Lexical Neighborhood Effects in Phonetic Processing

Rochelle S. Newman, James R. Sawusch, and Paul A. Luce
State University of New York at Buffalo

Previous research on spoken word recognition has demonstrated that identification of a
phonetic segment is affected by the lexical status of the item in which the segment occurs.
W. F. Ganong (1980) demonstrated that a category boundary shift occurs when the voiced
end of 1 voice-onset time continuum is a word but the voiceless end of another series is a
word; this is known as the "lexical effect." A series of studies was undertaken to examine how
lexical neighborhood, in contrast to lexical status, might influence word perception. Pairs of
nonword series were created in which the voiced end of 1 series had a higher frequency-
weighted neighborhood density, whereas the reverse was true for the other series. Lexical
neighborhood was found to affect word recognition in much the same way as lexical status.

Over the past few years, a number of studies have examined the "lexical effect" in spoken word recognition (Burton, Baum, & Blumstein, 1989; Connine & Clifton, 1987; Fox, 1984; Ganong, 1980; McQueen, 1991; Miller & Dexter, 1988; Pitt & Samuel, 1993). The lexical effect refers to the finding that perception of an ambiguous phonetic segment is affected by the lexical status of the spoken word in which the segment occurs. For example, a specially modified segment that is neither clearly a /b/ nor a /p/ will tend to be perceived as a /b/ when it is followed by /if/ (making the real word *beef* as compared to the nonword *peef*) but as a /p/ when it is followed by /is/ (making the real word *peace* as compared to *beace*). The lexical effect is often construed as evidence for the influence of higher level lexical information in phonetic perception.

In a typical experiment, listeners are presented with series of stimuli ranging from *beef* to *peef* and from *beace* to *peace*. The listeners are asked to classify the initial sound as being either a /b/ or a /p/. Listeners are more likely to classify intermediate, ambiguous stimuli in the *beef-peef* series as starting with /b/ and, conversely, to label intermediate stimuli in the *beace-peace* series as beginning with

/p/. In other words, there is a shift in the category boundary between the two series such that more items are classified as being members of whichever category makes them a real word.

Lexical effects (first reported by Ganong, 1980) were initially thought to reflect "postlexical" influences on phonetic categorization. It was hypothesized that the listeners often determined the identity of the phoneme only after having identified the word. In order to test this interpretation, Fox (1984) examined listeners' data in terms of their speed of response. He partitioned the data into slow, intermediate, and fast response sets. He found evidence for the lexical effect in the slower responses but not in the fast. Fox argued that when listeners respond quickly, their phonetic judgments are unaffected by lexical information. When listeners respond more slowly, however, processing proceeds to the point where an ambiguous item may match a lexical item, thus producing a lexical effect. Fox also suggested that the bias toward real words is not a simple response bias, because this would presumably occur regardless of the response time.

Connine and Clifton (1987) demonstrated distinct differences between the pattern of results found with a lexical bias and those found with a standard postperceptual bias. They asked listeners to classify the first phoneme in the series *dice-tice* and *dype-type* and found a category boundary shift similar to that found by Ganong (1980). They then examined the reaction times (RTs) both at each listener's category boundary (where the stimuli were ambiguous acoustically) and at the endpoints of the series (where the stimuli were unambiguous). They found that bias-consistent responses ("d" on the *dice-tice* series and "t" on the *dype-type* series) were significantly faster than bias-inconsistent responses at the category boundary but not at the endpoints. The authors compared these results to those found with a postperceptual bias (monetary reward and penalty contingencies). Here they found an RT advantage for bias-consistent responses both at the category boundary and at the endpoints of the continua. This difference supports the claim that the lexical bias is not simply a postperceptual bias

but may instead be the result of interactive processes between perceptual recognition and lexical knowledge.

Connine, Titone, and Wang (1993) showed that the lexical effect is also influenced by word frequency. They created series in which both endpoints were words but one had a higher frequency than the other. An example of this would be the series *best–pest*, where *best* is a much more common word than *pest*. In half of their series, the voiced end (e.g., *best*) was of higher frequency, whereas in the other half the voiceless end (e.g., *pest*) was of higher frequency. They found that listeners were more likely to classify ambiguous items as being an instance of the more frequent word. This finding demonstrates that lexical influences on classification are not limited to distinctions between words and nonwords but also involve information about the likelihood of the word.

Other research has suggested that the lexical effect is less pervasive than these studies might imply. Burton et al. (1989), using synthetic speech, reported a lexical effect with /d/–/t/ series. But they also found that if they used higher quality, naturally produced stimuli, the effect disappeared altogether, which suggests that the earlier results may have been the result of poor stimulus quality. When the acoustic properties of the stimuli are ambiguous, it is not surprising that top–down influences, such as lexical status, have a stronger effect. Burton et al. claimed that because the lexical effect disappeared with higher quality speech, it was unlikely to be an influence in normal perception. This argument assumes, of course, that "normal" listening conditions are analogous to the perception of isolated syllables produced in a quiet listening environment.

Recent results, however, suggest that stimulus quality may not have been the only factor contributing to Burton et al.'s (1989) null results. Pitt and Samuel (1993) reviewed the variability in lexical effects that had been reported and replicated some of this variation themselves. In spite of controlling factors such as the frequency of the word endpoints and their use of high-quality stimuli similar to those of Burton et al., they found robust effects with a /g/–/k/ series but little or no effect with a /d/–/t/ series. They suggested that for some reason, the effect was simply harder to obtain with /d/–/t/ series than with other voice-onset time (VOT) continua. However, this leaves some aspects of the lexical influence on phoneme recognition still to be explained. After all, there is no obvious reason why the /d/–/t/ stimuli should be unique in not showing a lexical influence (see, however, Paradis & Prunet, 1991a).

Much of the research on the lexical effect has attempted to explore the issue of interactive versus autonomous processing (see Connine & Clifton, 1987; McQueen, 1991; Pitt & Samuel, 1993). Thus, for instance, Connine and Clifton (1987) distinguished the lexical bias from a postperceptual bias as a way of supporting an interactive model of language comprehension. A hierarchical view would suggest that lexical information cannot directly influence perceptual analysis and thus would act in a way similar to other postperceptual biases. This issue has been difficult to resolve (see Pitt & Samuel, 1993) and is not our primary focus.

In all of the previous studies of the lexical effect, two potential sources of information from the mental lexicon have been available to listeners: lexical status and lexical neighborhood. Whereas lexical status has been the primary focus of these previous studies, the potential role of lexical neighborhoods in phoneme perception has been largely ignored. If one end of a series is similar to many words (has a high-density neighborhood), then listeners might respond to ambiguous stimuli with the label that corresponds to that end of the series, especially when these neighbors are high in frequency. Effects of neighborhood density on phonetic categorization are predicted by a number of models that propose multiple activation of lexical items in word recognition (see Luce, Pisoni, & Goldinger, 1990; McClelland & Elman, 1986).

A great deal of research in the past few years has examined the role of neighborhoods in auditory perception (Bard & Shillcock, 1993; Goldinger, Luce, & Pisoni, 1989; Luce, 1987; Luce et al., 1990). The general consensus seems to be that multiple lexical candidates compete with one another during word recognition (although there are a number of different definitions of which words in memory are likely to compete; see Bard & Shillcock, 1993, for a discussion of these differences). The parallel activation of these candidates has been shown to influence perception in a number of tasks, including word identification in noise (Luce, 1987), lexical decision (Luce, 1987), word naming (Luce, 1987), primed identification (Goldinger et al., 1989), and gated identification (Marslen-Wilson, 1990). Given this wealth of findings regarding the effect of lexical neighborhoods, it would not be surprising to find that neighborhoods can affect phonetic processing as well. However, most of the prior research involved word identification, whereas the lexical effect is found in phoneme identification for words and nonwords.

If lexical neighborhood were to have an effect on phoneme identification, then it is possible that part or all of the variability in prior research on the lexical effect could be explained by a failure to control for the neighborhoods of the items. When we examined the literature, we noticed a general tendency for studies that found large lexical effects to have used stimuli in which the word end of each series had a higher density of high-frequency lexical neighbors than the nonword end. An example is shown in the top panel of Table 1. The two series are *gift* to *kift* and *giss* to *kiss*. The neighborhood frequency of a target, shown for each end of the series, is the sum of the logarithms (base 10) of the word frequency (times 10) for each of the words that is similar to the target. We describe this computation of the frequency-weighted neighborhood density in more detail in the Method section. Simply note that the end of each series that is a word, shown in bold, also has the higher frequency-weighted neighborhood density, also in bold. For convenience, we refer to this frequency-weighted neighborhood density as the neighborhood frequency.

In contrast, those series that produced either a variable or no lexical influence had stimuli with a reversal of this pattern. In these series, the nonword ends of the two series had a higher neighborhood frequency than the word ends.

Table 1

Comparison of Lexical Status and Frequency-Weighted Neighborhood Density

Series	Frequency-weighted neighborhood density	
	Voiced end of series	Voiceless end of series
Lexical effect		
gift–kift	**11.1**	10.7
giss–**kiss**	26.5	**30.6**
No lexical effect		
deep–teep	31.1	**49.4**
deach–**teach**	**29.8**	28.3

Note. The end of each series that is a word is shown in boldface, as is the higher frequency-weighted neighborhood density for each series.

An example is shown in the bottom half of Table 1. Here, the word ends of the series *deep* to *teep* and *deach* to *teach* have a lower neighborhood frequency than their nonword counterparts. Interestingly, Pitt and Samuel (1993) noted that the lexical status effect was most often variable or absent for alveolar (/d/–/t/) series. Our analysis of the neighborhoods for stimuli used in previous experiments showed mismatches of the lexical status and neighborhood frequency in all of the /d/–/t/ series used in earlier studies but in none of the other series. Thus, at least part of the variability observed in previous studies may be due to uncontrolled variation in the frequency-weighted density of the lexical neighborhoods for the ends of the series.

