Journal of Experimental Psychology:
Human Perception and Performance
2006, Vol. 32, No. 1, 120–138

0096-1523/06/$12.00 DOI: 10.1037/0096-1523.32.1.120

Representational Specificity of Within-Category Phonetic Variation in the Long-Term Mental Lexicon

Min Ju
Université du Québec à Montréal

Paul A. Luce
University at Buffalo, State University of New York

This study examines the potential encoding in long-term memory of subphonemic, within-category variation in voice onset time (VOT) and the degree to which this encoding of subtle variation is mediated by lexical competition. In 4 long-term repetition-priming experiments, magnitude of priming was examined as a function of variation in VOT in words with voiced counterparts (*cape–gape*) and without (*cow–*gow*) and words whose counterparts were high frequency (*pest–best*) or low frequency (*pile–bile*). The results showed that within-category variation was indeed encoded in memory and could have demonstrable effects on priming. However, there were also robust effects of prototypical representations on priming. Encoding of within-category variation was also affected by the presence of lexical counterparts and by the frequency of counterparts.

Keywords: specificity, within-category phonetic variation, voice onset time

The speech signal exhibits a great degree of variability as a result of both linguistic variation, such as predictable differences among the articulation of segments as a function of phonetic position and context, and extralinguistic variation, such as differences in speaker identity, speaking rate, and affective state (see Pisoni, 1997; Pisoni & Luce, 1987). Understanding how the perceptual system deals with both types of variability in the mapping of acoustic signals to form-based representations has long been a goal of research on speech perception and spoken word recognition.

The traditional view of speech perception has been that listeners normalize speech, mapping a highly variable acoustic–phonetic waveform onto abstract representations such as phonemes (e.g., Fowler & Smith, 1986; Stevens, 2002; Studdert-Kennedy, 1976, 1987). Indeed, numerous studies have attempted to discover invariant, abstract units of the speech input in the form of acoustic features (e.g., Blumstein & Stevens, 1980; Stevens & Blumstein, 1978) or articulatory gestures (e.g., Fowler, 1986; Fowler & Rosenblum, 1991; Liberman & Mattingly, 1985). Theories of spoken word recognition have also typically assumed that words are represented as abstract, phonological codes (e.g., TRACE; McClelland & Elman, 1986; SHORTLIST; Norris, 1994; PARSYN; Jackson & Morton, 1984; Luce, Goldinger, Auer, & Vitevitch, 2000).

A contrasting view holds that the representations that subserve perception are not, in fact, highly abstract but instead preserve extralinguistic variability. For example, long-term auditory priming studies have demonstrated that some (as yet poorly understood) details of the speech input appear to be preserved in memory (e.g., Church & Schacter, 1994; Craik & Kirsner, 1974; Goldinger, 1996; Goldinger, Pisoni, & Logan, 1991; Luce & Lyons, 1998; Mullennix, Pisoni, & Martin, 1989; Palmeri, Goldinger, & Pisoni, 1993; Pilotti, Bergman, Gallo, Sommers, & Roediger, 2000; Sheffert, 1998; Sommers, 1999). In the long-term priming paradigm, two blocks of stimuli (a prime or study block and a target or test block) are presented. Some of the items are repeated across blocks, whereas others in the target block are new. The typical performance increment (in terms of both accuracy and speed) observed for repeated compared with new items is called the *long-term priming effect*. Changes in the perceptual details of the stimulus (e.g., in the voice of the talker) from prime to target typically result in a decrease in the magnitude of priming. This reduction in priming as a result of changes in the perceptual details of the stimulus is referred to as *specificity*. Specificity effects demonstrate that the perceptual detail in question is preserved in long-term memory.

In addition to the research on the preservation of extralinguistic variation in form-based lexical representation, recent work has demonstrated that changes in allophonic details between study and test (e.g., flapped→ nonflapped and nonflapped→ flapped) reduce priming effects (McLennan, Luce, & Charles-Luce, 2003). This research suggests that subphonemic, as well as extralinguistic, variability is preserved in memory. Research to date thus suggests that the memory system underlying the mental lexicon preserves specific linguistic and extralinguistic details of previously encountered spoken words. However, studies demonstrating that the specificity of the speech information is encoded in memory have involved perceptually salient changes in the input. Two questions thus follow: (a) How specific, in fact, are long-term memory representations for form-based word representations? (b) In par-

---

Min Ju, Département de Psychologie, Université du Québec à Montréal, Montréal, Québec, Canada; Paul A. Luce, Department of Psychology and Center for Cognitive Science, University at Buffalo, State University of New York.

This research was supported by National Institute on Deafness and Other Communication Disorders, National Institute of Health, Research Grant R01 DC 0265801. We thank Sheila Blumstein, Jan Charles-Luce, and James Sawusch for their careful reading of the manuscript.

Correspondence concerning this article should be addressed to Min Ju, Département de Psychologie, Université du Québec à Montréal, C. P. 8888, Succursale Centre-Ville, Montréal, Québec H3C 3P8, Canada. E-mail: c3440@er.uqam.ca

ticular, is fine-grained acoustic–phonetic information encoded in long-term form-based representations, or is this information mapped onto more abstract representations, be they featural or gestural?

An abundance of evidence suggests that listeners are sensitive in the short term to fine-grained speech information. In consonant perception, studies suggest that listeners are sensitive to fine-grained within-category differences in voice onset time (VOT). VOT is the brief interval between the stop release and the onset of voicing and is one of the primary features used in making phonetic category distinctions between voiced and voiceless stop consonants (/b/, /d/, /g/ vs. /p/, /t/, /k/) in English (e.g., Lisker & Abramson, 1970; Miller & Volaitis, 1989). Although VOT is a gradient property, perception of stop consonants that vary in VOT is often categorical (e.g., Jusczyk, 1986; Liberman, Harris, Hoffman, & Griffith, 1957).

Despite the categorical nature of consonant perception, within-category variation has a demonstrable impact on listeners' performance. Reaction times in both phoneme identification and discrimination tasks increase as stimuli move away from the prototypical range of VOT (Pisoni & Tash, 1974). Listeners' goodness judgments of stop consonants are also influenced by within-category variation: Ratings of category goodness decrease as stimuli approach the category boundary (Allen & Miller, 2001; Miller, 1997). Listeners' sensitivity to fine-grained acoustic information for consonants is also demonstrated in online spoken word recognition in monolingual speakers (e.g., McMurray, Tanenhaus, & Aslin, 2002) as well as in bilingual speakers (e.g., Ju & Luce, 2004). Studies that have investigated the effects of subcategorically mismatching information on short-term lexical activation have also demonstrated that lexical access is impaired by mismatching coarticulatory information (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Streeter & Nigro, 1979; Whalen, 1991), which further supports the view that the speech recognition system is, at least in immediate perceptual processing, sensitive to detail below the level of categorical units, such as phonemes.

Researchers have also demonstrated sensitivity to within-category variation using the semantic priming paradigm. Andruski, Blumstein, and Burton (1994) manipulated stimuli in a steplike manner by reducing the duration of VOT in spoken words with word-initial voiceless stops (e.g., *king*, *cat*) and examined the magnitude of semantic priming of targets as a function of three different durations of VOT of the primes (intact, one third of VOT removed, and two thirds of VOT removed). Andruski et al. found that primes containing intact VOTs activated their semantic lexical representation more strongly than those with modified VOTs (although the authors found a significant difference only between intact and two thirds primes). Researchers have since observed similar reductions in semantic facilitation for a number of temporal and spectral acoustic cues, including vowel duration and voicing during closure of word-final voiced stops (Utman, 1997; Utman, Blumstein, & Burton, 2000).

Clearly, listeners are sensitive to within-category (linguistic) variation in short-term priming. A critical question thus arises concerning the effects of within-category variation on long-term form-based representations. Two possibilities arise. First, subphonemic or allophonic details may be mapped onto abstract, categorical information stored in the long-term mental lexicon. In fact, most theories of spoken word recognition (e.g., TRACE; McClelland & Elman, 1986; SHORTLIST; Norris, 1994; PARSYN; Luce et al., 2000) assume abstract processing units. Thus, under abstractionist accounts, sensitivity to within-category variation (i.e., specificity effects) reflects the relative goodness of fit of the token to the prototypical representation of the perceptual category (e.g., phoneme).

Alternatively, phonological information may be stored as independent memory traces with specific veridical representations of each token preserved, as suggested by episodic theories of word perception (Goldinger, 1996, 1998). In this view, the within-category variation is encoded in word representations themselves in their specific form. Thus, specificity effects arise from the perceptual mismatches between studied and tested tokens. Theories that assume distributed representations (e.g., TRACE; McClelland & Elman, 1986) would also predict specificity effects, as a recently encountered token would have the most pronounced impact on connections among nodes representing the sound patterns of words. Hence, the mismatches in the perceptual information of words would attenuate priming.

In short, is fine-grained within-category variation encoded in long-term lexical representations? Although the answer to this question may reveal critical insights into the nature of form-based representations in the mental lexicon, no study has to date directly addressed the issue of whether within-category variation is preserved in long-term memory. The present study addresses this issue by examining whether within-category phonetic variation in VOT is encoded in long-term memory and has consequences for subsequent word recognition. That is, the present study examines whether mismatches in VOT of word-initial consonants have demonstrable effects on the subsequent recognition of words that contain word-initial voiceless stops. If specific details of within-category information are preserved, mismatches in VOT between studied and tested tokens should reduce long-term priming. If the speech recognition mechanism encodes phonological information strictly in long-term abstract, categorical representations, changes in within-category information should not have demonstrable effects on subsequent word recognition.

The present study also addresses the *gradience* of word representations in long-term memory. Andruski et al. (1994) observed effects of manipulating VOT only in the tokens that were substantially different (two thirds reduction in VOT) from the prototypical tokens, not in the tokens that underwent a lesser degree of manipulation (one third reduction). Their finding showed that small differences in VOT do not appear to influence lexical activation in short-term priming, suggesting that the representation of form-based word information may not be strictly gradient.[1] By contrast, McMurray et al. (2002) observed a gradient degree of short-term lexical activation as a function of VOT. Their finding suggests that within-category variation has continuous effects on short-term form-based lexical activation, unlike the finding from a semantic priming study by Andruski et al. (1994). A question then follows: To what extent do long-term form-based representations preserve the potentially gradient nature of within-category variability?

---

[1] There is also the possibility that the paradigm was not sensitive enough to pick up the small differences.

We also examine the hypothesis that listeners may encode within-category variation differently depending on whether they hear items with different VOTs (mixed presentation) or with the same VOT (blocked presentation). For example, if listeners hear only items with shortened VOTs in blocked presentation, they may shift their category boundary temporarily, treating these items as intact tokens. Under these circumstances, we may observe different degrees of sensitivity to within-category variation.

Finally, the present research also examines whether encoding of within-category variation is modulated by lexical competition. In particular, we examine whether the encoding of within-category variation may be affected by the presence of lexical counterparts. Within-category variation may have different perceptual consequences depending on whether voiceless words share the VOT space with a word or a nonword. If a word with a word-initial voiceless stop has a voiced counterpart (e.g., *clue–glue*), the variation requires listeners to discriminate between two different words. In contrast, in words without counterparts (e.g., *cow–*gow*), the variation requires listeners to make discrimination between words and nonwords. Because the consequences of within-category variation on making perceptual decisions are different as a function of the presence of a lexical counterpart, listeners may demonstrate differing degrees of sensitivity to within-category variation in the presence of lexical competition (see, e.g., Andruski et al., 1994; Ganong, 1980).

Moreover, greater specificity effects may arise for words whose counterparts are high- rather than low-frequency words. For example, consider the two minimal pairs *pest–best* versus *pile–bile*. Both pairs consist of real words, yet the words in the former pair are equal in frequency (Kučera & Francis, 1967), whereas in the latter pair the voiceless alternative is higher in frequency. If both alternatives are high-frequency words, lexical competition between voiceless and voiced alternatives should be greater (e.g., Luce & Pisoni, 1998), increasing the importance of VOT in lexical discrimination for the former pair over the latter pair. It is thus possible that listeners may process a greater amount of within-category information when the voiced alternative constitutes a high- rather than a low-frequency word.

