stimulus offset to response onset latencies as the dependent variable. Measuring response time (RT) from the offset of the stimulus to the onset of the response is arguably preferable, since this excludes the confound of stimulus duration, which is longer for low probability tokens in natural speech. However, using an onset-to-onset measure of RT for a subset of duration-matched stimuli, the authors did find a small (40 ms) but significant phonotactic facilitation effect for nonwords in the youngest group of children tested (mean age = 4;3).

One potential confound in the Munson et al. study, however, is that the response times were taken from an immediate shadowing task. While it has been argued that repetition latencies in immediate shadowing tasks are indicative of perceptual processing effects (e.g. Vitevitch & Luce, 1998), it has also been shown that phonotactic probability has a facilitatory effect on word production (Vitevitch, Armbrüster, & Chu, 2004). Thus if the phonotactic facilitation effect that Munson et al. found in their onset-to-onset analysis of nonwords turns out to be robust and replicable, it is unclear whether it should be attributed to facilitation in perception or production or both.

The question of whether to attribute any phonotactic facilitation effects to perception or production is an important one, as it can potentially inform our hypotheses regarding the origin and development of phonemic categories. If phonotactic facilitation occurs only in perception, this implies that young children have access to abstract sublexical categories during lexical access, as adults do. But if phonotactic facilitation occurs only in production, this implies that the phonotactic facilitation effect observed in Munson et al. (2005) is more accurately labeled a motor facilitation effect; children's articulators may simply have accumulated more practice producing frequent sound patterns, and this in turn could help drive the acquisition of psychologically real phonemic categories.

The present study, therefore, is intended to test whether phonotactic facilitation can be observed in a more purely perceptually-oriented task for subjects the same age as those in Munson et al. (2005), removing the confound of speech production in an attempt to localize any observed effects of phonotactic facilitation.

## Method

*Participants*

Participants were 12 children aged 3;10 – 4;10 (mean = 4;4.75), all in their first or second year of attendance at the same preschool. All were progressing normally through school and none had a history of speech, language, or hearing impairment, as reported by parents and teachers. Each child's parent completed a brief language background questionnaire prior to participation in the experiment. All of the data reported here are for children whose parents reported that the primary language spoken at home was English. However, due to the multicultural, multilingual nature of the community, many parents reported their children receiving at least some amount of exposure to a language other than English. In cases where parents noted a significant amount of exposure to a non-English language (such as a grandparent who speaks to the child in a different language), the teachers were consulted to make sure the child's English seemed to be developing at the same rate of his or her peers.

Order of participation was counterbalanced, such that a random half of the children first completed the nonword task, while the other half first completed the word task. Many of the children also participated in the second condition at a later date. However, this paper will be concerned with only the first condition in which each child participated.

*Stimuli*

There were 30 critical stimuli for each experimental condition (words vs. nonwords), consisting of 15 high neighborhood density/high phonotactic probability stimuli and 15 low neighborhood density/low phonotactic probability stimuli. (“Critical” stimuli are those that were presented in “same” trials, to be described in more detail below, and those for which reaction time is reported here.) For the purposes of this experiment, neighborhood density and phonotactic probability were highly correlated ($r = 0.86$ correlation between number of neighbors and mean segment probability based on the Child Mental Lexicon), such that there were no stimuli that had high neighborhood density but low phonotactic probability, or vice versa. For this reason, stimuli will henceforth be referred to simply as “high probability” vs. “low probability”. In addition to the 30 critical stimuli within each experimental condition, there were also 32 non-critical stimuli that were presented in the “different” trials. The non-critical stimuli also varied systematically by neighborhood density/phonotactic probability, but we do not report reaction time for the “different” trials here.

Summary statistics for the high vs. low probability stimuli are given in Table 1. Means are given for each metric based on the Child Mental Lexicon (Storkel & Hoover, 2010) and on the Hoosier

Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984). For both levels of lexicality (that is, words and nonwords), the differences in mean segment probability, mean biphone probability, mean number of phonological neighbors, and the sum of log frequencies for all phonological neighbors for high vs. low probability stimuli are statistically significant at well below the $p < .01$ level, as confirmed by a series of one-tailed *t*-tests. The differences in frequency and age of acquisition for high vs. low probability words was nonsignificant based on both the Child Mental Lexicon and the Hoosier Mental Lexicon (neither of these measures are applicable for nonwords).