Our experiments were designed to investigate possible effects of lexical neighborhood on phonetic categorization. Would listeners categorize an ambiguous item as belonging to the phonetic category that makes it more wordlike (or similar to more high-frequency words) even if none of the experimental items were themselves words? For example, the stimuli could be CVC (consonant–vowel–consonant) nonword syllables that vary within a series from an initial voiced stop (i.e., /b/) to a voiceless stop (i.e., /p/). A pair of such series would differ only in their final consonants such that for one series, the voiced (/b/) end would have a higher neighborhood frequency whereas for the other series, the voiceless (/p/) end would have a higher neighborhood frequency. According to neighborhood-based theories of word recognition, each item in the series should activate those words in the lexicon to which it is similar. Ambiguous items from near the category boundary will activate words from the neighborhoods of both endpoints of the series. That is, a syllable that is ambiguous between *beysh* (/beʃ/) and *peysh* (/peʃ/) will activate the neighbors of the voiced end of the series (such as *bake, base, bash,* etc.) as well as those of the voiceless end (*pace, paid, posh,* etc.). However, more of these neighborhood items will be similar to one endpoint than to the other. This should lead to higher activation for the initial phoneme that matches the greater number of more frequent neighbors and should result in a greater number of responses for the phoneme from the higher frequency neighborhood.

Experiment 1

Our goal in Experiment 1 was to examine whether lexical neighborhood influences phonetic perception. Specifically, could we find effects of lexical neighborhood similar to those found for lexical status? In this experiment the stimuli were pairs of VOT series in which the neighborhood frequency in one series was higher for the voiced end but in the other series was higher for the voiceless end. We chose the voicing distinction for a number of reasons. First, it is the distinction most often used in prior research on the lexical effect. Second, VOT series can be created by editing natural speech tokens, which produces highly intelligible, natural sounding stimuli.

Three sets of four CVC nonwords were selected. One of these sets is shown on the left in Table 2. Two members of each set (one voiced, and one voiceless) occurred in higher density neighborhoods, the other two in lower density neighborhoods. We computed neighborhoods by comparing each nonword to real words in an on-line dictionary. We defined an item's neighbors as being every real word that differed in only one phoneme from the item itself: either a one-phoneme substitution or the deletion or addition of one phoneme. Although some research in the past has used similarity matrices (Goldinger et al., 1989; Luce, Goldinger, Auer, & Vitevitch, 1996), there are several reasons why we chose this alternative method of neighborhood calculation. First, researchers have generally examined similarity matrices by placing the stimuli in white noise. This leads to a particular pattern of confusion data that may not generalize to stimuli presented in quiet. Second, we wished to use a female speaker to create our stimuli. Her higher fundamental would allow us to make a series with smaller acoustic differences between stimulus items. However, the similarity scaling was done on a male voice, and we do not know whether similarity scaling is consistent across talkers. Consequently, we chose the one-phoneme change approach because it seemed more appropriate to use a metric that is putatively talker independent than to use a measure whose relation to our talker's voice is unknown.

After we had determined the number of neighbors for each item, we weighted the neighbors by their log-transformed frequencies and summed them to yield a frequency-weighted neighborhood density.[1] We computed the frequency-weighted neighborhood density using all of the neighborhood words with a familiarity index of 6.0 or greater (Nusbaum, Pisoni, & Davis, 1984). We included only those words with a familiarity index of at least 6.0 in order to make sure our calculations were not based on

---

[1] For each word in a neighborhood, the influence of word frequency was computed as the logarithm (base 10) of the word frequency times 10. We multiplied the raw frequency (from Kucera & Francis, 1967) of each item by 10 so that items with a frequency of 1 would not end up with a log of 0 (and thus have no influence on the weighted frequency). If the frequency of a word in the on-line dictionary was 0, it was replaced with 1 before the computation.

Table 2
Computation of Frequency-Weighted Neighborhood Density

Series	Frequency-weighted neighborhood density		Target: gice (/gaIs/) neighborhood			
	Voiced	Voiceless	Word	Frequency	$\log_{10}$ (Frequency $\times$ 10)	Familiarity
gice–kice	**32.5**	24.9	dice	14	2.15	7.0
gipe–kipe	20.7	**27.5**	gas	98	2.99	7.0
			geese	3	1.48	7.0
			goose	4	1.48	7.0
			guess	56	2.75	7.0
			guide	36	2.56	6.5
			guise	26	2.41	7.0
			guy	51	2.71	7.0
			ice	45	2.65	7.0
			lice	2	1.30	6.7
			mice	10	2.00	7.0
			nice	75	2.88	7.0
			rice	33	2.52	7.0
			vice	42	2.62	6.8
			$\Sigma\log_{10}$ (Frequency $\times$ 10) = 32.5			

Note. The greater frequency-weighted neighborhood density in each series is shown in boldface.

neighbors that were unlikely to be in our listeners' lexicons. By weighting the words according to their frequency, we ensured that a neighbor that appeared very frequently in the language contributed more to the lexical neighborhood than did a rare word. This method of computing frequency-weighted neighborhood density has been previously described by Luce (1987). The neighborhood for the nonword *gice* (/gaIs/) is shown on the right side of Table 2 along with the raw frequency, log frequency, and familiarity of each word in the neighborhood. The frequency-weighted neighborhood density for *gice*, along with the computational formula, is shown at the bottom of Table 2. The frequency-weighted neighborhood densities for each of the four nonword endpoints used in these series are shown on the left side of Table 2. The *gice* and *kipe* ends of the two series have the greater neighborhood frequencies.[2] The target item (*gice*) has neighbors based on phoneme substitutions (*mice, gas, guide,* etc.) and deletions (*guy, ice*). However, there are no neighbors for *gice* that involve the addition of a phoneme. On the other hand, the target *gipe* does have a neighbor based on phoneme addition (*gripe*).

Using the on-line dictionary, we found pairs of VOT series with the largest possible differences in neighborhood frequency. We also specifically chose series in which the neighborhood bias was approximately symmetric. That is, the bias toward the voiced end of one series was approximately equal in size to the bias toward the voiceless end of the other series. Two additional criteria were used in choosing series. First, the final consonant could not be the same as either of the two initial consonants. Second, the final consonants of the two series in a set could not contrast in voicing. We used these two criteria in order to avoid situations in which a difference in the category boundary between two series could be caused by the use of a final consonant (or its voicing distinction) as a referent for judg-

ment of the initial consonant (see Simon & Studdert-Kennedy, 1978, or Sawusch, 1986).

In addition, we avoided pairs of series in which one series ended with an /l/, /r/, or nasal. These phonemes tend to color the vowel, so series involving them would differ in both vowel and final consonant. From a more practical standpoint, it would be almost impossible to cross-splice the final consonants in these series. There would be no point in a CV/r/ series where we could be assured of removing all (or most) information about the /r/ without removing all or most of the vowel. In order to avoid these complications, and to make editing of natural speech feasible, we decided to

---

[2] It should be noted that this is an approximate measure of neighborhood frequency for several reasons. First, it is impossible to be perfectly sure what words our listeners know. Although we can be fairly sure that they do know the words included in the neighborhood counts (because we only use words with rated familiarities of 6.0 or better), they may know additional words that are not part of our neighborhood calculations. Also, they may know slang terms, proper nouns, or other items that would not be listed in our dictionary yet would still have an influence as neighbors. Second, our method of choosing neighbors (one-phoneme substitution, deletion, or addition) is based on phonemes, not features, and thus does not take into account the similarity among various phonemes. This means that according to our calculations, both *tote* and *wrote* are equal neighbors to *coat*, even though the former differs from *coat* only in place of articulation whereas the latter differs in place, manner, and voicing. Similarly, phonetic confusability has also not been taken into account. While /w/ and /r/ are fairly confusable in some speakers' dialects, *one* is not given any greater weight as a neighbor to *run* than is *gun*. Unfortunately, there is no easy way around these difficulties. If we are able to find effects of lexical neighborhood despite these obvious problems, it would suggest that these neighborhood influences are fairly robust, even if they are small.

simply avoid CVCs ending with these consonants for the present experiments.

We generated three pairs of series corresponding to the three places of articulation of stop consonants in English. These items are shown in Table 3 along with their neighborhood frequency and density. The /b/–/p/ series involved the pairs *beyth–peyth* and *beysh–peysh* (/beθ/–/peθ/ and /beʃ/–/peʃ/). The /g/–/k/ series used *gice–kice* and *gipe–kipe* (/gaɪs/–/kaɪs/ and /gaɪp/–/kaɪp/) and the /d/–/t/ series used *dowv–towv* and *dowb–towb* (/daUv/–/taUv/ and /daUb/–/taUb/). For convenience, we refer to each pair of series by using an orthographic spelling of the voiced and voiceless endpoints with the higher neighborhood frequencies. According to this notation, our series were *beyth–peysh*, *gice–kipe*, and *dowv–towb*.

In each series, the variation in VOT was created by replacing successively longer portions of the voiced stop with the corresponding duration from the voiceless consonant. This method of cross-splicing allows us to make a natural-sounding series ranging from a voiced stop (/b/, /d/, /g/) to a voiceless one (/p/, /t/, /k/) and is the method that has been used in previous research on the lexical effect (Ganong, 1980). For all series, the final consonant from one of the two voiced stimuli (e.g., the /θ/ in /beθ/) was digitally removed and replaced by the final consonant of the other voiced stop syllable (e.g., the /ʃ/ in /beʃ/). This ensured that the acoustic–phonetic information in the beginning of the corresponding syllables in each series was the same and that the only difference between these two series of stimuli was at the end of the vowel and in the final consonant.

If we obtain a neighborhood effect for all three series, then we can attribute the general failure of previous studies to find consistent effects for /d/–/t/ series to the use of series in which the neighborhood frequency and lexical status conflicted. These opposite effects would presumably cancel one another and make any overall effect difficult to find. Alternatively, if we find a neighborhood effect for /b/–/p/ and /g/–/k/ series but not for the /d/–/t/ series, then some other factor must be involved in the lack of a consistent /d/–/t/ lexical effect, and the neighborhood results would seem to parallel lexical results. Finally, if no consistent influence of lexical neighborhood is found, then models involving the multiple activation of a set of word candidates and competition among the candidates would be open to question.

Method

Participants. The listeners were 92 undergraduate students from an introductory psychology course at the State University of New York at Buffalo who participated in the experiment for class credit. All were native speakers of English with no reported history of a speech or hearing disorder. During debriefing we discovered that 2 of our listeners were not native speakers of English; their data were not included in the analysis. Of the remaining participants, 39 listened to the *gice–kipe* series, 26 listened to the *beyth–peysh* series, and 25 heard the *dowv–towb* series. Nine listeners in the *gice–kipe* series group reported that some of the items sounded as if they began with /d/ or /t/. Because this change

in perception would presumably affect neighborhood activation, these listeners' data were omitted from the data analysis.[3] An additional 5 *gice–kipe* series listeners were excluded for failing to consistently classify one or more of the test series endpoints. Our criterion for consistent classification was a minimum of 80% correct responses on one of the two stimuli at each end of each series. One listener was dropped from the *beyth–peysh* series for not consistently classifying the endpoints. This left 25 participants in each of the three conditions.