In summary, the present research examines four questions: (a) Are fine-grained phonetic details preserved in long-term form-based representations of spoken words? (b) If within-category variation is encoded in long-term memory, to what extent is it gradient? (c) Does encoding of variation depend on the degree of variability to which the listener is exposed? (c) Is encoding of within-category variation modulated by lexical competition?

To address these four issues, we conducted four long-term repetition-priming experiments using the lexical decision task. The stimuli consisted of words containing word-initial voiceless stops. We manipulated the spoken words by reducing the duration of VOTs from the intact tokens to create three sets of stimuli that varied only in VOT. The first set consisted of intact tokens. We made the second set by reducing VOT of the original tokens by one third (the −1/3 set). Finally, we made the third set by reducing VOT by two thirds (the −2/3 set; see Figure 1). To ensure that participants perceived VOT-reduced words as intended (voiceless), we screened target words using phoneme categorization tasks (Experiments 1A, 2A, and 3B).[2]

Figure 1. Schematic presentation of stimuli presentation. The vertical line indicates that midpoint of voicing, the inner bracket indicates the portion reduced for one third stimuli, and the outer bracket represents the portion reduced for two thirds stimuli.

In Experiment 1B, participants made lexical decisions in both the prime and the target block. The word primes consisted of three different VOTs. The word targets consisted of the matching VOT items, mismatching VOT items, and unprimed items (e.g., matching: intact → intact, −1/3 → −1/3, −2/3 → −2/3; mismatching: intact → −1/3, intact → −2/3, −1/3 → intact, −1/3 → −2/3, −2/3 → intact, −2/3 → −1/3). We compared reaction times between matching and mismatching VOT conditions. Experiment 1C was identical to Experiment 1B except that the primes in this experiment were presented in blocked mode (only intact, one third, or two thirds).

In Experiment 2, we used intact and two thirds items, and participants again made lexical decisions in both the prime and the target block. The primes were presented in mixed mode. Half of the targets were words with voiced counterparts, and the other half were words without voiced counterparts. Experiment 3 was identical to Experiment 2 except that all targets in this experiment consisted of words with voiced counterparts: Half of the targets had high-frequency voiced counterparts, and the other half had low-frequency voiced counterparts.

Experiment 1: Within-Category Specificity and Gradient Lexicon

In Experiment 1 we examine whether fine-grained phonetic details are preserved in the long-term mental lexicon and whether these effects vary continuously as a function of the degrees of VOT reduction.

Experiment 1A: Phoneme Categorization

To ensure that all word-initial voiceless stops in targets used in Experiments 1B and 1C would be perceived as intended, we screened stimuli using a phoneme categorization task in which participants made forced-choice decisions on voiced–voiceless distinctions (e.g., /b/–/p/, /d/–/t/, /g/–/k/). Consistent with previous research that has shown gradient sensitivity to within-category differences, we expected that the greater the reduction in VOTs was, the slower and less accurate listeners would be at making phonemic decisions on the basis of the stimuli. Such a pattern of results would demonstrate that listeners are sensitive to fine-grained acoustic–phonetic differences when they have to make explicit phonemic decisions. If listeners are sensitive to within-category differences between different sets and yet do not show significant specificity effects in either set in long-term priming,

---

[2] Because VOT was reduced from the midpoint of the VOT range, there was a slight possibility that some part of the formant transitional information might also have been affected for some words, resulting in perception of a sound other than a voiced or voiceless stop consonant. We screened the stimuli in part to address this possibility.

this will suggest that long-term auditory priming is mediated by abstract representations.

Method

Participants. We recruited 24 participants from the University at Buffalo, State University of New York, community. They received partial credit for a course requirement. Participants were right-handed native speakers of American English, with no reported history of speech or hearing disorders.

Materials. The stimuli consisted of 63 spoken target words containing word-initial voiceless stops (e.g., *kite*, *top*, *pot*) and 63 spoken filler words containing word-initial voiced stops (e.g., *brush*, *dog*, *game*). We manipulated the VOTs of the target words as follows: We created −1/3 VOT items by removing one third of the VOT from the midpoint of the VOT range of intact items (see Figure 1). We created −2/3 VOT items by reducing the VOTs of intact items by two thirds in the same manner.

The stimuli were recorded in a sound-attenuated room by a male speaker of a Western New York dialect, low-pass filtered at 20 kHz, and digitized at a sampling rate of 44.1 kHz via a 16-bit analog-to-digital converter. All words were edited into individual files and stored on computer disk.

Design. We presented all three types of voiceless targets (intact, −1/3, −2/3) in three separate blocks (e.g., /g/–/k/, /b/–/p/, /d/–/t/). We also repeated voiced words three times to balance the total number of voiced and voiceless words.

Procedure. Participants were tested individually in a quiet room. They performed a forced-choice phoneme categorization task in which they were instructed to decide as quickly and accurately as possible whether the item they heard began with a voiced or voiceless sound (e.g., /b/ or /p/). They indicated their decision by pressing one of two appropriately labeled buttons (*/b/*, */d/*, or */g/* on the left and */p/*, */t/*, or */k/* on the right) on a response box positioned directly in front of them. In both the prime and the target blocks, the stimuli were presented binaurally over headphones. PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993), running on a Macintosh computer, controlled stimulus presentation and recorded participants' times in making phonemic categorizations. The stimuli were presented in three blocks (/b/–/p/ block, /d/–/t/ block, and /g/–/k/ block). All three types of VOT were presented in a given block. Presentation within each block was random for each participant. The order of the /k/, /p/, and /t/ blocks was counterbalanced via six lists.

A given trial proceeded as follows: A beep indicated the beginning of the trial. The participant was then presented with a stimulus word binaurally over the headphones. The participant was instructed to make a categorization decision as quickly and accurately as possible. Reaction times were measured from the onset of the presentation of the stimulus word to the onset of the participant's button press response. Stimulus duration was subtracted from the reaction times. After the participant responded, the next trial was initiated.

Results and Discussion

The stimuli were first scored for correct categorization. There were 51 items that met the predetermined criteria of 90% correct categorization, of which 24 were randomly selected for inclusion in the subsequent experiments.[3] Mean and range of VOTs of the 24 words are in Table 1.

The mean correct categorization rate for the initial stop in the 24 selected words was 98% (99% for intact and −1/3 stimuli, 98% for −2/3 stimuli). We performed subject (F_1) and item (F_2) one-way analyses of variance (ANOVAs; intact, −1/3, −2/3) on reaction times for correct responses and percentages correct of the 24 selected items. We replaced reaction times greater than 2,000 ms with the appropriate condition mean. We also replaced reaction times beyond two standard deviations from the condition mean with the appropriate condition mean. We replaced 3% of the total data.

Table 1

Mean and Range of VOTs (ms) for Intact /p,t,k/ Words Used in Lexical Decision Tasks

Experiment and Stimuli	*n*	Mean VOT	Range
Experiments 1B and 1C			
/p/	8	78	54–114
/t/	8	73	58–82
/k/	8	82	64–99
Experiment 2B			
/p/			
No CP	8	87	56–107
CP	8	86	64–105
/t/			
No CP	6	89	73–104
CP	6	94	67–113
/k/			
No CP	4	82	67–92
CP	4	94	76–107
Experiment 3C			
/p/			
Low	6	97	88–107
High	7	90	77–103
/t/			
Low	3	91	55–110
High	2	75	73–77
/k/			
Low	3	87	67–98
High	3	95	76–112

Note. *High* and *Low* refer to the frequency of the voiced counterpart. VOT = voice onset time; CP = lexical counterpart.

The mean reaction times to the three types of stimuli were 277 ms ($SD = 133$) for intact, 297 ms ($SD = 130$) for −1/3, and 375 ms ($SD = 150$) for −2/3. Listeners were sensitive to the VOT manipulation in the stimuli; that is, the more prototypical the VOT was, the faster was the categorization decision. In particular, reaction times to the three types of stimuli differed as a function of VOT conditions, $F_1(2, 46) = 78.34$, $p < .01$; $F_2(2, 46) = 103.12$, $p < .01$. Planned comparisons showed that listeners responded faster to the intact stimuli than to the −1/3 stimuli, $F_1(1, 46) = 5.95$, $p < .02$; $F_2(1, 46) = 6.59$, $p < .02$, and faster to the intact stimuli than to the −2/3 stimuli, $F_1(1, 46) = 140.46$, $p < .01$; $F_2(1, 46) = 195.09$, $p < .01$. Listeners also responded faster to the −1/3 stimuli than to the −2/3 stimuli, $F_1(1, 46) = 88.61$, $p < .01$; $F_2(1, 46) = 97.91$, $p < .01$. There were no significant differences in accuracy.

Experiment 1B: Mixed VOT Priming

In Experiment 1B we used a long-term repetition priming paradigm to examine whether within-category variation is preserved

---

[3] Because the design has 12 conditions, allowing 2 words for each condition resulted in a total of 24 words. Given that another 24 filler words were necessary for fully counterbalancing primed and unprimed items and matching and mismatching VOT items, 24 out of the 51 words that met the 90% accuracy criteria was the maximum number of words usable for the target.

in long-term memory. This paradigm enables us to determine whether two nominally different stimuli activate the same mental representation. In the present experiment, two blocks of stimuli (prime and target) containing words and nonwords were presented. In the prime block, all three types of stimuli were presented in equal number. In the target block, some of the items were presented again, whereas others were unprimed control items. Half of the primed items contained matching VOTs, and the other half contained mismatching VOTs.

Attenuation in priming for the prime–target pair that is mismatching on VOTs relative to the pair containing the same VOTs will be taken as evidence of *within-category specificity*. Within-category specificity effects indicate that within-category variation is encoded in long-term memory. If within-category specificity effects exist and their magnitude varies as a function of VOT reduction (i.e., a greater reduction between intact and −1/3 items than between intact and −2/3 items), this will suggest that form-based information in long-term memory is represented in a gradient, noncategorical manner. If smaller VOT variation does not result in specificity effects, whereas larger VOT variation does, this will suggest that long-term memory priming is mediated by prototypical representations.

Method

Participants. We recruited 60 participants from the University at Buffalo, State University of New York, community. They received partial credit for a course requirement. Participants were right-handed native speakers of American English, with no reported history of speech or hearing disorders.

Materials. The stimuli consisted of 24 spoken mono-, bi-, and trisyllabic words containing word-initial voiceless stops (e.g., *tiff*, *poster*, *cabinet*). There were three versions of each stimulus: intact, −1/3, and −2/3 VOT. We included 24 nonwords as fillers. (See Appendix A for the complete list.) We constructed the nonwords by changing the final syllables of the target words (e.g., *camel* vs. **camoid*) and recorded them separately.

Design. Two blocks of stimuli were presented, the first constituting the primes and the second the targets. Orthogonal combination of the four levels of prime type (intact, −1/3, −2/3, control) and three levels of target type (intact, −1/3, −2/3) resulted in 12 conditions, shown in Table 2. Across participants, each item participated in every possible condition. However, no single participant heard more than one version of a given word within a block. For example, if a participant heard the word *kite* in one of the blocks, he or she did not hear any version of that word again in the same block.

The prime block consisted of 24 words and 24 nonword fillers. Half of the word stimuli were in the matching VOT condition, and half were in the mismatching condition. The composition of the 12 matching stimuli was as follows: 6 critical primes (2 items per VOT type) and 6 filler primes.[4] The 12 mismatching stimuli were also composed of 2 items per condition, and there were six mismatching conditions orthogonally combined across three types of stimuli (see Table 2).

The target block consisted of 48 words and 48 nonword fillers. Half of the target words were the items presented in the prime block, and the other half were new. Among the 24 primed items, half of them appeared in matching VOT and the other half in mismatching VOT. The composition of the 12 VOT-matching targets was as follows: 6 critical targets (2 items per VOT type) and 6 filler targets. The 12 VOT-mismatching targets were also composed of 2 items per condition, and there were six conditions orthogonally combined across three types of stimuli (see Table 2). Among the 24 unprimed items, 6 were critical control items (2 items per VOT type), and 18 were filler controls (6 items per VOT type).

