Examining the Time Course of Indexical Specificity Effects in Spoken Word Recognition

Conor T. McLennan
Cleveland State University, c.mclennan@csuohio.edu

Paul A. Luce
University at Buffalo, luce@buffalo.edu

Publisher's Statement

Recommended Citation

Examining the Time Course of Indexical Specificity Effects in Spoken Word Recognition

Conor T. McLennan and Paul A. Luce
University at Buffalo, State University of New York

Variability in talker identity and speaking rate, commonly referred to as *indexical variation*, has demonstrable effects on the speed and accuracy of spoken word recognition. The present study examines the time course of indexical specificity effects to evaluate the hypothesis that such effects occur relatively late in the perceptual processing of spoken words. In 3 long-term repetition priming experiments, the authors examined reaction times to targets that were primed by stimuli that matched or mismatched on the indexical variable of interest (either talker identity or speaking rate). Each experiment was designed to manipulate the speed with which participants processed the stimuli. The results demonstrate that indexical variability affects participants' perception of spoken words only when processing is relatively slow and effortful.

Despite a highly variable speech signal, listeners recognize spoken words both quickly and accurately. According to the normalization hypothesis, variability in the speech waveform is treated as "noise" that is stripped away to contact an underlying or symbolic representation. Furthermore, according to the normalization hypothesis, form-based representations do not consist of the surface details attributed to variability (e.g., talker-specific details) but instead are thought to be idealized, abstract, and underspecified. Consequently, the normalization hypothesis is unable to account for empirical evidence demonstrating that variability has consequences for spoken language representation. Nonetheless, normalization is consistent with other evidence demonstrating that numerous types of variability, including allophonic and indexical variability, have consequences for spoken language processing.

The present series of experiments was motivated by past research examining the perceptual and representational consequences of allophonic and indexical variability. *Indexical variability* refers to variations in a spoken word that arise from differences among talkers, speaking rates, affective states, and so on (Abercrombie, 1967; Pisoni, 1997). Typically, indexical variability for a given word has no consequences for its denotation. Whether the noun *telephone* is spoken by a male or female, at a fast or a slow rate of speech, or in a happy or a sad emotional state has no implication for the fact that the word refers to a device used for communication over distances. In other words, indexical variability does not comprise part of the formal linguistic content of an utterance.

Allophonic variability (or, more commonly, *allophonic variation)* refers to articulatory and acoustic differences among speech sounds belonging to the same phonemic category (Ladefoged, 2000). The stop consonant /t/ is articulated somewhat differently—and hence has a different acoustic manifestation—before a vowel (as in *top*), after a vowel (as in *pot*), and in a consonant cluster (as in *stop*). These different versions are referred to as *allophones* of the phoneme /t/. Allophonic variability may result in lexical ambiguity, ultimately having consequences for lexical discrimination (see McLennan, Luce, & Charles-Luce, 2003). In contrast, indexical variability does not generally lead to lexical ambiguity and therefore typically has no consequence for lexical discrimination.

Although distinctions exist between allophonic and indexical variability (Luce, McLennan, & Charles-Luce, 2003), the perception of indexical and linguistic properties is fundamentally linked (Nygaard & Pisoni, 1998; Remez, Fellowes, & Rubin, 1997): Indexical information complements the processing of linguistic content during a spoken exchange. Nonetheless, little attention has been paid to this distinction between allophonic and indexical sources of variability.

Previous research has examined both the processing and the representational consequences of variability. Research on processing has demonstrated that increased variability in the signal places a higher demand on the normalization process at the time of perception (e.g., Mullennix, Pisoni, & Martin, 1989), and research on representation has examined the long-lasting effects of vari-

ability on the representations underlying language perception (e.g., Church & Schacter, 1994).

Processing

Peters (1955) compared the intelligibility of single-talker and multiple-talker messages in noise and found that single-talker messages were reliably more intelligible than were multiple-talker messages. Creelman (1957) compared the intelligibility of words spoken by either single or multiple talkers. He found an inverse relationship between identification performance and the number of talkers: As the number of talkers increased, identification performance decreased. More than 2 decades later, Pisoni and colleagues revisited research on the relationship between talker variability and the perception of isolated words. For example, Mullennix, Pisoni, and Martin (1989) found that participants' identification performance for blocks of familiar English words was both faster and more accurate in single-talker than in multiple-talker conditions (see also, McLennan, 2003; Nusbaum & Morin, 1992; Pisoni, 1990, 1992). Performance costs (measured in terms of decreased accuracy, increased reaction times, or both) associated with processing words spoken by multiple talkers relative to a single talker have also been obtained in preschool children (Ryalls & Pisoni, 1997), in hearing-impaired adults (Kirk, Pisoni, & Miyamoto, 1997), and in elderly adults (Sommers, 1996). Finally, identification of vowels (Verbrugge et al., 1976) and consonants (Fourcin, 1968) has been shown to be more accurate when produced by a single talker than when produced by multiple talkers.