Stimuli. A female native speaker of English (R.S.N.) recorded the syllables /gaɪs/, /gaɪp/, and /kaɪp/ in the context of running speech. She also recorded /beʃ/, /peʃ/, and /beθ/ and /daUv/, /taUv/, and /daUb/. All of the tokens were amplified, low-pass filtered at 9.5 kHz, digitized with a 12-bit, analog-to-digital converter at a 20-kHz sampling rate, and stored on computer disk. Each syllable was excised from the carrier sentence "Norton wrote _____ to me." We created an 8-item continuum ranging from /g/ to /k/ from the /gaɪs/ base by removing successively longer sections from the /g/ onset and replacing them with the corresponding sections of the /k/ (/kaɪs/) onset. The editing procedure used to produce these stimuli is essentially identical to that used by Ganong (1980) and results in a high-quality, relatively natural-sounding series. The first stimulus of the series consisted of the original /gaɪs/ syllable. We created the second stimulus by removing the /g/ release burst at the onset of /gaɪs/ and replacing it with the release burst from /kaɪs/, which resulted in a stimulus with the same VOT as the original /g/ but with the release burst of a /k/. We did all editing at zero crossings in the digital waveform to avoid audible clicks or other distortion. We made the third through eighth stimuli by removing the /g/ burst and 2, 4, 6, 8, 10, and 12 vocal pulses from the onset of the /gaɪs/ syllable and replacing them with the equivalent duration release, aspiration, and vocal pulses from the onset of /kaɪs/. The durations of the vocal pulses were not exactly equal, but averaged 4.5 ms, which resulted in a series with an approximately 9.0-ms VOT between stimuli. In the same manner, an 8-item continuum was created from the /beʃ/ base, and a 10-item continuum was created from the /daUv/ base. After pilot testing, the first stimulus of the /d/–/t/ sequence (the /d/ with the original /d/ burst) was dropped to yield a 9-item continuum. The VOT values for the stimuli in all three sets are shown in Table 4. The VOTs for the items in the three series are unequal because the stimuli are based on natural utterances, and burst and aspiration durations normally differ at the different places of articulation (see Fant, 1973, or Lisker & Abramson, 1970).

After the /gaɪs/–/kaɪs/ series was made, the /s/ portion of each syllable was removed and replaced with the /p/ from /gaɪp/. In order to make sure all of the /s/ was removed, and all of the /p/ added on, we examined spectral analyses of /gaɪs/ and /gaɪp/. The two syllables were essentially identical in their format frequencies up to the last few vocal pulses of the vowel. Consequently we included the last seven pitch pulses of the vowel along with the final segment. A released /p/ was used because it made the syllables sound more natural, and intelligible, when heard in isolation.

_____

[3] The *gice–kipe* series was the only series in which the data from such a high proportion of listeners had to be omitted. In order to investigate whether this may have altered the pattern of results, we had new tokens of these utterances recorded by the same talker and new series prepared using the same editing procedures. When listeners were tested with this series, the data for 3 (of 28) had to be omitted because of reports of other phonemes. The data for the remaining 25 listeners were essentially identical to those for the *gice–kipe* series reported in Experiment 1.

Table 3
Frequency-Weighted Neighborhood Density and Number of Neighbors for Series Used in Experiments 1–3   -

Series	Frequency-weighted neighborhood density		Number of neighbors	
	Voiced	Voiceless	Voiced	Voiceless
beysh–peysh	18.9	23.5	10	10
beyth–peyth	29.7	25.6	13	10
gice–kice	32.6	24.9	14	11
gipe–kipe	20.7	27.5	9	12
dowv–towv	11.0	7.2	4	3
dowb–towb	9.0	12.8	4	6
doish–toish	5.2	1.6	3	1
doif–toif	2.1	5.6	1	3
bowth–powth	18.4	14.9	7	6
bows–pows	19.7	24.9	9	11
bowth–powth	18.4	14.9	7	6
bowsh–powsh	4.1	9.4	3	5

Also, the released /p/ was the same duration as the /s/, so there were no duration differences between the two series.

Similarly, we altered the /beʃ/–/peʃ/ series by removing the /ʃ/ and replacing it with the /θ/ of /beθ/, and we altered the /daUv/–/taUv/ series by removing the /v/ and replacing it with the /b/ of /daUb/. We included the last seven pulses of the vowel with the final segment in the /b/–/p/ series, and the last five pulses of the vowel with the final segment in the /d/–/t/ series, again based on spectral analyses.

This resulted in six series overall. For each place of articulation there were two series that had identical acoustic values for the first phoneme and most of the vowel. For the velar series, one series ranged from /gaɪs/ to /kaɪs/ and the other from /gaɪp/ to /kaɪp/. For the labial place of articulation, one series ranged from /beʃ/ to /peʃ/ and the other from /beθ/ to /peθ/, and for the alveolar series, one ranged from /daUv/ to /taUv/ and the other from /daUb/ to /taUb/.

Procedure. Listeners were tested individually and heard only one pair of series. Stimulus presentation and response collection were controlled by a Digital Equipment Corporation PDP-11/34 computer. The stimuli, which were stored on disk, were converted to analog form by a 12-bit, digital-to-analog converter at a 20-kHz sampling rate, were low-pass filtered at 9 kHz, and were presented binaurally through TDH-39 headphones. The syllables were presented in random order. Listeners were asked to identify the initial phoneme as /g/ or /k/, /b/ or /p/, or /d/ or /t/ as quickly and accurately as possible by pressing one of two buttons on a computer-controlled response box. The mapping of response to hand was counterbalanced across listeners.

The presentation pace depended on the listeners' response speed. The next trial began 1.0 s after the listener had responded, or after an interval of 4.0 s from stimulus onset had elapsed, whichever came first. The listeners' response and RT were recorded for each stimulus. Responses from the first block of trials (three repetitions of each item) were considered practice and were not included in subsequent data analysis. After the practice set, stimuli were presented in blocks of 64 (or 72) trials (4 repetitions of each of the 16 or 18 items). All listeners participated in six blocks of experimental trials, which resulted in a total of 24 responses to each stimulus.

Results and Discussion

First, the overall data for each pair of series were examined. The percentage of voiced (/b/, /d/, or /g/) responses was determined for each stimulus in each series for every listener. Any response with an RT greater than 1,500 ms was eliminated. The data from each listener were examined for consistency in classifying the endpoints of their two series. Any listener who could not classify the endpoints consistently was dropped from the analysis, as described previously.

Table 4
Voice-Onset Times (in Milliseconds) for the Stimuli in Experiment 1

Series	Stimulus								
	1	2	3	4	5	6	7	8	9
beyth–peysh	7.2	7.2	15.2	26.0	33.0	42.0	49.6	49.6	
gice–kipe	21.5	21.5	30.2	39.4	48.2	54.3	54.3	54.3	
dowv–towb	8.7	16.3	25.5	34.7	44.2	53.6	63.0	72.4	81.9

The category boundary for each listener was then determined for each series by linear interpolation between the two stimuli on either side of the boundary. We chose this method of determining category boundaries because it is the method frequently used in earlier studies. We also tabulated the total percentage of voiced responses given by each listener to all of the stimuli in each series. Whereas the movement of the category boundary should indicate changes in the perception of ambiguous stimuli, the overall percentage of voiced responses to the series as a whole should include any changes away from the boundary of the series as well as those at the boundary. Thus, as noted by Samuel (1986) with respect to selective adaptation data, this overall percentage measure may be a more sensitive index of changes in perception (see also Pitt & Samuel, 1993). We conducted two paired t tests on the data (one using category boundaries and the other using percentages of voiced responses) to compare responses across the two series. Because we were predicting an effect in a specific direction (more responses for the nonword with a denser neighborhood), we used one-tailed tests.

We then partitioned data for each listener into three subsets based on RT. This partitioning was performed for several reasons. First, this analysis has frequently been done in the literature on the lexical effect, so it was necessary for comparison with prior research. Although the neighborhood effect we are investigating is a component of the lexical effect that has previously been reported, the time course of the neighborhood effect may not be the same as that of the lexical effect. Second, prior research has shown that the effect does not appear in all RT ranges. The lexical effect generally appears in the slow, or slow and intermediate, RT partitions but not in the fast responses (see Fox, 1984). To the extent that an effect is small and present in only some of a listener's responses, overall analyses can mask an underlying effect. Given that the effects we are looking for are likely to be quite small, breaking the responses into RT partitions may be the most sensitive approach to assessing neighborhood effects on phoneme perception. Thus, the partitioning by RTs is designed to allow us to find faint, transient effects that may be too weak to show up in an overall analysis. Finally, we do not expect that the different series will necessarily show the same pattern across RT partitions. Because listeners may respond faster overall to one series than to another, the analogous effects may appear in the intermediate partition for one series but only in the slow partition for another. The important issue is the presence or absence of an effect of lexical neighborhood on phoneme identification.

We assumed that each listener's RTs to each individual item in each series would approximate a normal distribution, and we divided this distribution into three equal portions. For each listener's responses to each stimulus, we determined the mean and standard deviation and then translated each individual RT into a z score. Scores of .43 and $-.43$ divide the distribution into three equal portions. Any RT equivalent to a z score less than $-.43$ was considered fast, any RT greater than a z score of .43 was considered slow, and any RT equivalent to a z score between $-.43$ and

.43 was considered intermediate. Other researchers have either divided each listener's responses into thirds (Burton et al., 1989; Miller & Dexter, 1988) or introduced RT ranges such as 0–500 ms, 500–800 ms, and >800 ms (Fox, 1984). We chose the method used here because of its sensitivity to the variability of RTs within each individual. Once each listener's data had been partitioned, identification functions, category boundaries, and percentages of voiced responses were determined for each series in each partition as described previously for the overall analysis. We used two one-way analyses of variance (ANOVAs) with RT partition as the single factor to examine the percentage of voiced responses and category boundary data. These were followed by paired t tests (planned comparisons) in each RT partition, as described previously for the overall analysis of each set of data.

We first examined the data from the gice–kipe series. Overall, there was a significant shift in the category boundary, $t(24) = 1.91$, $p < .05$, and in the percentage of "g" responses, $t(24) = 2.09$, $p < .05$. Listeners gave more voiced, "g" responses to the gice–kice series and more voiceless, "k" responses to the gipe–kipe series. This difference is consistent with the difference in neighborhood frequency. That is, listeners gave more responses to each series that were consistent with the endpoint that had the higher neighborhood frequency (gice and kipe). Figure 1 displays the mean identification functions for these two series across listeners and shows a small change in the locus of the category boundary for the two series.