Procedure. Participants were tested individually in a quiet room and were not told at the beginning of the experiment that there would be two blocks of trials. Participants performed a lexical decision task in which they were instructed to decide as quickly and accurately as possible whether the item they heard was a real English word or a nonword. They indicated their decision by pressing one of two appropriately labeled buttons (*word* on the right and *nonword* on the left) on a response box positioned directly in front of them. In both the prime and the target blocks, the stimuli were presented binaurally over headphones. PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993), running on a Macintosh computer, controlled stimulus presentation and recorded participants' lexical decision times. Stimulus presentation within each block was random for each participant.

A given trial proceeded as follows. A beep was played to indicate the beginning of the trial. The participant was then presented with a stimulus word or nonword binaurally over the headphones. The participant was instructed to make a lexical decision as quickly and accurately as possible. Reaction times were measured from the onset of the presentation of the stimulus word to the onset of the participant's button press response. Stimulus duration was subtracted from the reaction times. After the participant responded, the next trial was initiated. If the maximum reaction time (5 s) expired, the computer automatically recorded an incorrect response and presented the next trial.

After the first block, the participant was given a distractor task in which he or she was instructed to solve mathematical problems on a sheet provided. This task lasted for 5 min. The target block was then presented. The procedure for the target block was identical to that for the prime block.

Results

We performed a repeated measure ANOVA on reaction times for correct responses and percentages correct for the target word stimuli, with prime type (intact, −1/3, −2/3, control) and target type (intact, −1/3, −2/3) as within-subject factors. We replaced reaction times greater than 2,000 ms with the appropriate condition mean. We also replaced reaction times beyond two standard deviations from the condition mean with the appropriate condition mean. We replaced 3% of the total data.

Accuracy. Overall accuracy was 93% (intact, 98%; −1/3, 99%; −2/3, 81%) but differed as a function of target type, $F_1(2, 118) = 67.77$, $p < .01$; $F_2(2, 46) = 21.27$, $p < .01$. Accuracy for the −2/3 stimuli was lower than that for the intact and −1/3 stimuli, $F_1(1, 118) = 96.59$, $p < .01$; $F_2(1, 46) = 30.35$, $p < .01$; and $F_1(1, 118) = 106.49$, $p < .01$; $F_2(1, 46) = 33.38$, $p < .01$, respectively. No other effects were significant.

Reaction times. Reaction times as a function of prime and target type are plotted in Figure 2. Reaction times significantly differed both as a function of prime type, $F_1(3, 177) = 14.66$, $p < .01$; $F_2(3, 69) = 10.67$, $p < .01$, and as a function of target type, $F_1(2, 118) = 39.44$, $p < .01$; $F_2(2, 46) = 15.91$, $p < .01$. The interaction of prime type and target type was also significant, $F_1(6, 354) = 2.62$, $p < .02$; $F_2(6, 138) = 2.24$, $p < .05$. We conducted planned comparisons to examine the results in terms of priming and specificity.

Priming. Priming is indicated by the significant difference in reaction time between the primed condition (matching, mismatching) and its appropriate control condition. Priming patterns dif-

---

[4] Orthogonally combining mismatching conditions across three types of stimuli resulted in twice as many mismatching conditions as matching conditions. We added fillers to balance the number of items between matching and mismatching conditions.

Table 2
Experimental Conditions: Mixed (Experiment 1B) and Blocked (Experiment 1C) Mode of Prime Block

	Experiment 1B (mixed primes)	Experiment 1C (blocked primes)		
Condition	Prime → target	Prime → target (intact primes)	Prime → target (−1/3 primes)	Prime → target (−2/3 primes)
Matching	Intact → intact −1/3 → −1/3 −2/3 → −2/3	Intact → intact	−1/3 → −1/3	−2/3 → −2/3
Mismatching	Intact → −1/3 Intact → −2/3 −1/3 → intact −1/3 → −2/3 −2/3 → intact −2/3 → −1/3	Intact → −1/3 Intact → −2/3	−1/3 → intact −1/3 → −2/3	−2/3 → intact −2/3 → −1/3
Control (unprimed)	Intact −1/3 −2/3	Intact −1/3 −2/3	Intact −1/3 −2/3	Intact −1/3 −2/3

Note. −1/3 = one third of voice onset time removed; −2/3 = two thirds of voice onset time removed.

fered both by target types and by prime types. For intact targets, both intact primes and −1/3 primes significantly primed matching intact targets, $F_1(1, 354) = 10.51, p < .01$; $F_2(1, 138) = 9.55, p < .01$; and $F_1(1, 354) = 6.63, p < .02$; $F_2(1, 138) = 5.88, p < .02$, respectively. In contrast, −2/3 primes did not significantly prime intact targets, $F_1(1, 354) = 1.53$ and $F_2 < 1$.

For −1/3 targets, −1/3 primes did not significantly prime matching (F_1 and $F_2 < 1$), whereas both intact and −2/3 primes significantly primed mismatching. These effects were significant by subjects and approached significance by items, $F_1(1, 354) = 4.64, p < .04$; $F_2(1, 138) = 3.61, p < .06$; and $F_1(1, 354) = 4.21, p < .05$; $F_2(1, 138) = 2.89, p < .10$, respectively.

For −2/3 targets, all three prime types produced significant priming: −2/3 primes, $F_1(1, 354) = 30.19, p < .01$; $F_2(1, 138) = 18.80, p < .01$; intact primes, $F_1(1, 354) = 34.21, p < .01$; $F_2(1, 138) = 36.48, p < .01$; and −1/3 primes, $F_1(1, 354) = 9.65, p < .01$; $F_2(1, 138) = 13.81, p < .01$.

Specificity. Within-category specificity is indicated by a significant difference between the matching and mismatching VOT conditions for each target type. Within-category specificity effects were obtained for intact targets and −2/3 targets. That is, listeners responded significantly faster to intact targets when they were followed by matching intact primes than when they were followed by mismatching −2/3 primes, $F_1(1, 354) = 4.02, p < .05$; $F_2(1, 138) = 4.43, p < .04$. A smaller VOT difference did not have demonstrable effects on priming of intact targets, as indicated by the finding that listeners were no faster with matching intact primes than with mismatching −1/3 primes (both Fs < 1). The two mismatching VOT primes (−1/3 primes and −2/3 primes) also showed no significant difference, $F_1(1, 354) = 1.79$ and $F_2(1, 138) = 1.32$.

Listeners also responded significantly faster to −2/3 targets when they were followed by matching −2/3 primes than when they were followed by mismatching −1/3 primes. However, this effect was significant by subjects only, $F_1(1, 354) = 5.70, p < .02$ ($F_2 < 1$). We obtained no specificity effect between matching −2/3 primes and mismatching intact primes, $F_1 < 1$ and $F_2(1, 138) = 1.90$, which indicates that specificity effects existed only in a certain combination of primes and targets. Two mismatching VOT conditions (intact primes and −1/3 primes) differed signif-

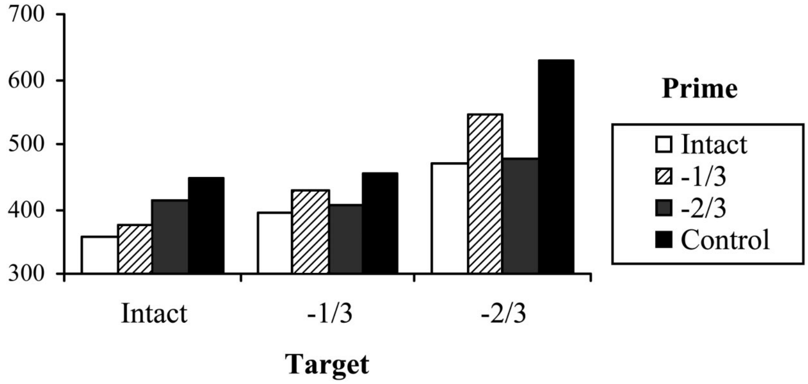

Figure 2. Lexical decision task (Experiment 1B): mixed prime block. Values in the Y axis represent response time (msec). −1/3 = one third of voice onset time was removed; −2/3 = two thirds of voice onset time was removed.

icantly, $F_1(1, 354) = 7.52$, $p < .01$; $F_2(1, 138) = 5.40$, $p < .03$, indicating that even a small VOT variation can have a demonstrable effect on long-term auditory priming. For −1/3 targets, there was no significant difference between matching and either of the mismatching VOT conditions (all *F*s <1).

Discussion

These results demonstrate that variation in VOT between prime and target can cause a reduction in long-term priming under certain conditions, resulting in within-category specificity effects. However, we obtained specificity effects only between intact primes and −2/3 primes for intact targets and between −2/3 primes and −1/3 primes for −2/3 targets. Examination of priming effects reveals a prominent role for prototypical VOT ranges of primes. Intact primes facilitated responses to targets with all types of VOTs, demonstrating that items with prototypical VOTs can prime targets even when VOTs do not match. In contrast, −2/3 primes facilitated responses to targets only when VOTs matched (−2/3 targets) or there was a one third reduction (−1/3 targets). When there was a greater than one third difference in VOT, these atypical −2/3 items failed to produce priming. However, a one third reduction in VOT did not appear to significantly attenuate priming, as evidenced by the fact that −1/3 primes facilitated processing of intact and −2/3 targets, both in mismatching conditions. Taken together, these results suggest that the degree of lexical activation varies as a function of the prototypicality of primes. Prototypical representations produce significant priming regardless of matches in VOT, whereas the less prototypical representations produce significant priming only when there is a small difference in VOT or when VOT ranges match precisely.

However, there was also an indication that long-term priming was mediated at least in part by specific memory representations. Note that mismatching −2/3 primes did not significantly prime intact targets and that the magnitude of priming produced by mismatching −1/3 targets for −2/3 targets was significantly lower than that produced by matching −2/3 primes. Taken together, these results suggest that long-term priming is mediated both by prototypical and by specific memory representations, although the prototypes appear to play a more prominent role in priming.

Although a one third reduction in VOT did not have demonstrable effects on the magnitude of long-term priming in other conditions, reaction times for the two mismatching VOT conditions (intact and −1/3 primes) for −2/3 targets differed significantly. This indicates that even a small variation in VOT can have a demonstrable effect on priming under some circumstances. Given that listeners were slowest on −2/3 targets, it is plausible that a small variation in VOT can, indeed, incur demonstrable processing costs when the baseline performance is slow. If this is true, it suggests that lexical activation in long-term priming may also be gradient, not unlike listeners' gradient sensitivity to fine-grained phonetic information (e.g., Pisoni & Tash, 1974). Clearly, listeners are sensitive to subtle variation in VOT, as demonstrated by the results from the phoneme categorization task, which showed that reaction times to all three types of stimuli differed. However, in a long-term priming task in which listeners were not required to make decisions regarding the categories of sounds explicitly, their sensitivity to fine-grained phonetic information seemed to be attenuated, as evidenced by the finding that even when VOTs did not match, words could prime each other (e.g., −1/3 → intact, intact → −2/3). Again, this points to the possibility that long-term priming may be primarily mediated by abstract representations. However, it is also evident from the results that within-category variation is preserved in long-term memory and influences priming. In the next experiment, we address the same issue while varying the mode of prime presentation to examine the effects of exposure to varying VOTs versus consistent VOTs on priming.

Experiment 1C: Blocked VOT Priming

Experiment 1C was identical to Experiment 1B except that in the present experiment, participants heard stimuli with the same VOTs in the prime block. We hypothesized that listeners may encode within-category variation differently depending on whether they hear items with different VOTs (mixed presentation) or whether they hear items with the same VOT (blocked presentation). To test this possibility, we presented all primes in this experiment in blocked mode (all intact, all −1/3, or all −2/3).

Method

Participants. We recruited 180 participants (60 for each type of prime) from the University at Buffalo, State University of New York, community. They received partial credit for a course requirement. Participants were right-handed native speakers of American English, with no reported history of speech or hearing disorders.

Materials. The same materials as in Experiment 1B were used.

Design. The design of the experiment was identical to that of Experiment 1B in every respect except that the primes within a given block consisted of a single type of VOT (all intact, all −1/3, or all −2/3). Thus, VOT was a between-subjects factor in this experiment. Each experiment within a given prime block consisted of six conditions: three primed conditions and three control conditions (see Table 2). As in Experiment 1B, for a given participant, two words occurred in each condition. We equated the number of primed and unprimed conditions and matching and mismatching conditions by adding an appropriate number of fillers.

Procedure. The same procedure was used as in Experiment 1B.