The processing work on talker variability clearly demonstrates that this source of indexical variability has perceptual consequences for spoken word recognition and is consistent with the notion of normalization mentioned earlier. If normalization is assumed to be a time-consuming and resource-demanding process, it follows that in cases in which normalization is required (e.g., when multiple talkers are present) perception of the spoken message should also be slower and less accurate. However, research on the representational aspect of indexical variability, particularly talker identity, provides compelling evidence against normalization.

Representation

Research on the representation of indexical variability demonstrates that surface details associated with indexical variability (e.g., talker identity) are represented in memory and have consequences for subsequent perception. Martin, Mullennix, Pisoni, and Sommers (1989) found that the recall performance of word lists spoken by a single talker was superior to the recall performance of word lists spoken by multiple talkers. Other studies have found that recognition accuracy decreases when a word is repeated in a different voice rather than when it is repeated in the same voice (see, e.g., Bradlow, Nygaard, & Pisoni, 1999; Craik & Kirsner, 1974; Goldinger, Pisoni, & Logan, 1991; Palmeri et al., 1990, 1993). Performance costs resulting from stimuli that mismatch on talker identity are referred to as *voice effects*.

Church and Schacter (1994) examined voice effects with the long-term repetition priming paradigm. In this paradigm, participants are presented with a block of spoken words to which they must respond (the study phase). After this initial exposure, partic-

ipants are presented with another block of words (the test phase). In the second block, some of the words from the first block are repeated. Typically, repeated words are responded to more quickly and accurately than are new words. This repetition priming effect presumably arises because repeated activation of form-based representations in memory facilitates processing. Any significant attenuation in priming for stimuli that mismatch on some dimension (e.g., rate of speech) is referred to as evidence for specificity.

In a long-term repetition priming experiment, Schacter and Church (1992; see also Church & Schacter, 1994) observed significant voice effects in their implicit stem-completion task but not in their explicit word-recognition or cued-recall tasks (see also, Jackson & Morton, 1984). More recently, Goldinger (1996) observed voice effects in both implicit and explicit tasks. Unlike Schacter and Church, Goldinger presented words in noise at both study and test, which may be responsible for the differences across the two sets of studies (see Franks, Bilbrey, Lien, & McNamara, 2000). A number of other studies have also obtained specificity effects, providing additional evidence that indexical information is stored in memory and has consequences for subsequent perceptual processing (see, e.g., Bradlow, Nygaard, & Pisoni, 1999; Fujimoto, 2003; Houston & Jusczyk, 2003 [in infants]; Yonan & Sommers, 2000 [in elderly adults]). These findings demonstrate that indexical information, particularly talker identity, is retained in memory and has a number of consequences for subsequent processing, contrary to the speaker normalization hypothesis.

The previous research provides evidence for both the representational and processing implications of indexical variability. Recent work also suggests that effects of variability appear to follow a time course, manifesting themselves at predictable points during perceptual processing. In particular, indexical specificity effects appear to emerge relatively late in processing. For example, research has demonstrated specificity effects for stimuli that are processed relatively slowly (i.e., lower frequency bisyllabic words; Luce, Charles-Luce, & McLennan, 1999) but not for stimuli that are processed more quickly (i.e., higher frequency monosyllabic words; Luce & Lyons, 1998).

Conversely, previous research has shown that allophonic specificity effects emerge rapidly but subsequently become attenuated as underlying representations come to dominate processing. In two separate experiments, we provide evidence for the representational status of flaps (McLennan et al., 2003; see also, Connine, in press). *Flaps* are neutralized allophones of intervocalic /t/s and /d/s. In casually produced American English, when a /t/ or a /d/ is produced between two vowels, as in the word *rater*, it is often realized as a flap, a segment that is neither exactly a /t/ nor exactly a /d/ (see Patterson & Connine, 2001). In our experiments on flapping, participants performed a lexical-decision task in which they were instructed to decide as quickly and as accurately as possible whether a given stimulus was a word or a nonword. We varied the degree of difficulty of word–nonword discriminations by manipulating the word likeness of the nonwords. In the easy discrimination experiment, the nonwords were "unwordlike" (e.g., *thushthudge)*. In the difficult discrimination experiment, the nonwords were wordlike (e.g., *bacov)*. Crucially, we found that participants made their lexical decisions more quickly in the easy discrimination experiment than in the difficult discrimination experiment.