The results of partitioning the RT data are shown in Figure 2. One-way ANOVAs for category boundary and percentage of voiced responses showed no significant effect of RT partition: $F(2, 48) = 1.66$, $p < .10$, and $F(2, 48) < 1.0$, respectively. In the planned comparisons, no significant effects of neighborhood frequency were found in the fast partition: $t(24) = 0.42$ by category boundaries; $t(24) = 0.93$ by percentages; both $ps > .10$. As shown in the top panel in Figure 2, the two fast RT partition classification functions lie on top of one another. The intermediate RT partition showed no effect of neighborhood frequency on the cate-
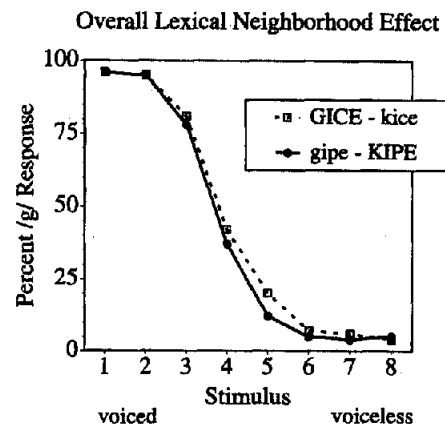

Figure 1. Group identification functions for the gice–kice and gipe–kipe series.
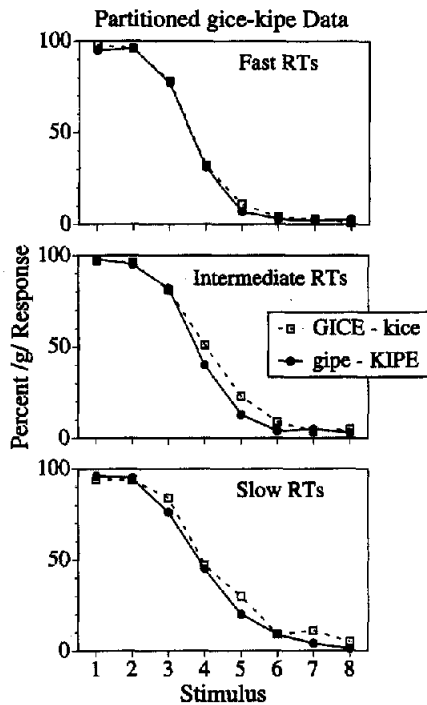
Figure 2. Group identification functions for the *gice–kice* and *gipe–kipe* series in three reaction time (RT) partitions: fast, intermediate, and slow.

gory boundaries, $t(24) = 1.19$, $p > .10$, but a marginal effect on the percentages, $t(24) = 1.47$, $p < .10$. In the slow partition there was a significant difference in the location of the category boundaries, $t(24) = 2.41$, $p < .05$, but only a marginal difference in percentages, $t(24) = 1.68$, $p < .10$. Inspection of Figure 2 shows the effect of neighborhood frequency on classification of the /g/–/k/ distinction in the middle and lower panels.

We examined the data for the other two series in the same manner. Figure 3 shows the overall results from the *beyth–peysh* series across listeners. As with the /g/–/k/ series, we determined the /b/–/p/ category boundary for each listener in each series and also tabulated the total percentages of "b" responses given by each listener to all of the stimuli in each series. There was a small, but reliable, effect of neighborhood frequency, as shown by a significant shift in the category boundary, $t(24) = 2.48$, $p < .05$, and a significant change in the percentage of "b" responses, $t(24) = 2.72$, $p < .01$. As in the *gice–kipe* series, the change in the category boundary and the percentage of "b" responses was to favor the end of the series with the higher neighborhood frequency (*beyth* and *peysh*).

We then partitioned the data into three subsets based on RT in the same manner as for the *gice–kipe* series. The one-way ANOVAs showed significant main effects of RT partition for both the category boundary data, $F(2, 48) = 3.89$, $p < .05$, and the percentage voiced data, $F(2, 48) = 5.65$, $p < .01$. As shown in the top panel of Figure 4, planned comparisons revealed no significant effects of

neighborhood frequency in the fast partition: $t(24) = 0.25$ for category boundaries and $t(24) = 0.02$ for percentages, both $ps > .10$. There was a significant effect of neighborhood frequency in the intermediate partition, both for the category boundary data, $t(24) = 4.31$, $p < .001$, and the percentage of "b" responses, $t(24) = 4.35$, $p < .001$. In the slow partition, there was a marginal effect of neighborhood frequency on the category boundaries, $t(24) = 1.37$, $p < .10$, but no effect on the percentage of "b" responses, $t(24) = 0.07$, $p > .10$. These data are shown in the bottom panel of Figure 4.

The *dowv–towb* series yielded somewhat different results. Figure 5 displays the overall identification functions for the two series across listeners. As for the other two series, we determined the /d/–/t/ category boundary for each listener and also tabulated the total percentage of "d" responses given by each listener to all of the stimuli in each series. There were no significant changes in either the category boundary data, $t(24) = -0.69$, $p > .10$, or in the percentage of "d" responses, $t(24) = -0.22$, $p > .10$.

We then partitioned the *dowv–towb* data into three sets based on RT in the same manner as before. These results are shown in Figure 6. One-way ANOVAs on the data yielded a mixed picture, with a significant effect of RT partition for the category boundary data, $F(2, 48) = 3.55$, $p < .05$, but no significant effect in the percentage voiced data, $F(2, 48) < 1.0$. We found a significant effect of neighborhood frequency on the category boundary in the intermediate speed condition (middle panel) but no other significant or marginal effects: $t(24) = -1.77$ by categories, and $t(24) = 0.86$ by percentages, both $ps > .10$ in the fast RT partition; $t(24) = 1.78$, $p < .05$ by categories, and $t(24) = -0.61$, $p > .10$ by percentages in the intermediate RT partition; and $t(24) = -1.59$ by category, and $t(24) = -0.51$ by percentages, both $ps > .10$ in the slow RT partition. Although the significant effect in the intermediate RT partition appears suggestive, it is balanced by equally large effects, in the opposite direction, in the fast RT partition and the slow RT partition. Also, even within the intermediate partition, an opposite effect can be seen in the percentage of "d" re-
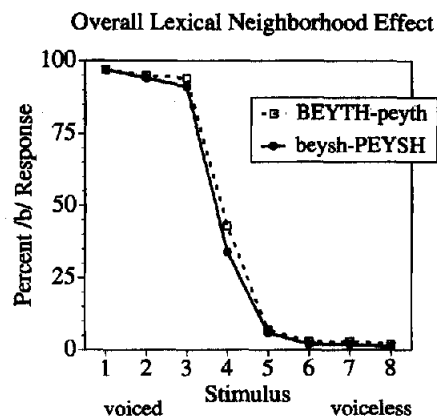

Figure 3. Group identification functions for the *beyth–peyth* and *beysh–peysh* series.
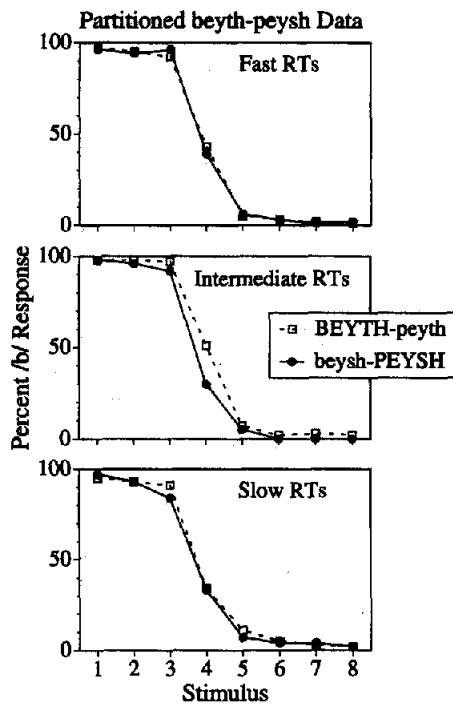
Figure 4. Group identification functions for the *beyth–peyth* and *beysh–peysh* series in three reaction time (RT) partitions: fast, intermediate, and slow.

sponses. Thus, although there may be a hint of a possible effect in the intermediate RT range, there appears to be nothing strong.

Figure 7 shows a summary of the effects of neighborhood frequency across the three series. The bars represent the difference in the overall percentages of voiced responses between the two series in each pair. A positive difference indicates that listeners were using the response that corresponded to the higher neighborhood frequency end of each series. Both the bilabial and velar series demonstrate reliable neighborhood effects, similar to the results from the lexical effects literature. For the bilabial series, the effects are concentrated in the intermediate RT partition, rather than in both the intermediate and slow partitions (or all three partitions) as is usually found for lexical effects. The importance of this is not that the effects are in the intermediate range per se, because some listeners are likely to be faster than others overall. What is "intermediate" for one listener might be "slow" for another. The more interesting aspect is that the results appear to be transient: They appear and then dissipate. Results from lexical series have always continued through the slow RTs. That is, previous accounts of lexical effects for initial stops have reported that if a lexical effect was present in one speed range, it was also present in all slower speed ranges. However, because this result is found only in the one pair of series, it is hard to make any substantive claims regarding this matter.

Together, these results for the bilabial and velar series show that the perception of an acoustic–phonetic sequence

is influenced by the number of words to which that sequence is similar. This is consistent with models of word recognition in which the acoustic–phonetic information at the beginning of a word activates a set of candidates that compete or interact with one another: a lexical neighborhood. The more neighbors an acoustic–phonetic sequence has, the higher the activation of the component phonetic units.

The alveolar series, however, gave little evidence for an effect of lexical neighborhood: There was no overall effect, and the effect in the intermediate partition was only significant in the category boundary data. Perhaps there is an effect of neighborhood here, but if so it appears to be much weaker than that found in the other series. This corresponds to the results found in the lexical studies, where the lexical effect was difficult to find with the alveolar place of articulation. However, given the hint of an effect in the intermediate RT partition, we felt it was necessary to examine a second /d/–/t/ series before making any strong claims regarding the uniqueness of alveolars.