Results and Discussion

We performed a repeated measure ANOVA on reaction times for correct responses and percentages correct for the target word stimuli, with prime block type (intact, −1/3, −2/3) as a between-subjects factor and target type (intact, −1/3, −2/3) and priming condition (primed, unprimed) as within-subject factors. We replaced reaction times greater than 2,000 ms with the appropriate condition mean. We also replaced reaction times beyond two standard deviations from the condition mean with the appropriate condition mean. We replaced less than 5% of the total data.

Accuracy. Overall accuracy was 92% (intact, 97%; −1/3, 95%; −2/3, 82%). Accuracy differed as a function of target type, $F_1(2, 354) = 60.35$, $p < .01$; $F_2(2, 138) = 20.67$, $p < .01$. Planned comparisons on the main effect showed that, as in Experiment 1B, listeners were significantly less accurate on −2/3 stimuli than on intact stimuli, $F_1(1, 354) = 97.95$, $p < .01$; $F_2(1, 138) = 55.88$, $p < .01$, and on −1/3 stimuli, $F_1(1, 354) = 82.44$, $p < .01$; $F_2(1, 138) = 81.17$, $p < .01$. No other effects were significant.

Reaction times. Reaction times as a function of prime block type, target type, and priming type are plotted in Figure 3. To show priming effects as a function of prime block type (between-

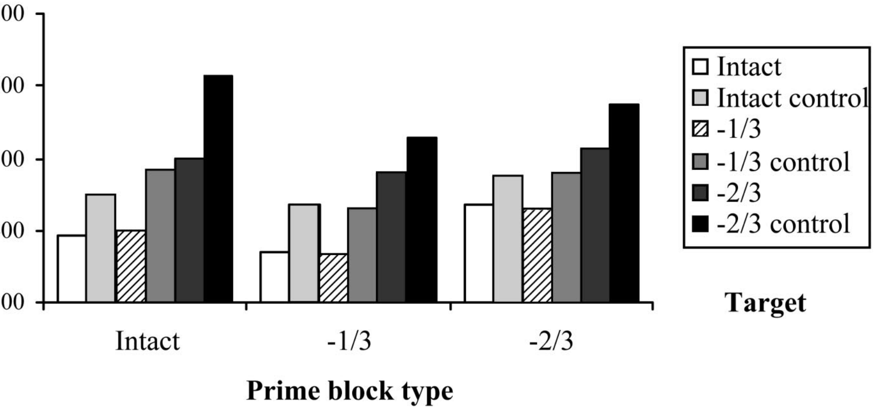

Figure 3. Lexical decision task (Experiment 1C): blocked prime block. Values in the Y axis represent response time (msec). −1/3 = one third of voice onset time was removed; −2/3 = two thirds of voice onset time was removed.

subjects manipulation), we have plotted the data with prime block type on the *x*-axis. Listeners responded faster to targets in primed than in unprimed conditions, $F_1(1, 177) = 42.32$, $p < .01$; $F_2(1, 71) = 31.21$, $p < .01$. Reaction times also differed as a function of target type, $F_1(2, 354) = 77.40$, $p < .01$; $F_2(2, 142) = 46.22$, $p < .01$. Planned contrasts revealed that listeners were significantly slower on −2/3 targets than on both intact and −1/3 targets, $F_1(1, 354) = 120.97$, $p < .01$; $F_2(1, 142) = 66.98$, $p < .01$; and $F_1(1, 354) = 121.41$, $p < .01$; $F_2(1, 142) = 71.59$, $p < .01$, respectively. However, there was no significant main effect of prime block types (intact, −1/3, −2/3), which indicates that the magnitude of priming did not differ as a function of prime block types. We found no other significant effects.

Priming. We conducted planned contrasts to examine priming effects in greater detail. Overall, in the blocked prime mode, all prime types resulted in significant priming relative to their own controls. In the intact prime block, all targets showed significant priming relative to their respective controls: intact targets, $F_1(1, 118) = 4.10$, $p < .05$; $F_2(1, 46) = 4.35$, $p < .05$; −1/3 targets, $F_1(1, 118) = 7.87$, $p < .01$; $F_2(1, 46) = 5.52$, $p < .03$; and −2/3 targets, $F_1(1, 177) = 16.04$, $p < .01$; $F_2(1, 46) = 15.17$, $p < .01$. In the −1/3 prime block, intact targets showed significant priming relative to their own controls, $F_1(1, 118) = 7.34$, $p < .01$; $F_2(1, 46) = 4.05$, $p < .05$. In addition, −1/3 targets and −2/3 targets showed significance by subjects, $F_1(1, 118) = 6.50$, $p < .02$, and $F_1(1, 118) = 4.35$, $p < .04$, respectively, and approached significance by items, $F_2(1, 46) = 3.03$, $p < .09$, and $F_2(1, 46) = 3.45$, $p < .07$, respectively.[5] In the −2/3 prime block, intact targets showed weak priming, and the effect approached significance by subjects only, $F_1(1, 118) = 3.27$, $p < .08$; $F_2(1, 46) = 1.98$. We found that −1/3 targets showed significant priming by subjects, $F_1(1, 118) = 4.49$, $p < .04$, and this was marginally significant by items, $F_2(1, 46) = 3.66$, $p < .07$. In addition, −2/3 targets showed significant priming both by subjects and by items, $F_1(1, 118) = 7.96$, $p < .01$; $F_2(1, 46) = 7.12$, $p < .02$.

Specificity. To compare the magnitude of priming across different conditions with different baselines, we normalized reaction times by the duration of their own control conditions: (target reaction time − baseline reaction time)/baseline reaction time.[6] We performed a repeated measure ANOVA on the basis of the normalized reaction times, with prime block type (intact, −1/3, −2/3) as a between-subjects factor and target type (intact, −1/3, −2/3) as a within-subject factor. We obtained no significant effects, which suggests that the magnitude of priming was similar in all conditions and within-category specificity was not as evident as in the mixed prime experiment. The only suggestion of within-category specificity effects was provided by the fact that in the −2/3 prime block, −2/3 targets (matching VOT) showed significant priming, whereas intact targets (mismatching VOT) did not, which is consistent with results in the same condition from the mixed prime experiment.

Specificity effects in the blocked experiments were not as pronounced as in the mixed experiment. The only evidence for specificity in the blocked experiment comes from the finding that −2/3 targets showed significant priming in the matching VOT condition, whereas they showed only marginally significant priming in the mismatching condition with intact primes. In all other conditions, however, the magnitude of priming did not significantly differ between matching and mismatching conditions. This contrasts with the findings from the mixed experiment, in which both intact and −2/3 targets showed significant specificity effects. This presents a puzzling question concerning why listeners may encode within-category variation differently depending on whether VOTs of primes are consistent or mixed.

Two possibilities arise. First, presenting primes with mixed VOT may increase the probability of within-category variation being encoded in long-term memory, perhaps because experience with variability encourages listeners to attend to differences between different target types. Conversely, exposure to consistent VOTs may encourage less detailed analysis of VOT, attenuating

---

[5] In both Experiments 1B and 1C, −1/3 targets did not show significant priming with matching primes. Some listeners may adjust to a one third reduction in VOT and respond faster, which lowers reaction times for control items. Thus, although there was a numerical trend toward significance, the variability in participants' baseline performance prevented this condition from being significant.

[6] Normalization was necessary because the magnitude of priming can vary as a function of baseline performance. In particular, Chapman, Chapman, Curran, and Miller (1994) showed that slower baseline performance can lead to heightened priming difference scores (unprimed − primed).

within-category specificity effects. This explanation finds support from studies of consonant learning. Lively, Logan, and Pisoni (1993) demonstrated the importance of exposure to within-category variability in perceptual learning. In their study of native Japanese speakers' learning of the English /r/–/l/ contrast, listeners trained with high-variability stimulus sets generalized better to new tokens than listeners exposed to a single talker. Similarly, Maye, Werker, and Gerken (2002) showed that exposure to consonants at category boundaries enhanced the learning of stop consonants in infants. In their study, the infants exposed to tokens with the atypical VOTs as well as the prototypical VOTs showed better discrimination performance later than those exposed only to the prototypical VOTs. Thus, exposure to within-category variability in the mixed prime block might have enhanced listeners' sensitivity to changes in VOT between primes and targets. This may explain why we obtained within-category specificity in the mixed prime experiment and not in the blocked prime experiment.

In summary, the results from Experiments 1B and 1C demonstrate that listeners indeed encode within-category variation in long-term memory under certain circumstances. The results also suggest a prominent role for prototypical representations in long-term priming, as evidenced by the finding that the items with prototypical VOTs (intact) resulted in significant priming regardless of matches in VOT. In contrast, aprototypical items produced significant priming only when VOTs matched precisely.

Experiment 2: Effects of Lexical Counterparts

In Experiment 2 we examine whether having a voiced counterpart influences the encoding of within-category specificity in words containing voiceless word-initial stops. Within-category variation can have different perceptual consequences for lexical discrimination depending on whether voiceless words share the VOT space with a word or a nonword. We thus hypothesized that listeners might show different degrees of sensitivity to within-category variation as a function of the presence of a lexical counterpart. Because within-category specificity effects were more prominent in the mixed prime block than in the blocked prime block, the prime block in this experiment contained both intact and $-2/3$ VOT items. Further, because a gradient lexicon hypothesis was no longer a focus, we did not use $-1/3$ VOT items in subsequent experiments.

Experiment 2A: Phoneme Categorization

In this categorization experiment, we made a change to further ensure that the reduced VOT items were perceived as voiceless. Because the reduced items had a substantial portion of their original VOT removed, there was a small possibility that the word-initial consonants in the targets with reduced VOT might be perceived as sounds other than voiceless or voiced. Thus, instead of a forced choice between voiced and voiceless, we gave listeners in this experiment three choices for a button response (voiced, voiceless, and other).[7] The rest of the procedure remained identical to Experiment 1A.

Method

Participants. We recruited 24 participants from the University at Buffalo, State University of New York, community. They received partial credit for a course requirement. Participants were right-handed native speakers of American English, with no reported history of speech or hearing disorders.

Materials. The stimuli consisted of 48 spoken mono- and bisyllabic words containing word-initial voiceless stops and 48 spoken filler words containing word-initial voiced stops. Half of the targets had voiced counterparts (e.g., the target, *cape*, has a voiced counterpart, *gape*), and the other half did not have voiced counterparts (e.g., the target, *kite*, has no voiced counterpart).

Results and Discussion

The stimuli were first scored for correct categorization. As in the previous phoneme categorization experiment, only those stimuli that were correctly categorized as voiceless by at least 90% of the total number of listeners were included for the lexical decision experiment. There were 38 items that met the criteria. Thirty-six were randomly selected for the lexical decision experiments. Mean and range of VOTs of the 36 words are given in Table 1.

Accuracy. Mean correct categorization for the initial stop in the 36 selected /p/–/t/–/k/ words was 98%. We performed a repeated measure ANOVA on reaction times for correct responses and percentages with target types (intact, $-2/3$) and counterpart types (with, without) as within-subject factors. Listeners were significantly less accurate on words with counterparts, $F_1(1, 23) = 8.79$, $p < .01$; $F_2(1, 17) = 14.40$, $p < .01$.

Reaction times. We replaced reaction times greater than 2,000 ms with the appropriate condition mean. We also replaced reaction times beyond two standard deviations from the condition mean with the appropriate condition mean. We replaced less than 3% of the total data.

The mean reaction times to the two types of stimuli were 427 ms ($SD = 167$) and 506 ms ($SD = 160$) for $-2/3$ in the no-counterpart condition and 432 ms ($SD = 150$) and 534 ms ($SD = 158$) for $-2/3$ in the counterpart condition. Listeners' reaction times did not significantly differ as a function of counterpart, $F_1(1, 23) = 2.30$ and $F_2 < 1$), but listeners responded more slowly on $-2/3$ targets than on intact targets, $F_1(1, 23) = 49.54$, $p < .01$; $F_2(1, 17) = 34.01$, $p < .01$. The interaction of target type and counterpart type was not significant, $F_1(1, 23) = 1.02$; $F_2(1, 17) = 2.12$. No other effects were significant.