| | **Mean Segment Probability** | | **Mean Biphone Probability** | | **Mean # of Phonological Neighbors** | | **Mean Frequency** | | **Mean Sum of Log Frequency of Neighbors** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **CML** | **HML** | **CML** | **HML** | **CML** | **HML** | **CML** | **HML** | **CML** | **HML** |
| **high probability words** | 0.0716 | 0.0659 | 0.0062 | 0.0047 | 20.1 | 28.0 | 3.23 | 2.79 | 51.84 | 6490.0 |
| **low probability words** | 0.0392 | 0.0355 | 0.0018 | 0.0012 | 6.8 | 12.8 | 2.79 | 2.52 | 15.04 | 686.2 |
| **high probability nonwords** | 0.0731 | 0.0680 | 0.0066 | 0.0060 | 20.7 | 28.8 | 0 | 0 | 55.26 | 10480.4 |
| **low probability nonwords** | 0.0396 | 0.0372 | 0.0013 | 0.0011 | 6.6 | 13.1 | 0 | 0 | 15.41 | 589.6 |

*Table 1: Summary statistics for the critical stimuli (those for which reaction times are analyzed) in the present experiment.*

Stimuli were carefully selected to be as phonetically balanced as possible; all were CVC, and all four types (2 levels of lexicality x 2 levels of probability) were matched as closely as possible for voicing and manner of the onset consonant and identity of the rime. This was largely achieved by selecting a phonetically matched set of high vs. low probability words whose age of acquisition was reported as less than 4;0 (Gilhooly & Logie, 1980; MRC Psycholinguistic Database), and then scrambling the onsets and rimes to produce nonwords with the desired phonotactic characteristics. This method provided a good starting point, but some modifications were necessary in order to avoid producing actual words, and to produce sufficiently high or low probability phonotactic patterns. The 60 critical stimuli are provided below in Table 2.

| high probability words | | low probability words | | high probability nonwords | low probability nonwords |
|---|---|---|---|---|---|
| fæt | *fat* | tʃɚtʃ | *church* | fot | tʃʌʃ |
| hɛd | *head* | fʌdʒ | *fudge* | hek | fedʒ |
| her | *hair* | haʊl | *howl* | hɛt | fɚtʃ |
| kek | *cake* | kedʒ | *cage* | hɪn | hoⁱs |
| kɪk | *kick* | pedʒ | *page* | kir | kɔʃ |
| pɛt | *pet* | ʃɚt | *shirt* | kæk | pɔs |
| pɪn | *pin* | sɔŋ | *song* | pɛd | ʃom |
| sʌn | *sun* | θʌm | *thumb* | ser | θel |
| tir | *tear* | bʊʃ | *bush* | tɛt | bus |
| bal | *ball* | dɔg | *dog* | baⁱn | dɚm |
| bæk | *back* | dor | *door* | bɪk | dedʒ |
| dæd | *dad* | gɚl | *girl* | dæt | dɚs |
| laⁱn | *line* | dʒus | *juice* | lɛn | dʒor |
| mɛn | *men* | nɚs | *nurse* | nɪn | mʌdʒ |
| ren | *rain* | rum | *room* | rɪn | raʊl |

*Table 2: Critical stimuli (those for which reaction times are analyzed) in the present experiment.*

### *Data Collection*

Data collection took place in a quiet room at the children's preschool. Children were seated at a small table with the researcher, who controlled the presentation of the stimuli. Sound files were played at a comfortable listening volume from a Dell laptop with two external loudspeakers. The presentation order was randomized for each child by the software used to present the stimuli, E-Prime.

The experiment consisted of 8 blocks of 6 trials each, and children were rewarded with a picture of a cartoon animal on the computer screen after each block. Children wore a wireless Sony lapel microphone whose receiver was connected to a button box attached to the laptop. For each trial, the experimenter asked the question, “Are these the same?” and cued the playing of the stimuli by pressing one of the buttons on the button box. The children then heard two stimuli, separated by 750 milliseconds of silence, and responded either “yes” or “no”[1]. Reaction time was recorded by the computer as the first sound detected by the lapel microphone following the offset of the second stimulus. A research assistant was seated just outside the room and could observe through a one-way mirror and listen in with an audio amplification system. The research assistant coded each trial for whether the child got the right answer, and whether the reaction time detected by the microphone was

1 A small number of children adopted slightly different response strategies, such as always saying, “that's the same” vs. “that's not the same”. Since these children (n=2) were consistent across trials and were otherwise very successful at the task, their data have not been excluded.

appropriate; 5.1% of all trials were excluded because children gave the wrong answer, began their response with "uhh" or another hesitation noise, or in some cases they coughed, touched the microphone, or gave an otherwise inappropriate response. Thus all reaction time data reported here come from "same" trials where the child's response and the reaction time recorded by the computer were correct.