Experiment 2

Our purpose in the second experiment was to provide additional evidence concerning the effects of lexical neighborhoods on the perception of the /d/–/t/ contrast. Specifically, we wished to replicate the results from the alveolar series in Experiment 1. Series ranging from *doish* to *toish* and from *doif* to *toif* (/doIʃ/–/toIʃ/ and /doIf/–/toIf/) were constructed. In these series, the *doish* and *toif* ends had higher frequency-weighted neighborhood densities (see Table 3). If a reliable effect of lexical neighborhood is observed with these stimuli, then it would suggest that all three places of articulation show evidence of neighborhood frequency influencing phonetic perception. In turn, this would be consistent with our proposal that variability in previous studies of the lexical effect with /d/–/t/ series is due to neighborhood and lexical status effects that conflict and largely cancel one another. If our second /d/–/t/ series, like
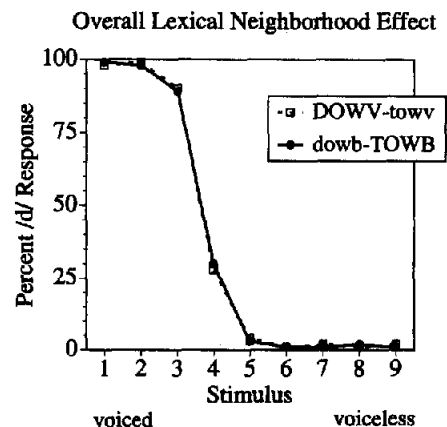

Figure 5. Group identification functions for the *dowv–towv* and *dowb–towb* series.
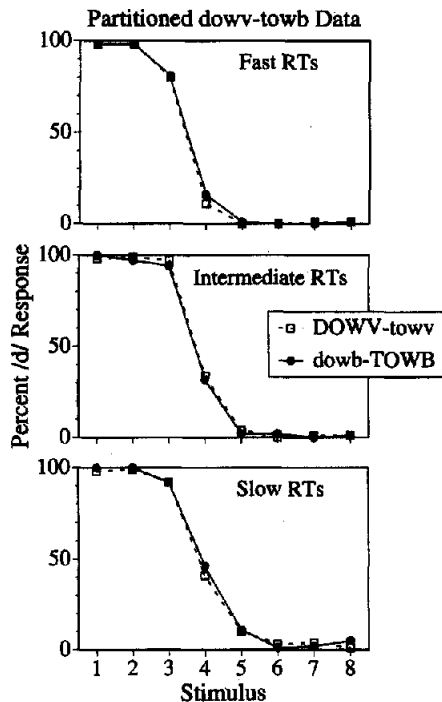
Figure 6. Group identification functions for the *dowv–towv* and *dowb–towb* series in three reaction time (RT) partitions: fast, intermediate, and slow.

the *dowv–towb*, shows no effect of neighborhoods, it would suggest that there is some other aspect of /d/–/t/ series that makes them different from labial and velar series. It would also suggest that the neighborhood effect exactly parallels the lexical effect, appearing in labial and velar series but not in alveolar ones.

Method

Participants. The listeners were 28 undergraduate students from an introductory psychology course at the State University of New York at Buffalo who participated in the experiment for class credit. All were native speakers of English with no reported history of a speech or hearing disorder. One listener's data were omitted from the experiment because he reported hearing some of the syllables as ending in /s/. A 2nd listener's data were omitted on the basis of her reporting some of the items as *dice* and *tice*. A 3rd listener's data were omitted for failure to accurately classify the endpoints of one series at 80% or better. This left 25 participants.

Stimuli. The same female native talker of English recorded the syllables /doIʃ/, /toIʃ/, and /doIf/ in the context of running speech. The tokens were amplified, low-pass filtered at 9.5 kHz, digitized with a 12-bit, analog-to-digital converter at a 20-kHz sampling rate, and stored on computer disk. Each syllable was excised from the carrier sentence "Norton wrote _____ to me." As in Experiment 1, we created a 9-item continuum ranging from /d/ to /t/ from the /doIʃ/ base by removing successive sections from the /d/ onset and replacing them with the corresponding sections of the /t/ onset. The first section consisted of the 9.8-ms /d/ burst. This yielded a stimulus with the same VOT as the /d/ end but with a burst more characteristic of a /t/ than a /d/. Subsequent sections consisted of

the /d/ burst plus 2, 4, 6, 8, 10, and 12 vocal pulses. The durations of these vocal pulses were not exactly equal, but were quite close, averaging around 3.9 ms. The editing process resulted in a series with VOTs of 9.8, 9.8, 16.7, 25.0, 32.5, 39.6, and 45.5 ms for each of the last three items. (These last items had the initial vocal pulses of the /t/ replacing those of the /d/; they differed in the amount of the /d/ that was replaced with the /t/ but did not differ in VOT.)

After this series was made, the /ʃ/ portion of each syllable was removed and replaced with the /f/ from /doIf/. In order to ensure that all of the /ʃ/ was removed, and all of the /f/ added on, we examined a spectral analysis of each syllable. On the basis of these analyses, we included the last six pitch pulses of the vowel along with the final frication. This resulted in two series, one ranging from /doIʃ/ to /toIʃ/ and the other from /doIf/ to /toIf/, that had identical acoustic values for the initial phoneme and most of the following vowel.

Procedure. The procedure was identical to that in Experiment 1. Listeners were tested individually and were asked to identify the initial phoneme as quickly as possible by pressing a button on a computer-controlled response box. Responses from the first block of 54 trials were considered practice and were not included in subsequent data analysis. After the practice set, stimuli were presented in blocks of 72 trials (4 repetitions of each of the 18 items), and listeners received six blocks. This yielded a total of 24 responses to each stimulus for each listener.

Results and Discussion

Figure 8 displays the overall identification functions for the two alveolar series across listeners. As in Experiment 1, we determined the /d/–/t/ category boundary for each listener and also tabulated the total percentages of "d" responses given by each listener to all of the stimuli in each series. Directional, paired *t* tests were conducted on the data, one using category boundaries and the other using the percentage of "d" responses. There were no significant effects of lexical neighborhood for either the category boundary data, $t(24) = -1.06$, $p > .10$, or the percentage of voiced responses, $t(24) = -1.37$, $p > .10$.

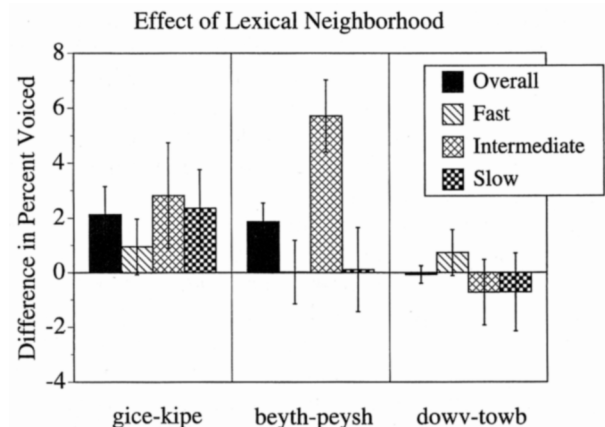We then partitioned the data into three subsets based on

Figure 7. Neighborhood frequency effect on percentage of voiced responses for *gice–kipe, beyth–peysh,* and *dowv–towb* series in both the overall data and each of the three reaction time partitions.
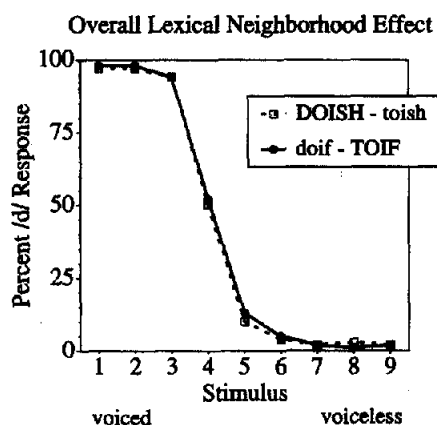
Overall Lexical Neighborhood Effect

Figure 8. Group identification functions for the *doish–toish* and *doif–toif* series.

RT in the same manner as in Experiment 1. One-way ANOVAs revealed no significant effects of RT partition in either the category boundary data or the percentage of voiced responses, both $Fs(2, 48) < 1.0$. Planned comparisons showed no significant effects of neighborhood frequency in any of the RT partitions: For category boundaries, $t(24) = 0.64$ at fast RTs, $t(24) = -1.73$ at intermediate RTs, and $t(24) = -0.82$ at slow RTs, all $ps > .10$; for percentage of "d" responses, $t(24) = 0.44$ at fast RTs, $t(24) = -1.28$ at intermediate RTs, and $t(24) = -0.37$ at slow RTs, all $ps > .10$. The classification functions in each of the three partitions were essentially identical to the overall functions shown in Figure 8.

The lack of any effect of neighborhood frequency on the *doish–toif* series is consistent with the minimal effects of lexical neighborhood on the *dowv–towb* series in Experiment 1. Overall, the results of Experiments 1 and 2 with nonwords parallel those of the word–nonword studies remarkably well. With /d/–/t/ series, there is little or no effect of neighborhood frequency. The studies of lexical status influences on phoneme perception show a similar overall null result for /d/–/t/ series (see Pitt & Samuel, 1993). In contrast, both previous word–nonword series and our own nonword series show consistent effects of lexical status and lexical neighborhood on other voicing continua.

The obvious question, then, is why the /d/–/t/ series consistently show little or no influence of higher level information. There are a number of possibilities, which we will consider in turn. One possibility has to do with the spacing of stimuli in a series and our listeners' perception of the stimuli. If the step size between stimuli in the /d/–/t/ series was perceptually larger than that between stimuli in the other series, then there may not have been any truly ambiguous stimuli in the /d/–/t/ series. If none of the stimuli were ambiguous, then the lack of a neighborhood effect would simply reflect insensitive measurement.[4] This lack of any /d/–/t/ effect could arise even if the physical (acoustic) spacing in the /d/–/t/ series was similar to that in the /b/–/p/ and /g/–/k/ series. All that is required is that the category boundary for the /d/–/t/ series be sharper (more of a step

function) than that of the other series. In a sense, this is the argument of Burton et al. (1989). If the stimuli are high in quality and natural sounding, then perceptual processing is basically data driven and no higher level influences will be found.