These results demonstrate that listeners are sensitive to the VOT manipulation in the stimuli—namely, the more prototypical the VOT was, the faster the categorization decision was. The presence of a voiced counterpart also influenced phoneme categorization decision by slowing reaction times for the VOT-reduced items relative to the equivalent condition in the no-counterpart condition.

Experiment 2B: Priming

In Experiment 2B, we address the question of whether encoding of within-category variation differs as a function of having voiced counterparts.

---

[7] In case the number of button choices might have substantially influenced listeners' categorization behaviors, we also reran the phoneme categorization task of Experiment 1A using three button responses. In this experiment, for the 24 targets used in the lexical decision task, 2 items failed to meet 90% accuracy (*piston*, 71%; *tomato*, 83%). When we reanalyzed the data from both Experiments 1B and 1C without these 2 items, the pattern of results did not change.

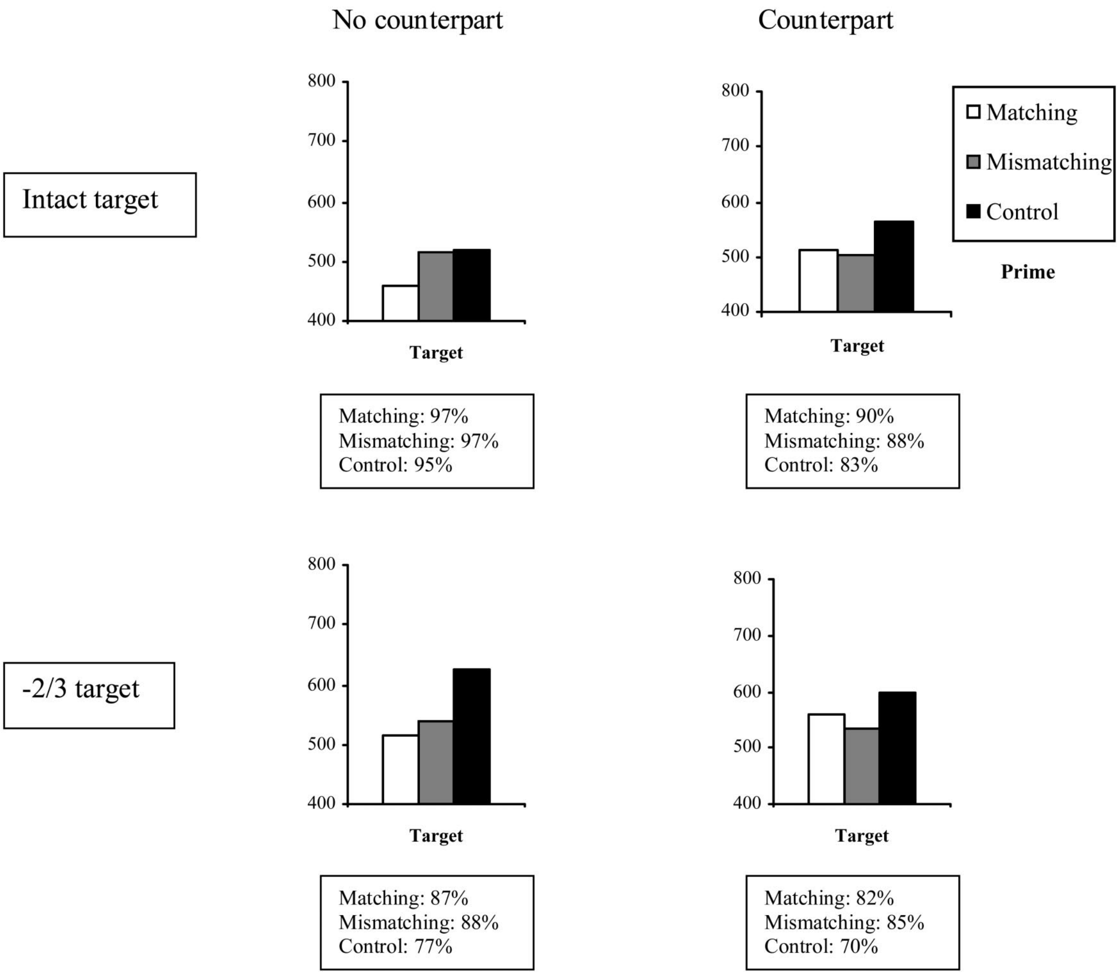

Figure 4. Lexical decision task (Experiment 2B). Values in the Y axis represent response time (msec). −2/3 = two thirds of voice onset time was removed.

Method

Participants. We recruited 60 participants from the University at Buffalo, State University of New York, community. They received partial credit for a course requirement. Participants were right-handed native speakers of American English, with no reported history of speech or hearing disorders.

Materials. The stimuli consisted of 18 spoken words containing word-initial voiceless stops that have voiced counterparts (e.g., *cash*), 18 words containing word-initial voiceless stops without voiced counterparts (e.g., *camel*), and nonwords. We again constructed nonwords by changing the final syllables of the target words (e.g., *camel* vs. **camoid*). A list of the stimuli for this experiment is given in Appendix B. Mean word frequency (Kučera & Francis, 1967) and familiarity ratings (Nusbaum, Pisoni, & Davis, 1984) of the targets in the no-counterpart condition were 9 (*SD* = 10) and 6.9 (*SD* = 2.1), respectively, and those of the targets in the counterpart condition were 7 (*SD* = 7) and 6.8 (*SD* = 3.8), respectively. *t* tests confirmed that the two conditions did not differ significantly either in frequency or in familiarity.

We created two sets of target words in the same manner as in the previous experiments. One set contained intact tokens, whereas the other set consisted of the original tokens with two thirds of the VOT removed from the midpoint of the voicing segment.

Design. The design was identical to that of Experiment 1B. Orthogonal combination of the three levels of prime type (match, mismatch, control), two levels of target type (intact, −2/3), and two levels of counterpart types (with, without) resulted in 12 conditions. For a given participant, three words occurred in each condition.

Procedure. The same procedure was used as in the Experiments 1B and 1C.

Results

We performed a repeated measure ANOVA on reaction times for correct responses and percentages correct for the target word stimuli, with prime type (matching, mismatching, control) and target type (intact, −2/3) as within-subject factors. We replaced reaction times greater than 2,000 ms from the offset of the target with the appropriate condition mean. We also replaced reaction times beyond two standard deviations from the condition mean with the appropriate condition mean. We replaced less than 5% of the total data.

Accuracy. Overall accuracy was 87%. Percentages correct as a function of prime and target type are given in Figure 4. We performed a repeated measure ANOVA with counterpart type (counterpart, no counterpart), target type (intact, −2/3), and prime type (matching, mismatching, control) as within-subject factors. Accuracy did not differ as a function of counterpart type (F_1 and $F_2 < 1$). However, it differed as a function of target type: Listeners were less accurate on −2/3 targets than on intact targets, $F_1(1, 59) = 52.12, p < .01$; $F_2(1, 34) = 5.44, p < .03$. Reaction times significantly differed by prime type, $F_1(2, 118) = 3.60, p < .04$; $F_2(2, 68) = 6.12, p < .01$. Planned contrasts showed that listeners were less accurate on control items than on matching and mismatching items, $F_1(1, 118) = 7.19, p < .01$; $F_2(1, 68) = 12.10, p < .01$; and $F_1(1, 118) = 6.12, p < .01$; $F_2(1, 68) = 14.37, p < .01$, respectively. There was also a significant interaction effect of counterpart type and target type, $F_1(1, 59) = 12.93, p < .07$; $F_2(2, 68) = 3.91, p < .03$. Planned contrasts on the interaction showed that listeners were less accurate on −2/3 targets than on intact targets in the no-counterpart condition, $F_1(1, 59) = 54.63, p < .01$; $F_2(1, 17) = 7.01, p < .02$, as well as in the counterpart condition, $F_1(1, 59) = 5.32, p < .03$ ($F_2 < 1$). Accuracy on intact targets was higher in the no-counterpart condition than in the counterpart condition by subjects only, $F_1(1, 59) = 6.08, p < .02$ ($F_2 < 1$). Accuracy on −2/3 targets was higher in the counterpart condition than in the no-counterpart condition by subjects only, $F_1(1, 59) = 6.86, p < .02$ ($F_2 < 1$). No other effects were significant.

Reaction times. Reaction times as a function of prime and target type are plotted in Figure 4. There was a significant main effect of counterpart type: Listeners were slower at making lexical decisions in the counterpart condition than in the no-counterpart condition, $F_1(1, 59) = 5.62, p < .03$; $F_2(1, 34) = 11.98, p < .01$. Reaction times were shorter on intact targets than on −2/3 targets, $F_1(1, 59) = 46.42, p < .01$; $F_2(1, 34) = 8.29, p < .01$. Reaction times also differed as a function of prime type, $F_1(2, 118) = 24.95, p < .01$; $F_2(2, 68) = 15.84, p < .01$. There was also a significant interaction effect of counterpart type and prime type, $F_1(2, 118) = 3.74, p < .03$, subjects, $F_2(2, 68) = 1.89$. No other effects were significant.

Priming. To examine priming effects in greater detail, we conducted planned contrasts. In the no-counterpart condition, there was significant priming for intact targets in the matching condition, $F_1(1, 118) = 8.98, p < 01$; $F_2(1, 34) = 8.68, p < .01$, but not in the mismatching condition (F_1 and $F_2 < 1$). For −2/3 targets, there was significant priming in both matching and mismatching conditions, $F_1(1, 118) = 30.13, p < .01$; $F_2\ (1, 34) = 13.29, p < .01$; and $F_1\ (1, 118) = 19.25, p < .01$; $F_2\ (1, 34) = 8.71, p < .01$, respectively. In the counterpart condition, there was significant priming in both matching and mismatching conditions for intact targets, $F_1(1, 118) = 7.45, p < .\ 01$; $F_2(1, 34) = 17.14, p < .01$; and $F_1(1, 118) = 9.71, p < .03$; $F_2(1, 34) = 10.67, p < .01$, respectively. For −2/3 targets, the matching condition showed only weak priming (marginally significant by subjects only), $F_1(1, 118) = 3.59, p < .07$, $F_2(1, 34) = 1.64$, whereas the mismatching condition showed significant priming, $F_1(1, 118) = 9.20, p < .01$; $F_2(1, 34) = 6.96, p < .02$.

Specificity. To examine specificity effects, we conducted planned contrasts on the two-way interaction of counterpart type and prime type. The results demonstrated within-category specificity effects in the no-counterpart condition: Listeners responded faster in the matching condition than in the mismatching condition, $F_1(1, 118) = 7.33, p < .\ 01$; $F_2(1, 34) = 8.68, p < .01$. In contrast, there was no significant difference between the matching and mismatching conditions in the counterpart condition. As in Experiment 1B, −2/3 targets did not show specificity effects, as there was no significant difference between the matching and mismatching conditions: Intact tokens with prototypical VOTs primed all items regardless of matches in VOT.

Discussion

These results demonstrate that the presence of lexical counterparts modulates the encoding of within-category variation in long-term memory (see Figure 4). Once again, we found evidence of a prominent role for prototypical representations in long-term auditory priming, as evidenced by the finding that intact items with prototypical VOTs produced significant priming regardless of matches in VOT. We also found within-category specificity effects for intact targets, but only in the no-counterpart condition, which replicates the results from Experiment 1B.

In the counterpart condition, we found no difference between matching and mismatching VOT conditions for both types of targets, which indicates that within-category variation has different consequences for long-term priming depending on whether the opposite end of the VOT continuum constitutes another word or a nonword. Given that overall reaction times to intact targets in the counterpart condition were longer than in the no-counterpart condition, the discrepancy regarding specificity effects between the counterpart and the no-counterpart conditions might have been due to a different time course of processing. That is, the slower the processing is, the more time listeners may have for mapping −2/3 primes onto the prototype. Conversely, if listeners make a fast response, they may not have enough time for mapping −2/3 primes onto the prototype, which results in attenuated priming (i.e., a specificity effect). However, this possibility is unlikely given the finding that reaction times for −2/3 targets in two counterpart-type conditions were not significantly different in either prime or target blocks (all Fs < 1). We revisit the possibility that a time course was responsible for the lack of specificity in the counterpart condition in the following experiment, in which we examine specificity effects only in words with counterparts while manipulating the frequency of their voiced counterparts. Because we predict that the high-frequency counterpart condition will show slower processing, if the time course hypothesis is true, specificity should emerge only in the low-frequency counterpart condition. Further, the following experiment also provides precise information concerning the effects of voiced counterparts on specificity effects.