The experiment began with a practice block consisting of a pre-determined maximum of 20 practice trials. However, most children needed only 4 – 6 trials (mean = 4.92) before it was clear they understood the task. The order of the stimuli for the practice trials was fixed and consisted of either words or nonwords (depending on the condition they were participating in) that were different from the experimental stimuli. For the practice block, children first heard 2 same trials, then 1 different trial, then roughly alternating same and different trials until the practice block was terminated at the experimenter's discretion[2]. Adequate feedback and guidance were given during the practice block, but feedback was kept to a minimum during the experimental trials.

Presentation order was fully randomized for the experimental stimuli. For the purposes of this experiment, we were interested in only "same" trials, but we needed to add an adequate number of "different" trials to keep the children on task while not making the experiment too long. Pilot testing revealed that a balance of 1 different trial for every 1 same trial resulted in an experiment that simply took too long, so we settled on 1 different trial for every 2 same trials. This meant that each block of 6 trials consisted of 4 same trials and 2 different trials, resulting in a maximum of 4 same trials in a row. High versus low probability stimuli were fully randomized, such that there were no restrictions on which type of "same" trials occurred in each block. In other words, some blocks may have consisted of only high probability stimuli or only low probability stimuli by chance.

As mentioned above, feedback during the experimental trials was kept to a minimum while still keeping the children engaged. The feedback for correct responses was generally "Good job!", while the feedback for incorrect responses was generally, "Are you sure? Let's try another one."

## Results

As mentioned above, the reaction time (RT) measures reported here all come from "same" trials where both the child's response and the response time recorded by the computer were accurate.[3] Because the standard deviation in RT was so high (747 ms), the standard outlier removal technique of

2 The precise order of the practice trials was as follows: same, same, different, same, different, different, same, different, same, different.

3 While Munson et al. found that offset-to-onset RT yielded slightly different results from onset-to-onset RT, as of yet there appear to be no qualitative differences between these two measures in my data.

excluding data points greater than 2.5 SD from the mean was not applicable here; thus no outliers have been removed from the data.

A 2x2 ANOVA examined the effects of lexicality (word vs. nonword) and phonotactic probability/neighborhood density (high probability vs. low probability). No significant effects were found in the by-subjects analysis, likely due to the low number of subjects tested thus far (n = 12), though the by-subject means do show the same pattern as the by-item means. In the by-items analysis, there was a main effect of phonotactic probability/neighborhood density ($F(1,59) = 4.34$, $p < .05$) and a marginally significant main effect of lexicality ($F(1,59) = 3.92$, $p = .05$). Reaction times were significantly slower for high probability stimuli, and tended to be slower for nonwords. There was no interaction between lexicality and phonotactic probability/neighborhood density.

A box and whisker plot depicting these results is shown in Figure 3. The black dot represents the median RT for each type of stimulus (word vs. nonword, high vs. low probability), the box shows the range where 50% of the data lie (25% above the median and 25% below the median), and the whiskers extend to 1.5 times the interquartile range. Suspected outliers are RTs more than 1.5 x IQR higher than the 3rd quartile (the top of the box).