The average changes in VOT for the bilabial and velar series stimuli were 8.8 and 9.0 ms, respectively. The alveolar series had step sizes of 9.4 and 7.8 ms for *dowv–towb* and *doish–toif*, respectively. Thus, although the four series had slightly different step sizes, the two /d/–/t/ series had both the largest step size of the four series and the smallest. It therefore seems unlikely that differences in the step size were in any way responsible for the lack of the /d/–/t/ effect. However, the real key to this question is in our listeners' perception of the stimuli. To assess the degree to which the series did or did not contain ambiguous stimuli near the boundary, we fit a psychometric function to the data from each of the two series for each listener. The analysis used here for fitting a normal ogive to the data was described by Engen (1971). From each psychometric function, a slope was derived. This slope represents the steepness of the listener's classification function for each series. Thus, it also captures the degree to which stimuli were ambiguous. High slopes correspond to classification functions that are more steplike. Low slopes suggest that stimuli from the middle of the series are ambiguous. The mean slopes for each of our eight series, collapsed across listeners, are shown in Table 5.

As can be seen in Table 5, the slopes of the *dowv–towb* series were steeper than those of the bilabial and velar series, but those of the *doish–toif* series fall between those of the *beyth–peysh* and *gice–kipe* series. We conducted a two-way ANOVA on these slope data with the higher density end of the series (voiced vs. voiceless) and series as factors. There was a significant effect of series, but follow-up tests (Newman–Keuls) showed that none of the pairs were significantly different. Examination of the slopes produced no evidence to support the idea that the /d/–/t/ stimuli did not show a neighborhood effect because the intermediate stimuli were less ambiguous. The *dowv–towv* and *dowb–towb* series (which showed an effect of neighborhood on category boundaries in the intermediate partition) had the steepest ogives of any of the series; but the *doish–toif* series (which showed no effect at any RT partition) had an intermediate steepness, one between those of the labial and the velar series. If the steepness of the slopes caused the lack of effect for the /d/–/t/ series, we would expect that all four /d/–/t/ series would have had steeper slopes than those for the other series. Because the *doish–toif* series had less steep slopes than the *beyth–peysh* series and yet showed no effect, whereas the labial series did show a significant effect, it appears that steepness of the slope is not a factor.

A second possible reason for the general lack of lexical and neighborhood effects for /d/–/t/ series concerns the status of /d/ and /t/ in perception. These two phonemes are among the most common consonants in English. If there is

---

4 We would like to thank Dick Pastore for suggesting this possibility.

Table 5

Mean Slopes for Psychometric Functions Fit to the Listeners' Responses to the Four Pairs of Voicing Series in Experiments 1 and 2

Higher density end	beysh–peyth	gice–kipe	dowv–towb	doish–toif
Voiced	1.72	1.04	1.87	1.47
Voiceless	1.52	1.26	2.23	1.33

an influence of phoneme probability in perception, perhaps this influence overwhelms any higher level effect of neighborhood or lexical status. A related idea concerns the possibility that the default place of articulation may be coronal. This ties in to current theories of underspecificity. According to these theories, coronals may have a "default" place of articulation (see Stemberger & Stoel-Gammon, 1991, for a discussion). Because this default value will be assumed whenever there is no other information, there is no need to mark coronals for their place of articulation, and therefore these consonants are underspecified in terms of place of articulation. That is, the underlying representations for coronals do not contain any information at all about their place of articulation (see Paradis & Prunet, 1991a). It is not clear what implications underspecificity might have for theories of neighborhood activation; the details of such an approach have not been worked out sufficiently to allow us to use it in computing neighborhood frequency. Nevertheless, it is possible that the unique role of coronals in underspecification theory will have some implications for perceptual research such as that described here, especially because it has already been shown to be important for some other areas of speech research (see Stemberger & Stoel-Gammon, 1991).

A third possibility has to do with the phonotactics of our series. All of the initial consonant–vowel (CV) sequences in our series in Experiments 1 and 2 occur in American English. However, some of the vowel–consonant (VC) sequences do not occur within a syllable, and some do not occur at all (either within a syllable or across a syllable boundary within a word), in the corpus represented in our on-line dictionary. In Experiment 1, the /eʃ/ in the *beysh–peysh* series does not occur within a syllable. Similarly, neither /aUv/ nor /aUb/ from the *dowb–towb* and *dowv–towv* series nor /oIʃ/ from the *doish–toish* series occurs within a syllable. Both /eʃ/ and /aUb/ do occur across a syllable boundary, such as in the words *ratio, patience,* and *cowboy.* In addition, sequences such as /aUb/ do occur in names, such as Weintraub. If the phonotactic legality of the VC sequences in our series were influencing our data, then we would have expected the *beyth–peysh* and *doish–toif* sets to yield similar results, because both sets have one series in which the VC sequence does not occur within a syllable in American English. Because lexical neighborhood effects were found in the *beyth–peysh* set but not in the *doish–toif* set, a phonotactics-based explanation for the lack of any neighborhood effect with alveolars does not appear particularly compelling. Nevertheless, for the moment, phonotactic influences remain a possibility that we will return to later.

The final possibility that we will consider is that neighborhood frequency does have an effect on the perception of the /d/–/t/ contrast but that this effect was not observed in Experiments 1 and 2 because the sizes of the lexical neighborhoods for our /d/–/t/ nonwords were too small. Further examination of the neighborhoods for all of our series revealed an interesting finding. The /d/–/t/ series we used had fewer neighbors than did the /b/–/p/ and /g/–/k/ series. Specifically, none of the /d/–/t/ endpoints in either set of series had more than six neighbors, whereas none of the endpoints for the /b/–/p/ or /g/–/k/ series had fewer than nine. When we reexamined the /d/–/t/ series used in previous research, we found that the neighborhood densities for the /d/–/t/ word–nonword series were much higher. Perhaps there needs to be a minimum number of neighbors before a neighborhood effect can show up. Not having enough neighbors could have resulted in the null effects for the /d/–/t/ series we found in Experiments 1 and 2. In prior experiments, researchers have used series with much larger numbers of neighbors, and neighborhood frequency effects may have been present. Because these effects went in the opposite direction from the lexical effect, they may have resulted in an overall null result.

In order to test this, we searched the lexicon for a pair of /d/–/t/ nonword series with a larger number of neighbors. Unfortunately, we could not find any series in the lexicon that met our stimulus requirements. Alveolars are so common in English that there were few series where all four endpoints were nonwords. Although some /d/–/t/ nonword series with large numbers of neighbors exist, many of these end with /d/ or /t/ and thus we could not use them for fear of producing an identity contrast. Also, many series ended with /l/ or with nasals; these tend to color the vowel and make editing impossible. Given our inability to directly test this idea that the number of neighbors was playing a role in our lack of /d/–/t/ effects, we decided to look for a /b/–/p/ series with a small number of neighbors to see whether we could make the neighborhood effect disappear solely on the basis of number of neighbors. If the lack of /d/–/t/ effects was caused by the small number of neighbors, a /b/–/p/ series with an equally small number of neighbors should likewise show no effect.

Experiment 3

In this experiment, we used series ranging from *bowth* to *powth, bows* to *pows,* and *bowsh* to *powsh* (/baUθ/–/paUθ/, /baUs/–/paUs/, /baUʃ/–/paUʃ/). The largest number of neighbors for any of these endpoints was 11, and the fewest was 3. The *bows–pows* series, with 11 and 9 neighbors, should be large enough to produce an effect, because its size is comparable with those of our earlier bilabial and velar series, which did show effects. The *bowth* and *powth* endpoints have 6 and 7 neighbors, respectively. This places them at a point intermediate between the previous alveolar and nonalveolar series. The *bowsh* and *powsh* endpoints have only 5 and 3 neighbors, respectively, similar to the number of neighbors in the *dowv–towv* series. This series,

then, should not be able to produce an effect if the effect is mediated by number of neighbors in addition to differences in the neighborhood frequency. Overall, if number of neighbors is playing a role, we might expect to see a small effect when listeners hear the *bowth–powth* and *bows–pows* stimuli but no effect when they hear the *bowth–powth* and *bowsh–powsh* stimuli. The frequency-weighted neighborhood densities and the number of neighbors for the endpoints of these series are shown in Table 3.

It should be noted that the /aUʃ/ VC sequence of the *bowsh–powsh* series does not occur within American English words, either within a syllable or across a syllable boundary (based on a search of the on-line dictionary). Thus, pairing the *bowth–powth* and *bowsh–powsh* series yields a set of stimuli whose phonotactics are similar to those of the *doish–toif* set. If an effect of lexical neighborhood (series) is found for the *bowth–powsh* set, then it would seem reasonable to conclude that phonotactic legality is not the reason for our null results with alveolars.

Method

Participants. The listeners were 59 undergraduate students from an introductory psychology course at the State University of New York at Buffalo who participated in the experiment for class credit. All were native speakers of English with no reported history of a speech or hearing disorder. Twenty-nine participants listened to the *bowth–pows* series. The data of 3 listeners were omitted from the experiment after they reported hearing other syllables (2 reported hearing some VC syllables, *outh* and *ous*, and 1 reported hearing some hVC syllables, *house* and *houth*). The data of 1 additional listener were omitted for failure to reach 80% correct at the endpoints of the series. This left a total of 25 listeners for the *bowth–pows* series. Thirty participants listened to the *bowth–powsh* series. Of these, the data of 1 listener were omitted for reports of hearing syllables ending in /s/. Four additional listeners' data were omitted for failure to reach 80% correct at one or more endpoints. This left a total of 25 listeners in the *bowth–powsh* series group.

Stimuli. The same female native talker of English recorded the syllables /baUθ/, /paUθ/, /baUs/, and /baUʃ/ in the context of running speech. The tokens were amplified, low-pass filtered at 9.5 kHz, digitized with a 12-bit, analog-to-digital converter at a 20-kHz sampling rate, and stored on computer disk. Each syllable was excised from the carrier sentence "Norton wrote _____ to me." As in Experiment 1, we created a 7-item continuum ranging from /b/ to /p/ from the /baUθ/ base by removing successive sections from the /b/ onset and replacing them with the corresponding sections of the /p/ onset. The first section consisted of the 5.5-ms burst plus one vocal pulse and was replaced with the release burst from /paUθ/. Each subsequent section contained two additional vocal pulses from the /baUθ/. The durations of these vocal pulse sections were not exactly equal but were quite close, averaging 9.3 ms. The VOTs for the seven stimuli in each series were 5.5, 11.7, 20.9, 30.1, 39.2, 48.6, and 58.1 ms.

After this series was made, the /θ/ portion of each syllable was removed and replaced with the /s/ from /baUs/ and the /ʃ/ from /baUʃ/. On the basis of spectral analyses and to ensure that all of the /θ/ was removed and all of the /s/ or /ʃ/ was added on, we included the last six pitch pulses of the vowel along with the final consonant.