Experiment 3: Effects of Competitor Frequency

This experiment examines whether the frequency of voiced counterparts influences the encoding of within-category variation in words with word-initial voiceless stops. If the degree of lexical competition affects the degree of within-category information encoded, we should observe differential magnitude of specificity effects as a function of frequency of counterparts.

Experiment 3A: Frequency Rating

The stimuli were divided into two conditions according to the word frequency (Kučera & Francis, 1967) of the voiced counterparts of the voiceless targets: high and low. The high-frequency counterpart condition contained words whose voiced counterparts are higher in frequency (e.g., *pig–big*). The low-frequency counterpart condition contained words whose voiced counterparts are lower in frequency (e.g., *cash–gash*). To independently validate the frequency disparity, we administered a frequency-rating questionnaire prior to a phoneme categorization task.

Method

Participants. Five hundred thirty-three participants from the University at Buffalo, State University of New York, community completed a mass testing questionnaire in return for partial credit for a course requirement. Participants were native speakers of American English, with no reported history of speech or hearing disorders.

Materials. The stimuli consisted of 44 spoken mono- and bisyllabic words containing word-initial voiceless stops. All stimuli had voiced counterparts.

Procedure. In the questionnaire, participants were presented with each pair of words (e.g., *cape–gape*) and asked to judge which word in the pair they thought they heard more frequently. They were given three choices: (a) voiceless more often than voiced (e.g., *peer* more often than *beer*), (b) voiced more often than voiceless (e.g., *beer* more often than *peer*), or (c) voiceless and voiced equally often (e.g., *peer* and *beer* equally often).

Results and Discussion

Words were ranked from high to low in terms of the frequency of the “voiceless more often” response and the frequency of the “voiced more often” response. Nineteen words ranked on the top of each category were chosen for the phoneme categorization task (38 in total). The mean percentages of response for all three response categories for these 38 words are given in Table 3.

Table 3
Frequency Ratings (Experiment 3A): Mean Percentages of Ratings and Standard Deviations

	Rating		
Voiced counterpart	Voiceless more often	Voiced more often	Equally often
Low frequency (%)			
M	70	11	19
SD	10	4	7
High frequency (%)			
M	13	63	24
SD	6	11	11

Experiment 3B: Phoneme Categorization

Method

Participants. We recruited 24 participants from the University at Buffalo, State University of New York, community. They received partial credit for a course requirement. Participants were right-handed native speakers of American English, with no reported history of speech or hearing disorders.

Materials. The stimuli consisted of 38 spoken mono- and bisyllabic words containing word-initial voiceless stops and 38 spoken filler words containing word-initial voiced stops. Half of the voiceless targets consisted of high-frequency counterpart types, and the other half consisted low-frequency counterpart types.

Procedure. The procedure for the phoneme categorization task remained identical to that of Experiment 2A.

Results and Discussion

As in the previous phoneme categorization experiment, only those stimuli that were correctly categorized as voiceless by at least 90% of the listeners were included in the lexical decision experiment. In all, 32 items met the criteria. Twenty-four were randomly selected for the lexical decision experiments. Mean and range of VOTs of the 24 words are given in Table 1.

The mean accuracy for the 24 selected /p/–/t/–/k/ words was 99%. We performed a repeated measure ANOVA on reaction times for correct responses and percentages correct of the 24 selected items, with target type (intact, −2/3) and counterpart type (high, low frequency) as within-subject factors. We replaced reaction times greater than 2,000 ms with the appropriate condition mean. We also replaced reaction times beyond two standard deviations from the condition mean with the appropriate condition mean. We replaced less than 3% of the total data.

Accuracy. There was no overall difference in accuracy as a function of counterpart frequency. However, listeners were more accurate on intact targets than on −2/3 targets, $F_1(1, 23) = 6.18$, $p < .03$; $F_2(1, 11) = 4.23$, $p < .07$. There was also a significant interaction between counterpart type and target type, $F_1(1, 23) = 7.44$, $p < .02$; $F_2(1, 11) = 9.43$, $p < .02$. Planned contrasts showed that listeners were less accurate on −2/3 targets than on intact targets in the high-frequency counterpart condition but not in the low-frequency counterpart condition, $F_1(1, 23) = 8.83$, $p < .01$; $F_2(1, 11) = 10.61$, $p < .01$. No other effects were significant.

Reaction times. The mean reaction times to the two types of stimuli were 416 ms (*SD* = 177) and 459 ms (*SD* = 166) for −2/3 in the low-frequency counterpart condition and 441 ms (*SD* = 169) and 563 ms (*SD* = 165) for −2/3 in the high-frequency counterpart condition.

Reaction times were longer in the high-frequency counterpart condition than in the low-frequency counterpart condition, $F_1(1, 23) = 20.71$, $p < .01$, by subjects only ($F_2 < 1$). Reaction times were also longer on −2/3 targets than on intact targets, $F_1(1, 23) = 28.09$, $p < .01$; $F_2(1, 17) = 22.37$, $p < .01$. The Counterpart Type × Target Type interaction was significant by subjects only, $F_1(1, 23) = 9.22$, $p < .01$ ($F_2 < 1$). Planned comparisons showed that reaction times were longer for −2/3 targets in the high-frequency counterpart condition than in the low-frequency counterpart condition, $F_1(1, 23) = 32.06$, $p < .01$, by subjects, and this was marginally significant by items, $F_2(1, 17) = 3.23$, $p < .09$. No other effects were significant. These results again demonstrate that the more prototypical the VOT was, the faster was the categorization decision. Moreover, listeners were slower to −2/3 targets when the counterparts were high frequency than when they were low frequency.

Experiment 3C: Priming

In Experiment 3C, we address the question of whether encoding of within-category variation differs as a function of frequency of voiced counterparts.

Method

Participants. We recruited 60 participants from the University at Buffalo, State University of New York, community. They received partial credit for a course requirement. Participants were right-handed native speakers of American English, with no reported history of speech or hearing disorders.

Materials. The stimuli consisted of 12 spoken mono- and bisyllabic words containing word-initial voiceless stops and having high-frequency voiced counterparts (e.g., *pig–big*), 12 words containing word-initial voiceless stops and having low-frequency voiced counterparts (e.g., *cab–gab*), and nonwords. We constructed nonwords by exchanging the final syllables of the target words (e.g., *pig* vs. **pid*). (See Appendix C for the final list of the stimuli.)

The mean word frequency (Kučera & Francis, 1967) of these 24 targets in the high-frequency counterpart condition was 26 (*SD* = 59) for voiceless words and 754 (*SD* = 1,491) for their voiced counterparts. The mean word frequency of the targets in the low-frequency counterpart condition was 36 (*SD* = 49) for voiceless words and 9 (*SD* = 18) for their voiced counterparts. Two-tailed *t* tests showed that the frequency of voiced counterparts was significantly higher in the high-frequency counterpart condition than in the low-frequency counterpart condition, $t(11) = 2.47$, $p < .04$, whereas the frequency of voiceless targets did not differ significantly.

The mean familiarity rating scores (Nusbaum et al., 1984) of the 24 targets in the high-frequency counterpart condition were 6.8 (*SD* = 5.7) for voiceless words and 6.9 (*SD* = 2.7) for their voiced counterparts. The mean familiarity rating scores of the targets in the low-frequency counterpart condition were 6.9 (*SD* = 1.6) for voiceless words and 6.0 (*SD* = 1.0) for their voiced counterparts. Two-tailed *t* tests showed that the familiarity ratings of voiced counterparts were significantly higher in the high-frequency counterpart condition than in the low-frequency counterpart condition, $t(11) = 4.21$, $p < .01$, whereas the familiarity ratings of voiceless targets did not differ significantly.

We created two sets of target words in the same manner as in the previous experiments. One set contained intact tokens, whereas the other set consisted of the original tokens with two thirds of the VOT removed from the midpoint of the voicing segment.

Design. The design was identical to Experiment 2B. Orthogonal combination of the three levels of prime type (matching, mismatching, control), two levels of target type (intact, −2/3), and two levels of stimulus types (high-frequency voiced counterparts, low-frequency voiced counterparts) resulted in 12 conditions. The primes were presented in the mixed mode as in Experiment 2B.

Procedure. The same procedure was used as in the Experiment 2B.

Results

We performed a three-way repeated measure ANOVA on reaction times for correct responses and percentages correct for the target word stimuli, with prime type (matching, mismatching, control), target type (intact, −2/3), and counterpart frequency (high-frequency counterpart, low-frequency counterpart) as within-subject factors. We replaced reaction times greater than 2,000 ms from the offset of the target with the appropriate condition mean. We also replaced reaction times beyond two standard deviations from the condition mean with the appropriate condition mean. We replaced less than 5% of the total data.

Accuracy. Overall accuracy was 87%. Listeners were less accurate in the high-frequency counterpart condition than in the low-frequency counterpart condition, $F_1(1, 59) = 62.64$, $p < .01$; $F_2(1, 22) = 4.36$, $p < .05$. Listeners were also less accurate on −2/3 targets than on intact targets, $F_1(1, 59) = 18.35$, $p < .01$; $F_2(1, 22) = 14.63$, $p < .01$. No other effects were significant.

Reaction times. Reaction times as a function of prime type, target type, and counterpart frequency are plotted in Figure 5. Reaction times were longer in the high-frequency counterpart condition than in the low-frequency counterpart condition, $F_1(1, 59) = 115.12$, $p < .01$; $F_2(1, 22) = 7.72$, $p < .02$. Reaction times also differed significantly as a function of prime type, $F_1(2, 118) = 11.66$, $p < .01$; $F_2(2, 44) = 13.78$, $p < .01$, and target type, $F_1(1, 59) = 20.82$, $p < .01$; $F_2(1, 22) = 7.47$, $p < .02$. There was a significant two-way interaction of counterpart type and prime type by subjects only, $F_1(2, 118) = 5.71$, $p < .03$; $F_2(2, 44) = 1.98$. There was also a significant two-way interaction of target type and prime type, $F_1(2, 118) = 6.76$, $p < .01$; $F_2(2, 44) = 5.44$, $p < .01$. A three-way interaction of counterpart frequency type, target type, and prime type was also significant, $F_1(2, 118) = 7.47$, $p < .01$, $F_2(2, 44) = 4.27$, $p < .03$.

Priming. To examine priming effects in detail, we conducted planned comparisons on the three-way interaction. In the low-frequency counterpart condition, intact targets did not show significant priming in either the matching or the mismatching conditions. We found significant priming for −2/3 targets both in the matching condition, $F_1(1, 118) = 5.82$, $p < .02$; $F_2(1, 22) = 3.55$, $p < .08$, and in the mismatching condition, $F_1(1, 118) = 10.40$, $p < .01$; $F_2(1, 11) = 6.51$, $p < .02$. Priming patterns were very different in the high-frequency counterpart condition. Intact targets in this condition showed significant priming in the matching condition, $F_1(1, 118) = 6.76$, $p < .01$; $F_2(1, 22) = 5.07$, $p < .04$, but not in the mismatching condition, $F_1(1, 118) = 2.80$; $F_2(1, 22) = 1.62$. Priming patterns for −2/3 targets were also different from those in the low-frequency counterpart condition. This time, only the mismatching condition showed significant priming, $F_1(1, 118) = 31.11$, $p < .01$; $F_2(1, 22) = 14.50$, $p < .01$.

Specificity. We found no significant effects between the matching and mismatching conditions in the low-frequency counterpart condition for either type of target (all *F*s < 1), which indicates no within-category specificity effects in the low-frequency counterpart condition. In the high-frequency counterpart condition, although there was no significant difference between the matching and mismatching conditions for intact targets, $F_1(1, 118) = 1.81$; $F_2 < 1$, their priming patterns suggested specificity effects. Intact targets showed significant priming only in the matching condition, not in the mismatching condition. It is interesting to note that reaction times for −2/3 targets in the high-frequency counterpart condition were significantly longer in the matching condition than in the mismatching condition, $F_1(1, 118) = 43.35$, $p < .01$; $F_2(1, 22) = 23.62$, $p < .01$, an opposite effect of specificity effects (i.e., attenuated priming in the matching condition).