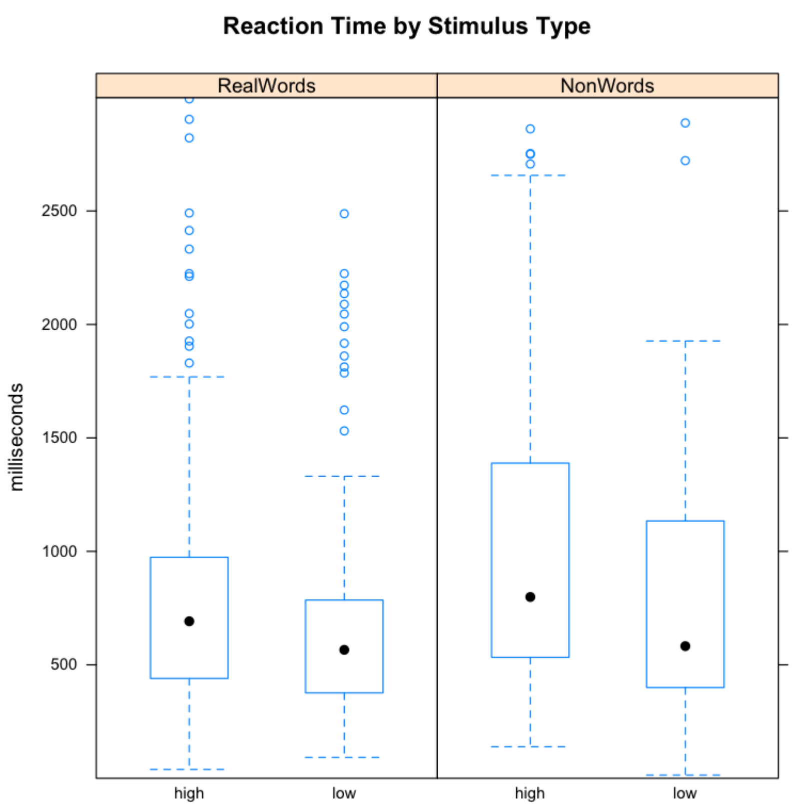


*Figure 3: Median offset-to-onset reaction times in the present experiment, as a function of stimulus type.*

Because the ANOVA suggested that a main effect of phonotactic probability/neighborhood density was the only factor influencing response time, a post-hoc, two-tailed *t*-test using a Welch correction for unequal variances was conducted, comparing mean RTs for high vs. low probability stimuli (that is, words and nonwords were collapsed together). Despite the huge variability in response times, this test was highly significant ($t(368.7) = 2.42$, $p = .01$, mean for high probability stimuli = 993.8 ms, mean for low probability stimuli = 807.1 ms), indicating that the main effect of phonotactic probability/neighborhood density revealed in the ANOVA was highly unlikely to have been observed by chance. A box and whisker plot depicting response times to high vs. low probability stimuli is shown in Figure 4.

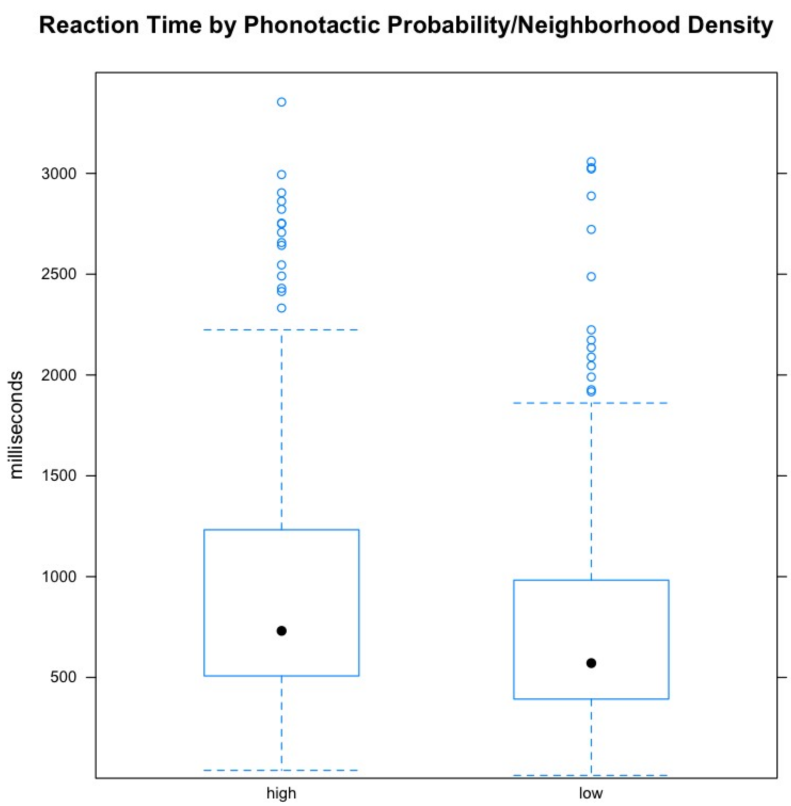


*Figure 4: Reaction time as a function of high vs. low phonotactic probability/neighborhood density in the present experiment.*

## Discussion

In the present experiment, contrary to the findings of Munson et al. (2005), no effects of phonotactic facilitation were observed. However, there was a significant main effect of phonological neighborhood density; both words and nonwords from dense areas of the lexicon elicited significantly slower response times. On the one hand, this suggests that the phonotactic facilitation effect found by Munson et al. may have been attributable to production, and not perception. Moreover, the present findings indicate that young children process all incoming stimuli on the lexical level; whether this is due to a word-bias or a simple lack of sublexical units available for processing during speech perception is currently unknown. What is clear is that lexical competition appears to be the operative force in young children's speech perception.