This resulted in three series, one ranging from /baUθ/ to /paUθ/,

a second from /baUs/ to /paUs/, and the last from /baUʃ/ to /paUʃ/, all of which had identical acoustic values in the corresponding stimuli for the initial phoneme and most of the following vowel. Listeners heard the *bowth–powth* series and either the *bows–pows* or *bowsh–powsh* series, but not both.

Procedure. The procedure was identical to that in Experiment 1. Listeners were tested individually and were asked to identify the initial phoneme as quickly and accurately as possible by pressing a button on a computer-controlled response box. Responses from the first block of 42 trials (3 repetitions of each of 14 stimuli) were considered practice and were not included in subsequent data analysis. After the practice set, stimuli were presented in blocks of 56 trials (4 repetitions of each of the 14 items), and listeners received six blocks, which resulted in 24 responses to each stimulus. A computer failure resulted in the loss of one block of data for 1 listener in the *bowth–pows* group, so 1 listener's data consisted of only 20 responses to each stimulus.

Results and Discussion

The overall identification functions for both pairs of series are shown in Figure 9. As in the earlier experiments, we determined the /b/–/p/ category boundary for each listener and also tabulated the total percentages of "b" responses given by each listener to all of the stimuli in each series. Directional, paired t tests were conducted on the data, one using category boundaries and the other using the percentage of "b" responses. The data for the *bowth–pows* group are shown in the top panel of Figure 9. There was an overall effect of neighborhood frequency in both the category boundary data, $t(24) = 4.91$, $p < .001$, and the percentage of "b" responses, $t(24) = 3.43$, $p < .005$.

We then partitioned the *bowth–pows* data into three subsets based on RT in the same manner as in Experiment 1. The one-way ANOVAs on the category boundary data and
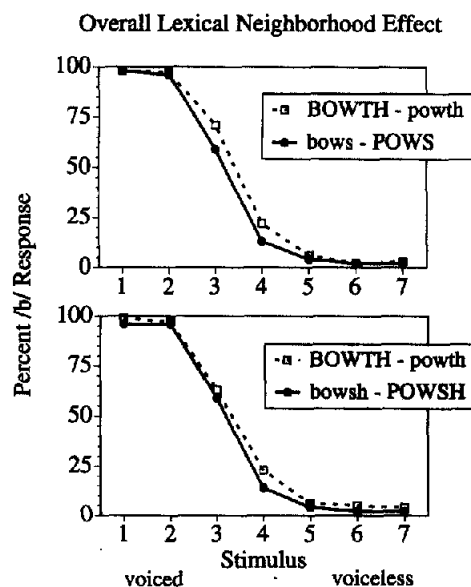

Figure 9. Group identification functions for the *bowth–powth* and *bows–pows* series and the *bowth–powth* and *bowsh–powsh* series.

percentage of voiced responses both showed no significant effect of RT partition: $F(2, 48) < 1.0$ and $F(2, 48) = 1.36$, $p > .10$, respectively. Planned comparisons revealed significant effects in both the category boundary data and the percentage of "b" responses in the fast and intermediate RT partitions but only in the category boundary data in the slow RT partition: For the fast RTs, $t(24) = 3.09$, $p < .005$ by categories, and $t(24) = 2.40$, $p < .05$ by percentages; for the intermediate RTs, $t(24) = 2.70$, $p < .01$ by categories, and $t(24) = 2.97$, $p < .005$ by percentages; and for the slow RT partition, $t(24) = 2.03$, $p < .05$ by categories, and $t(24) = 1.18$, $p > .10$ by percentages.

The overall results from the bowth–powsh series stimuli are shown in the bottom panel of Figure 9. There were significant effects of neighborhood frequency in both the category boundary data, $t(24) = 2.26$, $p < .05$, and in the percentage of "b" responses, $t(24) = 3.09$, $p < .005$.

As in the earlier experiments, we then partitioned the results into three RT ranges. The one-way ANOVAs showed a significant effect of RT partition in the category boundary data, $F(2, 48) = 4.65$, $p < .05$, and a marginal effect in the percentage of voiced responses, $F(2, 48) = 3.03$, $p < .10$. There were significant effects in both of our response measures in the intermediate and slow RT partitions, but not at the fast RT partition: For the fast RT partition, $t(24) = 0.44$, $p > .10$ by categories, and $t(24) = 0.80$, $p > .10$ by percentages; for the intermediate RT partition, $t(24) = 2.21$, $p < .05$ by categories, and $t(24) = 3.44$, $p < .005$ by percentages; and for the slow RT partition, $t(24) = 3.21$, $p < .005$ by categories, and $t(24) = 2.50$, $p < .01$ by percentages.

Figure 10 shows a summary of the effects of neighborhood frequency across the two /b/–/p/ series. The doish–toif data from Experiment 2 are also shown for comparison. The bars represent the difference in the overall percentages of voiced responses between the two series in each pair. A positive difference indicates that listeners were using the response that corresponded to the higher neighborhood fre-
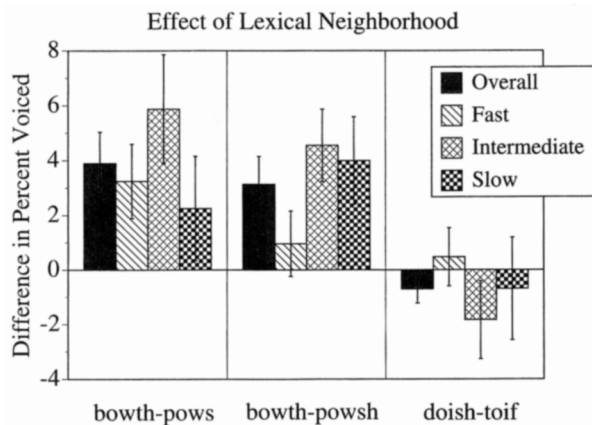
quency end of each series. The left and middle panels show the effects of neighborhood frequency for both pairs of /b/–/p/ series and suggest that number of neighbors is unlikely to be the reason for the lack of /d/–/t/ effects shown earlier. Given that the bowth–powth and bowsh–powsh stimuli have no more neighbors than the dowv–towb series (which did not show an effect of neighborhood frequency), it appears that a low number of neighbors for the /d/–/t/ series is not the cause for the lack of a neighborhood effect. In order to examine this issue in more detail, we compared the magnitude of the influence of neighborhood frequency across the two pairs of series in this experiment. If the number of neighbors does influence the size of the effect that we obtained separately from the difference in frequency-weighted neighborhood density, then perhaps the magnitude of the effect that we obtained would differ between the two pairs of series. There was no significant effect of the size of the neighborhood across the two /b/–/p/ series: $t(48) = 0.97$ by category boundaries, and $t(48) = -0.50$ by percentages, both $ps > .10$. That is, the category boundary shifts and changes in percentages of responses were no larger (or smaller) for the bowth–pows pair (with more neighbors) than they were for the bowth–powsh pair (with fewer neighbors). Thus, within the range of neighborhood sizes used in these experiments, we can find no evidence for an effect of the number of neighbors an item has.

It is not obvious why the effects were mostly in the fast and intermediate RT partitions for the bowth–pows series (with some in the slow partition) but only in the intermediate and slow partitions for the bowth–bowsh series. One possibility is that listeners were responding at different overall speeds. Upon further examination, however, this does not seem to be the case. The average endpoint RT for the bowth–pows group was 574 ms, and that for the bowth–powsh group was 573 ms.[5] There does not seem to be any consistent difference in RTs between the two groups in the different RT partitions, either. In the fast RT partitions, mean endpoint RT was 453 ms for bowth–pows listeners and 441 ms for bowth–powsh listeners. In the intermediate RT partitions, bowth–pows listeners averaged 564 ms and bowth–powsh listeners averaged 571 ms. In the slow RT partitions, bowth–pows listeners averaged 746 ms at the endpoints and bowth–powsh listeners, 760 ms. Thus, we are left with no clear reason for this difference. However, the most important aspect is not that the findings occur in different RT ranges in the two series, but that they occur in both series.

Finally, both series in the bowth–pows set contained phonotactically legal phoneme sequences. In contrast, the /aU∫/ in the bowth–powsh set does not occur in American English. In spite of this, substantially similar neighborhood effects were found for both sets. Thus, it seems reasonable to conclude that the phonotactic legality of one series in a set does not affect the influence of lexical neighborhood on phoneme perception.

Figure 10. Neighborhood frequency effect on percentage of voiced responses for bowth–pows, bowth–powsh, and doish–toif series in both the overall data and each of the three reaction time partitions.

---

[5] Only the RTs for correct responses were used here.

General Discussion

Our results demonstrate that frequency-weighted neighborhood density influences phoneme identification in a fashion very similar to the lexical effects previously reported. In VOT continua where neither endpoint was a word, there was a shift in the category boundary such that more items were classified as belonging to whichever end had a greater frequency-weighted neighborhood density. In a sense, this means that Burton et al. (1989) were wrong: Lexical influences *are* ubiquitous in speech perception. In fact, the items need not even be words to demonstrate these effects.

The effects of lexical neighborhood on phoneme perception in nonwords parallel those of lexical status on phoneme perception. Neighborhood effects were found with bilabial and velar stops varying in voicing. However, like the studies of lexical status, we found little evidence for any effect of lexical neighborhood on the perception of alveolar stops varying in voicing. Across the three experiments, we have consistently found an effect of neighborhood with /g/–/k/ and /b/–/p/ series but never with /d/–/t/ series.

In our final analysis, we compared the magnitudes of the neighborhood effect across the three places of articulation. To be fair, a significant effect might occur simply because the series differ in terms of their apparent step size. We do not claim that the spacing between members of any one series is equivalent, perceptually or acoustically, to the spacing between members of other series. In fact, the psychometric functions for the series from the first two experiments, which we discussed at the end of Experiment 2, suggest that our spacings were not equivalent. These types of differences could lead to a significant cross-series effect. However, if no significant difference in the effect of neighborhood across the three places of articulation were to be found, it would certainly suggest that the /d/–/t/ series are not as different from the other series as we have been claiming. If the /d/–/t/ series are showing a different pattern than the other places of articulation, we should see this as significant in an overall ANOVA. To this end, we calculated the difference scores for each of the 25 participants in each group in all of the experiments reported so far (the /g/–/k/ series and its replication from Footnote 3, the three /b/–/p/ series from Experiments 1 and 3, and the two /d/–/t/ series from Experiments 1 and 2). These difference scores were entered into a one-way ANOVA with three levels representing the three places of articulation for the initial stops. There was a significant overall effect of place of articulation in both the category boundary data, $F(2, 172) = 8.64, p < .001$, and the percentage voiced data, $F(2, 172) = 4.03, p < .02$. Newman-Keuls post hoc tests showed that by category boundaries, the /d/–/t/ series was significantly different from both the /b/–/p/ series and the /g/–/k/ series, whereas the latter two did not differ from one another. By percentages, the /d/–/t/ series was significantly different from the /b/–/p/ series, but there were no other significant differences.