Discussion

As in the previous two experiments, the results demonstrate that within-category variation is encoded in long-term memory and affects magnitude of priming. The crucial finding is that long-term priming of voiceless words was modulated by the frequency of their voiced counterparts, demonstrating the effect of lexical competition on long-term priming, not unlike that in short-term lexical activation (e.g., Luce & Pisoni, 1998). Indeed, the effect of counterpart frequency was robust. In all conditions, listeners' reaction times were longer in the high-frequency counterpart condition than in the low-frequency counterpart condition, although the difference was most amplified for −2/3 targets in the matching condition. Reaction times for −2/3 targets in the matching condition were particularly long. It is interesting to note that these inhibitory

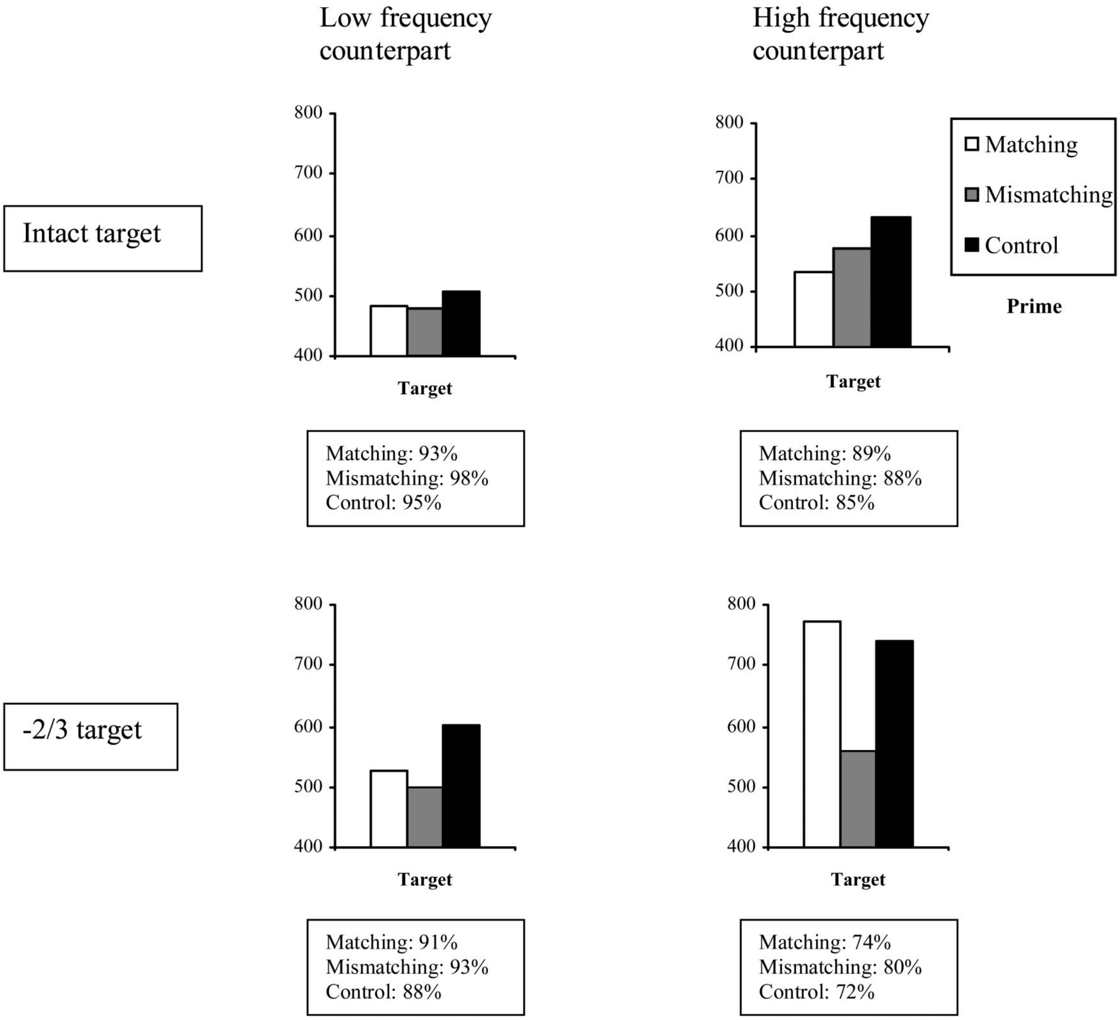

Figure 5. Lexical decision task (Experiment 3C). Values in the Y axis represent response time (msec). −2/3 = two thirds of voice onset time was removed.

effects of lexical competition on long-term priming in the matching condition were also present (albeit less prominently) in Experiment 2B (see Figure 4), in which −2/3 targets showed weaker priming in the matching condition (marginally significant) than in the mismatching condition (significant). However, the data patterns from words with counterparts contrasted with those from words with no counterparts (see Figures 2 and 4), where all matching conditions for −2/3 targets showed significant priming. Thus, these results also confirm that the presence of counterparts and their competition effect attenuated priming in the matching condition for −2/3 targets relative to cases when there was no such competition effect.

It is interesting that we obtained within-category specificity effects for intact targets in the counterpart condition (high frequency), unlike in Experiment 2B (see Figure 4), in which only the no-counterpart condition showed specificity effects. To compare results from all counterpart conditions, we replotted the data from Experiments 2B and 3C for words with counterparts in Figure 6. Note that we obtained no specificity for intact targets in either the counterpart condition in Experiment 2B or the low-frequency counterpart condition in Experiment 3C. However, specificity emerged again when the counterpart frequency and, thus, lexical competition were high. Thus, the fact that we did not obtain specificity in the counterpart condition in Experiment 2B appears to have been due to differences in the composition of the stimuli between the two experiments—in particular to the difference in the frequency of voiced counterparts. A close inspection of the stimuli in Experiment 2B indeed revealed that the majority contained low-frequency counterparts (high, 39%; low, 61%).

The fact that specificity emerged with competition from voiced competitors poses an interesting question. Thus far, specificity has been present in words without counterparts and in words with high-frequency counterparts, not in words with low-frequency counterparts. What might make listeners more sensitive to within-category variation in those two types of stimulus? One possible explanation is a differing degree of lexical discriminability for the voiceless targets. Note that we obtained specificity when listeners were given −2/3 primes (as well as −1/3 primes in Experiment

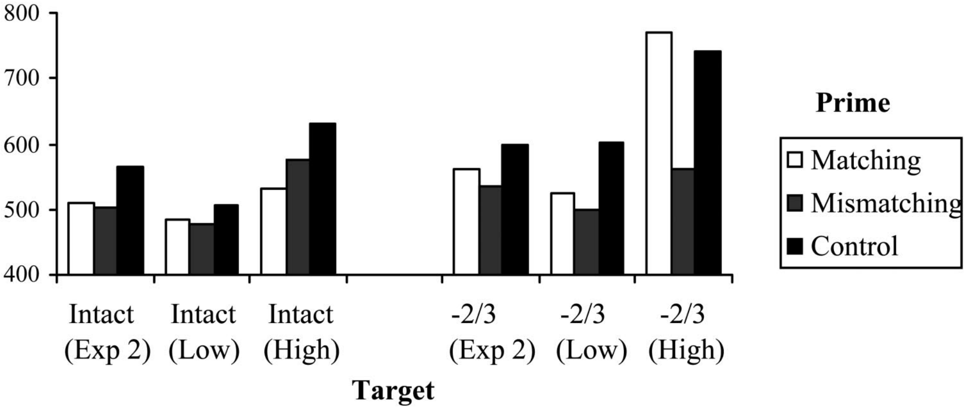

Figure 6. Counterpart conditions (Experiments 2B and 3C combined). Values in the Y axis represent response time (msec). Exp = Experiment; −2/3 = two thirds of voice onset time was removed.

1B) followed by intact targets. That is, VOT-reduced primes resulted in a lesser degree of priming effects than did intact primes.

In the no-counterpart condition, VOT-reduced items might have led to the recognition of nonwords, whereas in the high-frequency counterpart condition, these items might have activated strongly voiced lexical competitors. In short, when the counterpart constitutes a nonword or a strong competitor, the competition attenuates priming and, hence, specificity effects. By contrast, words with low-frequency counterparts pose a lesser degree of competition in the process of recognizing the voiceless target because the other end of the VOT continuum is a word (hence, there is a lower risk of making a wrong, nonword decision), and competition from counterparts is marginal because of their low frequency. Nevertheless, these results are consistent with those from the previous experiments in that they suggest a role for both prototypical and specific memory representations in long-term priming.

General Discussion

Although numerous studies have shown that listeners are sensitive to fine-grained acoustic–phonetic information in spoken word recognition (e.g., Andruski et al., 1994; Dahan et al., 2001; Ju & Luce, 2004; Marslen-Wilson & Warren, 1994; McMurray et al., 2002; McQueen et al., 1999; Streeter & Nigro, 1979; Utman, 1997; Utman et al., 2000; Whalen, 1991), it is as yet unclear whether this sensitivity extends to long-term memory. Two possibilities exist. The fine-grained acoustic–phonetic information may be mapped onto long-term abstract representations. Alternatively, the system may store veridical representations of words, in which case within-category variation is represented in memory traces of words themselves. To tease apart these two possibilities, we hypothesized that if changes in the VOTs of word-initial consonants have consequences for subsequent recognition of a word containing the consonant, this will provide evidence for the encoding of within-category variation. Conversely, if within-category variation does not influence subsequent word recognition, this will suggest that listeners encode speech information in terms of abstract, categorical representations.

The results of this study reveal that listeners do indeed encode within-category variation in long-term memory and that the subphonemic variation has consequences for subsequent spoken word recognition. In particular, a substantial mismatch in VOT resulted in attenuation in priming effects, providing evidence for within-category specificity effects. These specificity effects were present under the following circumstances: −2/3 primes → intact targets, −1/3 primes → −2/3 targets in words with no counterparts (see Figures 2 and 4), and −2/3 primes → intact targets in words with high-frequency counterparts (see Figure 5). These findings demonstrate that within-category information is preserved in long-term form-based representations.

Listeners were also sensitive to relatively subtle variation in VOT, which suggests that the magnitude of specificity effects may vary in a gradient manner as a function of degree of VOT reduction. In Experiment 1B, for example, two mismatching VOT primes (intact, −1/3) produced a significantly different magnitude of priming with −2/3 targets. Although the same effect was not significant with intact targets, the numerical trend was consistent. The magnitude of priming decreased as a function of VOT reduction. That is, the greater the reduction was, the greater was the degree of attenuation in priming, suggesting that form-based word representations in long-term memory may also be gradient, further supporting the view that word representations are not strictly categorical.

The finding that listeners are sensitive to subtle variation in VOT in long-term auditory priming contrasts with Andruski et al.'s (1994) results demonstrating no significant difference between intact and −1/3 stimuli in the magnitude of semantic facilitation. This discrepancy might have occurred for two reasons. First, Andruski et al.'s study might have lacked the power to detect a difference. Another, nonmutually exclusive possibility is that this discrepancy is due to the differences in the modalities of priming examined in their study and ours. It is possible that subtle variation does not play as strong a role in mediating semantic priming, unlike in form-based priming. This possibility finds support from a study of short-term, form-based lexical activation that found gradient effects of within-category variation (McMurray et al., 2002). Thus, it appears that priming in both short-term and long-term memory can show gradient effects of within-category variation.

Note, however, that much of the specificity was obtained in the mixed prime experiment, whereas there was only weak evidence for specificity in the blocked experiment. To determine whether specificity effects are the norm rather than the exception, one may ask which of these conditions more closely approximates natural listening conditions. Given that listeners are exposed to the speech stream consisting of different VOTs from different speakers (Allen, Miller, & DeSteno, 2003) and that VOTs, even within a single talker, can vary substantially depending on speaking rate and syllable duration (Kessinger & Blumstein, 1997; Miller & Baer, 1983; Miller, Green, & Reeves, 1986; Miller & Volaitis, 1989; Summerfield, 1981; Volaitis & Miller, 1992), the findings from the mixed prime experiment do indeed shed an interesting insight into how variation in VOTs may be treated during spoken word recognition. Indeed, speech input contains highly variable signals (for a review, see Pisoni & Luce, 1987) occurring in much noisier contexts than isolated words presented in carefully controlled experimental settings. Thus, the specificity effects obtained in the mixed prime experiment may more accurately reflect normal listening conditions.