Within the acquisition literature, it has long been proposed that phonemic awareness is "emergent"; that is, that children's early language processing is based on whole word-sized units, and that segment-based processing emerges over time as children begin to create abstract phonemic categories based on their distributions within words. This idea was articulated as early as Reichling (1935, cited in Vihman & Croft, 2007) and has been repeatedly invoked in case studies of individual children's acquisition of phonological systems. More recently, Vihman and Croft's (2007) "radical templatic phonology" has presented a more formally argued, elaborated version of the emergent phoneme hypothesis as a way of thinking about cross-linguistic variation in adult phonological systems. Vihman and Croft argue that analyzing phonemes as categories that emerge from abstraction over a language-specific lexicon and that incorporate knowledge about the positional distribution of sounds within words can shed light on many of the "problems" in current phonological theory.

The idea that phonemic categories gradually emerge over time has also been gaining traction within the phonetics and psycholinguistics literature, and accords nicely with the present findings. Pierrehumbert (2003) posits that phonemic categories may be initiated during infancy as babies become sensitive to the statistical regularities in sound distributions, but that these categories only become robust as information about positional contrasts becomes available through generalization over the growing lexicon. This hypothesis has also been supported by experimental work carried out by Munson and colleagues (Edwards, Beckman, & Munson, 2004; Munson, Edwards, & Beckman, 2005), who argued that young children with larger vocabularies are better at accurately shadowing novel nonsense words because their larger lexicons have allowed them to create more robust, context independent generalizations about sublexical units.

To our knowledge, however, this is the first study that has demonstrated in an online processing task that children make sense of unknown speech stimuli by activating lexical items in memory. These findings indicate that young children's speech perception is fundamentally different from that of infants and adults; low-level phonetic processing may be essentially discontinued when infants' attention shifts to assigning meaning to lexical items, later being reimplemented as lexical competition necessitates finer grained phonological representations (Werker & Stager, 2000; Pierrehumbert, 2003). We conclude that young children both possess and privilege word-sized units in online processing, and that competition among lexical items is a plausible driving force for the eventual emergence of robust phonological categories.

## References

Edwards, Beckman, & Munson (2004). "The Interaction Between Vocabulary Size and Phonotactic Probability Effects on Children's Production Accuracy and Fluency in Nonword Repetition." *Journal of Speech, Language, and Hearing Research* 47(2), 421-436.

Jusczyk, Luce, & Charles-Luce (1994). "Infants' Sensitivity to Phonotactic Patterns in the Native Language." *Journal of Memory and Language* 33, 630-645.

Munson, Edwards, & Beckman (2005). "Relationships between Nonword Repetition Accuracy and other Measures of Linguistic Development in Children with Phonological Disorders." *Journal of Speech, Language, and Hearing Research* 48(1), 61 – 78.

Munson, Swenson, & Manthei (2005). "Lexical and Phonological Organization in Children: Evidence from Repetition Tasks." *Journal of Speech, Language, and Hearing Research* 48(1), 108-124.

Pierrehumbert, J. (2003). "Phonetic Diversity, Statistical Learning, and Acquisition of Phonology." *Language and Speech* 46(2-3), 115-154.

Storkel & Lee (in press). "The Independent Effects of Phonotactic Probability and Neighborhood Density on Lexical Acquisition by Preschool Children." *Language and Cognitive Processes*.

Vihman & Croft (2007). "Phonological Deveopment: Toward a "Radical Templatic Phonology". *Linguistics* 45(4), 683 – 725.

Vitevitch & Luce (1998). "When Words Compete: Levels of Processing in Perception of Spoken Words." *Psychological Science* 9(4), 325-329.

Vitevitch & Luce (1999). "Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition." *Journal of Memory and Language* 40, 374-408.

Werker & Stager (2000). "Developmental Changes in Infant Speech Perception and Early Word Learning: Is There a Link?" In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, ed. Michael B. Broe and Janet B. Pierrehumbert, 181-193.