Even though previous studies of the lexical effect had used /d/–/t/ series in which lexical neighborhood and lexical status would have produced opposite effects, it does not appear that this conflict is the cause of the difficulty in finding lexical effects for /d/–/t/ series. Because we found little evidence of a neighborhood effect for alveolar stops, some other explanation for the resistance of alveolars to higher level influences in perception is needed.

We are not sure why /d/–/t/ series are different in this respect. However, other areas of research have also demonstrated unusual effects for these phonemes (see Paradis & Prunet, 1991a). This research has shown, for example, that coronals (which include /d/ and /t/) are more frequently involved in exchange and substitution errors (Stemberger & Stoel-Gammon, 1991) as well as being transparent to feature spreading in vowel harmony languages (Paradis & Prunet, 1991b). There are a number of reasons we can suggest post hoc as to why the /d/–/t/ series might be different in our neighborhood and lexical status effects. First, it is possible that /d/ and /t/ are underspecified in their underlying representations. To the extent that the representations of coronals are somehow different from those for labial and velar stops, what constitutes a neighbor for these phonemes may also be different. Second, it is possible that the very high frequency of occurrence of /d/ and /t/ in words in English overwhelms neighborhood effects and makes these neighborhood influences very difficult to find. We cannot at present distinguish between these different ideas, but it would be interesting to determine whether the same lack of a neighborhood effect would occur with /s/–/t/ sequences. Davis (1991) argued that /s/ is not underspecified, even though it shares the alveolar place of articulation with /t/. Further, the frequency of occurrence in English of /s/ is similar to that for /d/ and /t/. Thus, using a series such as this may allow us to distinguish between the notions of phoneme frequency and underspecification.

We should point out that our method of computing frequency-weighted neighborhood density includes effects of both the number of neighbors and their frequency. Combining the results of Experiments 1 and 3, we found that the difference in neighborhood frequency in our series produces changes in phoneme perception across a range of neighborhood sizes. However, it is still possible that word frequency and density have distinct effects that occur at different points in processing. We chose to examine weighted density because of prior work suggesting its importance (Bard & Shillcock, 1993; Luce, 1987), but further work will be necessary to determine what aspects of the lexical neighborhood are most important and how they produce their influence on phoneme perception.

Our method of computing neighborhoods is based on all items that differ from the target by one phoneme. It does not place any special emphasis on neighbors that share their initial phoneme. That is, *mice* and *guide* are considered to be equally good neighbors of *gice*. This is contrary to a number of published models of word recognition, such as COHORT (Marslen-Wilson & Welsh, 1978), which suggest that word recognition begins with the activation of only those words that share the same beginning. However, this difference between theories does not alter the neighborhood differences that were used to select the stimuli for the

studies presented here. While our analysis of the nonword *gice* does include a number of neighbors that do not share the initial phoneme (such as *mice*), these items are also neighbors of the opposite endpoint of the series, *kice*. That is, these neighbors do not contribute to the difference in neighborhoods between the endpoints of a series. Whether we choose to include them as neighbors (as does the neighborhood activation model) or decide not to do so (as does COHORT), we end up with the same neighborhood difference for one endpoint over the other. So both models make equivalent predictions in this regard. The results here cannot be used to distinguish between these two ways of calculating neighborhoods.

Bard and Shillcock (1993) suggested that neighborhood effects may actually be governed by a single high-frequency neighbor. In our three /b/–/p/ series, the larger neighborhoods also contained the highest frequency words. However, for the /g/–/k/ series, the relation between neighborhood frequency and the presence of a high-frequency word broke down. The word frequency of the most frequent neighbors of both *kice* (*case*, frequency of 362) and *kipe* (*keep*, frequency of 264) are larger than those of *gice* (*gas*, frequency of 98) and *gipe* (*type*, frequency of 200). That is, the highest frequency neighbor favors the /k/ end of both series, rather than showing the crossing pattern predicted by a neighborhood account. The bias for /k/ is not equally large in both series, so it is still possible that there would be more "k" responses in one series than in the other. But the series that would presumably produce the most "k" responses would be the *gice–kice* series, the series that actually produced more "g" responses. Thus, a theory based on a single high-frequency neighbor would predict either no effect for the /g/–/k/ series or an effect opposite to the one that actually occurred. This is not to say that the presence of a single high-frequency neighbor cannot produce additional neighborhood effects. However, our results here suggest that such an account would not be sufficient in and of itself. Further work would be needed to determine whether additional effects of a single neighbor could be found.

Further work will also be necessary to examine the time course of lexical activation. Models such as TRACE and the neighborhood activation model suggest that when a word (or nonword) is heard, there is first activation of the general area in word space in which the item belongs (i.e., activation of all neighbors), followed by inhibition of incorrect items until only one item (the word), or no items, remains active. This suggests that neighborhood effects may precede lexical status effects in time. There is some evidence among our findings that suggests that this might be the case. Our *beyth–peysh* series showed neighborhood influences primarily in the intermediate reaction time partition, whereas lexical status effects are commonly found in both the intermediate and slow partitions. However, some of our other series showed effects in the slow partition as well as the intermediate partition. More detailed work examining this time course is necessary before any definite statements can be made.

Our research suggests that multiple potential words are routinely activated during the process of speech perception and word recognition. This type of effect falls out quite naturally from interactive, TRACE-like models such as the neighborhood activation model. It is worth examining how an autonomous model, on the other hand, might explain these results. Generally, recent autonomous models (such as Shortlist, see Cutler & Norris, 1979; Norris, 1994) depend on a race between two primary mechanisms of phoneme identification: data-driven phonemic analysis and competition within the lexicon (or within a small number of potential candidates). Phonological information can thus be accessed from the lexicon, as well as from the phonological analysis, but only when the lexical route results in faster response times than the phonemic route.

When this type of autonomous model is presented with a word–nonword series, an ambiguous item will activate appropriate lexical items, even though the bottom-up, phonemic information is not sufficient to clearly identify the target. Because only one end of the series is a word, listeners reading out phonological information from the lexicon will respond with the phoneme appropriate for that word. Thus, with a /bif/–/pif/ series, whenever the lexical route "wins" the race, the listener will respond "b," whereas when the phonological route wins, the listener is equally likely to respond "b" and "p." This leads to the overall "b" bias. However, the only reason the lexical route is likely to win the race is in fact because there is a word in memory that is consistent with the input. The existence of the word *beef* is what allows the lexical route to win. If the series did not include a word as one of the two endpoints, the data-driven, phonemic analysis route would consistently win the race.

How might such a process explain the neighborhood results reported here? In our series, none of the items were words. If the autonomous model depends on a race between two processing routes and one of these routes is lexical access, then our data present a problem. There is no longer any reason for the lexical route to win the race, because there is no lexical item that matches the input. The phonological route is unaffected by lexical information and would not result in a bias one way or the other. Thus, while neighborhood effects fall out quite easily from interactive models, they seem to require revisions or additional assumptions from autonomous models that explain lexical influences on phoneme perception in terms of a race between two processes. This may be the first real evidence supporting a distinction between autonomous and interactive models.

In summary, we found consistent evidence that lexical neighborhood influences phoneme perception in nonwords. For both velar and bilabial stops, listeners classified ambiguous tokens with the phonetic label that yielded a nonword syllable from a higher frequency-weighted neighborhood in the lexicon. These results parallel those previously reported for the effects of lexical status on phoneme perception. We also found little evidence that the perception of syllable-initial alveolar stops is affected by lexical neighborhood. Although the explanation for this difference between bilabial and velar stops on the one hand and alveolars on the other remains unclear, this difference parallels findings on lexical status effects on phoneme perception. Taken as a

whole, our results suggest that lexical influences on phonetic perception are ubiquitous and must be incorporated into models of auditory word recognition.

References

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*. Hillsdale, NJ: Erlbaum.

Burton, M. W., Baum, S. R., & Blumstein, S. E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 567–575.

Connine, C. M., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 291–299.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 81–94.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Davis, S. (1991). Coronals and the phonotactics of nonadjacent consonants in English. In C. Paradis & J.-F. Prunet (Eds.), *Phonetics and phonology: Vol. 2. The special status of coronals. Internal and external evidence* (pp. 49–60). San Diego, CA: Academic Press.

Engen, T. (1971). Psychophysics I: Discrimination and detection. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth and Schlosberg's experimental psychology*. New York: Holt, Rinehart & Winston.

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 526–540.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110–125.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28*, 501–518.

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

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967*. Prague, Czech Republic: Academia.

Luce, P. A. (1987). Neighborhoods of words in the mental lexicon. *Dissertation Abstracts International, 47* (12-B, Pt. 1), 5078.

Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., & Vitevitch, M. S. (1996). *Form-based inhibition priming in spoken word recognition*. Unpublished manuscript.

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*. Cambridge, MA: MIT Press.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing*. Cambridge, MA: MIT Press.

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.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 433–443.

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 369–378.

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* (Research on Speech Perception Progress Report No. 10, pp. 357–376). Bloomington: Indiana University.

Paradis, C., & Prunet, J.-F. (Eds.). (1991a). *Phonetics and phonology: Vol. 2. The special status of coronals. Internal and external evidence*. San Diego, CA: Academic Press.

Paradis, C., & Prunet, J.-F. (1991b). Introduction: Asymmetry and visibility in consonant articulations. In C. Paradis & J.-F. Prunet (Eds.), *Phonetics and phonology: Vol. 2. The special status of coronals. Internal and external evidence* (pp. 1–28). San Diego, CA: Academic Press.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 699–725.

Samuel, A. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology, 18*, 452–499.

Sawusch, J. R. (1986). Auditory and phonetic coding of speech. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Vol. 1. Speech perception* (pp. 51–88). New York: Academic Press.

Simon, H. J., & Studdert-Kennedy, M. (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. *Journal of the Acoustic Society of America, 64*, 1338–1357.

Stemberger, J. P., & Stoel-Gammon, C. (1991). The underspecification of coronals: Evidence from language acquisition and performance errors. In C. Paradis & J.-F. Prunet (Eds.), *Phonetics and phonology: Vol. 2. The special status of coronals. Internal and external evidence* (pp. 49–60). San Diego, CA: Academic Press.