Another important finding of the present study is that the encoding of within-category variation in long-term memory was modulated by the lexical activation of other competing words—in particular, the presence and frequency of voiced counterparts. Listeners were more tolerant of within-category variation when there was a voiced counterpart. That is, when both ends of the voiceless–voiced continuum constituted a word, variation in VOT did not seem to crucially influence long-term priming, resulting in a lesser degree of sensitivity. However, when faced with stronger competition from a high-frequency counterpart, listeners showed sensitivity to within-category variation (see Figure 6), demonstrating the integral role for lexical competition in long-term priming. Taken together, these findings suggest that when listeners encode within-category variation, they not only take the lexicality of the other end of the VOT continuum into consideration but also weigh the overall level of lexical activation, including that of lexical competitors. When there is strong competition, variation in VOT has a greater impact on priming than when there is a weaker competition effect. The findings that the frequency of counterparts and the lexicality of voiceless targets modulated encoding of within-category variation demonstrate that the speech recognition system is flexible and adjusts the degree of processing of the stimulus depending on the speech context.

Finally, our findings demonstrate that long-term priming was also mediated by prototypical representations, consistent with theories of phonetic prototypes (e.g., Grieser & Kuhl, 1989; Kuhl, 1991; McQueen, Cutler, & Norris, 2004; Samuel, 1982). Kuhl (1991) reported a similar perceptual effect of prototypicality on vowel perception—that is, the *perceptual magnet effect*. She proposed that phonetic category structure is warped in such a way that prototypes draw in the stimuli close to them in the acoustic space. The finding of the present study showing that $-1/3$ primes facilitated priming of intact targets, whereas $-2/3$ primes at the edge of the category did not, is consistent with aspects of Kuhl's theory. In a similar vein, Samuel (1982) demonstrated that prototypical stop consonants yielded better selective adaptation results than nonprototypes. In dichotic listening tasks, Miller (1977) and Repp (1977) also found that prototypical stop consonants proved to be more efficient dichotic competitors than nonprototypes. In perceptual learning experiments conducted in Dutch, McQueen, Cutler, and Norris (2004) also found evidence for prototypical segmental representations.

The evidence for the role for both prototypical and specific memory representations in long-term priming is consistent with two primary theories in spoken word recognition: episodic theories (Goldinger, 1996; Pisoni, 1990) and distributed system theories (e.g., distributed cohort model; Gaskell & Marslen-Wilson, 2002). Both theories assume that specific acoustic details are stored in memory and can account for both the specificity and the generality of word memory traces. However, they differ in their basic representational assumptions. That is, episodic theories assume that each time a word is encountered, a memory trace for the word is stored. Consequently, long-term representations are specific and veridical. However, prototype abstraction can also occur under episodic theories (albeit epiphenomenally), because word perception depends on an aggregate of relevant, individual memory traces at retrieval. Thus, this view accounts for both the specificity and the generality of memory traces using only exemplar representations for form-based word information. Sensitivity to within-category variation in short-term priming arises from goodness of fit to the best exemplar in the category. By contrast, within-category specificity in long-term memory priming arises from mismatches between primes and targets.

Distributed system theories (e.g., distributed cohort model; Gaskell & Marslen-Wilson, 2002) differ from episodic theories in their representational assumptions. Under this view, words are represented in terms of the distributed patterns of activity in the connectionist network. This theory can account for our findings in two ways. First, as in episodic theories, abstract representations emerge epiphenomenally from the most prototypical patterns of activation for a given word. These prototypical representations are more strongly activated than atypical ones, reflecting sensitivity to within-category variation in short-term lexical activation. In long-term priming, the more recently activated patterns have a stronger impact than other patterns. Within-category specificity thus arises from the fact that the activation patterns for the tested token do not precisely match the patterns that were previously stored from the prime. Alternatively, words may be stored as normalized lexical phonological representations in the network, which could explain the pervasive role for the prototypical representations in our results. Listeners would map acoustic signals onto these normalized lexical phonological representations, and variation among the signals that map onto the same normalized lexical phonological representations would then be reflected in the connection weights between the two levels of representations, hence specificity effects.

In short, the speech recognition system appears to be a flexible, hybrid one that preserves both the generality and the specificity of representations depending on the perceptual distance of a stimulus from its prototypical representation. One potential explanation for the mixed effects of prototypes and specific memory representations is that long-term repetition priming is, in part, affected by the episodic memory trace from the prime, as in short-term repetition priming (e.g., Forster & Davis, 1984; Fowler, Napps, & Feldman, 1985). That is, although the abstract lexicon drives long-term priming, variation is also reflected in the magnitude of priming via the episodic trace from the prime. Indeed, we predict that the precise role for abstract and specific memory representations in the long-term mental lexicon will become an increasingly pressing question for researchers attempting to model representation and process in spoken word recognition.

References

Allen, J. S., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics, 63,* 798–810.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America, 113,* 544–552.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52,* 163–187.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America, 67,* 648–662.

Chapman, L. J., Chapman, J. P., Curran, T. E., & Miller, M. B. (1994). Do children and the elderly show heightened semantic priming? How to answer the question. *Developmental Review, 14,* 159–185.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 496–509.

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

Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology, 26,* 274–284.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16,* 507–534.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 680–698.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3–28.

Fowler, C. A., Napps, S. E., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13,* 241–255.

Fowler, C. A., & Rosenblum, L. D. (1991). Perception of the phonetic gesture. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory* (pp. 33–59). Hillsdale, NJ: Erlbaum.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123–136). Hillsdale, NJ: Erlbaum.

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

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology, 45,* 220–266.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1521–1562.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech sounds by infants: Support for speech sound prototypes. *Developmental Psychology, 25,* 577–588.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition, 12,* 568–574.

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science, 15,* 314–318.

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 1–35). Hillsdale, NJ: Erlbaum.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics, 25,* 143–168.

Kučera, F., & Francis, W. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93–107.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358–368.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. *Cognition, 21,* 1–36.

Lisker, L., & Abramson, A. S. (1970, August). *The voicing dimension: Some experiments in comparative phonetics.* Paper presented at the International Congress of Phonetic Sciences, Prague, Czechoslovakia.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94,* 1242–1255.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62,* 615–625.

Luce, P. A., & Lyons, E. (1998). Specificity of memory representation for spoken words. *Memory & Cognition, 26,* 708–715.

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

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101,* 653–675.

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82,* B101–B111.

McClelland, J. L., & Elman, J. L. (1986). Interactive processes in speech recognition: The TRACE model. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 58–121). Cambridge, MA: MIT Press.

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

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86,* B32–B42.

McQueen, J. M., Cutler, A., & Norris, D. (2004). *Decoding speech via abstract representations of speech sounds: Evidence from perceptual learning.* Manuscript submitted for publication.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1363–1389.

Miller, J. L. (1977). Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America, 62,* 641–648.

Miller, J. L. (1997). Internal structure of phonetic categories. *Language and Cognitive Processes, 12,* 865–869.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *Journal of the Acoustical Society of America, 73,* 1751–1755.

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica, 43,* 106–115.

Miller, J. L., & Volaitis, L. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46,* 505–512.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365–378.

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. In *Research on Speech Perception Progress Report* (Vol. 10, pp. 357–376). Bloomington: Speech Research Laboratory, Indiana University.

Palmeri, T., Goldinger, S., & Pisoni, D. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 309–328.

Pilotti, M., Bergman, E. T., Gallo, D. A., Sommers, M., & Roediger, H. L., III. (2000). Direct comparison of auditory implicit memory tests. *Psychonomic Bulletin & Review, 7,* 347–353.

Pisoni, D. B. (1990). Effects of talker variability on speech perception: Implications for current research and theory. In H. Fujisaki (Ed.), *Proceedings of the 1990 International Conference on Spoken Language Processing* (pp. 1399–1407). Tokyo, Japan: Acoustical Society of Japan.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego, CA: Academic Press.

Pisoni, D. B., & Luce, P. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25,* 21–52.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15,* 285–290.

Repp, B. H. (1977). Dichotic competition of speech sounds: The role of acoustic stimuli structure. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 37–50.

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics, 31,* 307–314.

Sheffert, S. M. (1998). Voice-specificity effects on auditory word priming. *Memory & Cognition, 26,* 591–598.

Sommers, M. S. (1999). Perceptual specificity and implicit memory priming in older and younger adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1236–1255.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America, 111,* 1872–1891.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64,* 1358–1368.

Streeter, L. A., & Nigro, G. N. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America, 65,* 1533–1541.

Studdert-Kennedy, M. (1976). Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243–293). New York: Academic Press.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 67–84). San Diego, CA: Academic Press.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074–1095.

Utman, J. A. (1997). *Effects of subphonetic acoustic differences on lexical access in neurologically intact adults and patients with Broca's aphasia.* Unpublished doctoral dissertation, Brown University.

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics, 62,* 1297–1311.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 92,* 723–735.

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics, 50,* 351–360.

(*Appendixes follow*)

Appendix A

Stimuli Used in Experiments 1B and 1C

Target words	Nonwords	Word fillers	Nonword fillers
cabinet	kæbɪnəm	canteen	kænten
canon	kæmən	capsule	kæpʃoɪl
consolation	kandoleʃən	pebble	pɛbənt
cruise	kɹuv	pigeon	pɪʃəm
casualty	kæʒʊəbɪ	tulip	tulɪs
chaos	kemɨs	turtle	tɚtɪp
coffin	kavɪn		
copper	kadɚ		
pardon	parsən		
pendulum	pɛnmudəm		
perjury	pɚʃəɹɪ		
patron	pætɹɨp		
piston	pɪzmən		
pliers	ploɪɚz		
polish	palɪʃ		
pollen	padɪn		
temper	tɛmlɚ		
tortoise	tɔɹtlɨs		
torture	tɔɹtɹaɪd		
tomato	tɔmetəm		
temptation	tɛmpmeʃən		
tunnel	tɨlən		
turmoil	tɚpoɪl		
twilight	twaɪlut		

Appendix C

Stimuli Used in Experiment 3C

Target words	Nonwords	Word fillers	Nonword fillers
cab	kib	cape	keʃ
cable	kepʟ	clamor	klæməl
cash	kædʒ	curl	kɚ
pan	pæf	peach	pidʒ
pass	pæv	pest	pɛsp
pay	poɪd	teal	tip
pile	paɪm	tear	tiʃ
pillow	pɪlloɪd	toe	taʊ
plot	plɪt		
tank	tænʃ		
towel	taʊvəl		
tub	tɨp		
cave	keb		
cobble	kamʟ		
cot	kav		
pack	pædʒ		
pad	pæv		
peer	pɪʃ		
pest	pɛsp		
pie	paɪd		
pig	pɪd		
pill	pɪm		
town	taʊʃ		
tusk	tɨst		

Note. Voiceless words with a high-frequency counterpart appear in italics.

Appendix B

Stimuli Used in Experiment 2B

Target words	Nonwords	Word fillers	Nonword fillers
cane	keb	cage	keʃ
cape	keʃ	cot	kav
cave	ked	crane	kɹem
clamor	klæməl	curl	kɚ
peach	pidʒ	teal	tip
pear	paɪɚ	toe	taʊ
peer	piʃ	*calf*	kæz
pest	pɛsp	*cart*	kaɹk
pig	pɪd	*cast*	kæsp
pox	pas	*coil*	koɪm
puff	pɨt	*taint*	tenʃ
punt	pɨnʃ	*tweed*	twɪs
tab	tæm		
tear	tib		
tire	taɪg		
tub	tɨl		
tuck	tɨt		
tusk	tɨst		
capsule	kæpʃoɪl		
clown	kaʊg		
corpse	kɔɹp		
kite	kaɪp		
pearl	pɚn		
peel	pim		
pint	paɪnʃ		
plumb	plɨt		
plunge	plɨnʃ		
port	pɔɹg		
prop	pɹat		
pup	pɨd		
tar	tɚ		
tights	taɪgs		
tilt	tɪlk		
tool	tudʒ		
torch	tɔɹʃ		
twist	twɪsp		

Note. Voiceless words with a high-frequency counterpart appear in italics.

Received June 21, 2004 ■