Our results demonstrate that when responses were rapid, the more specific allophonic representation dominated processing. However, when responses were slowed, evidence for more abstract, underlying representations (i.e., /t/ and /d/) emerged. Thus, we observed specificity effects in the easy but not in the difficult discrimination experiment, a pattern of results consistent with the hypothesis that the more frequent allophonic information dominates initial processing (i.e., as in the easy discrimination experiment) and that effects of the less frequent underlying information emerge only after some delay (i.e., as in the difficult discrimination experiment). McLennan et al. (2003) have proposed an account of the time course of allophonic variability in terms of the adaptive resonance theory (ART) framework (Grossberg, 1986).

According to the ART framework, acoustic–phonetic input activates chunks of features corresponding to sublexical and lexical representations. A *chunk* is a learned set of associated features that may vary in size; a given chunk may correspond to an individual feature, an allophone, or a word. Chunks resonate with the input, with the resonance between input and chunk constituting the percept. Furthermore, more frequent features and combinations of features (i.e., chunks) in a pattern establish resonance with the input more easily and more quickly than less frequent features. Although we assume that eventually all chunks and features matching the input will establish resonances and will affect perceptual processing at some point, it may nevertheless be possible to tap into processing before particular resonances have been established.

Differential frequency of features accounts for our results on allophonic specificity in flaps. In American English, the flap is more frequent in intervocalic contexts than the underlying /t/s and /d/s (see Charles-Luce, 1997; Patterson & Connine, 2001). Consequently, one reason we obtained specificity effects in our easy discrimination task is that the easy discrimination allowed us to tap into the system at a point when only the more frequent features had established a resonance with the input. On the other hand, the difficult discrimination task presumably taps into the recognition process somewhat later, after the features corresponding to the underlying /t/ and /d/ representations have also established a resonance with the input.

We propose that this adaptive resonance framework also accounts for the time course of indexical specificity effects.[1] In particular, we again propose that chunks are composed of acoustic–phonetic features that vary in frequency. Roughly, phonetic or phonological features are abstract, whereas features capturing indexical variability are less so. Typically, more abstract and general features are higher in frequency, whereas less abstract and more specific features are lower in frequency. As a result, abstract features initially resonate with the input, with features representing indexical information only resonating after some delay. To clarify, chunks are composed of both abstract features and features representing indexical information. Furthermore, because these different features establish a resonance with the input at different rates, and the resonance between input and chunk constitutes the percept, the role that these different features play during word recognition depends, at least in part, on the time course of processing.

The design and logic behind the current experiments examining the time course of indexical variability follow from our earlier work examining allophonic variability and specificity effects (McLennan et al., 2003). In each of the present experiments, we used the long-term repetition priming paradigm to investigate indexical specificity. Because of the differences in the nature and frequency of the features associated with allophonic and indexical information, we posit a time course for indexical specificity effects that is distinct from that for allophonic specificity effects. On the basis of our time-course hypothesis, we predicted that there would be attenuated indexical specificity effects when processing is fast, and, as processing unfolds over time there would be more pronounced indexical specificity effects.

There are two fundamental reasons allophonic and indexical specificity effects might follow a different time course. First, the more specific allophonic representations may occur more frequently than the underlying forms (as is the case with flaps in intervocalic context in American English). On the other hand, the more specific indexical representations are generally less frequent than are the more abstract underlying forms. Recall that the feature frequency account predicts that the more frequent representations should resonate with the input first. Hence, specific allophonic representations may resonate with the input before underlying, abstract representations. Second, although allophones are specific relative to underlying abstract phonemic representations, allophonic variation in speech constitutes fairly abstract information, especially compared with indexical variability. Therefore, because allophonic and indexical variability are not equivalent on a number of dimensions, especially feature frequency, we predicted different time courses for allophonic and indexical specificity effects.

Again, because underlying abstract information is more frequent and more predictive than more variable surface indexical information, underlying information should therefore resonate with the input before indexical information does. One potential reason linguistic information is considered more predictive than indexical information is that these two sources of information may map onto qualitatively distinct representations in memory. Whereas information associated with allophonic variability maps onto discrete, idealized, abstract categorical representations or symbols, information associated with indexical variability may map onto more continuous representations (see, e.g., Ryan, Chasaide, & Gobl, 2003). For example, indexical information associated with talker identity may be represented on a continuous scale in perceptual space (e.g., from high to low fundamental frequency values), rather than in categorical format. Therefore, our account not only views the representations associated with linguistic and indexical information as varying quantitatively in terms of frequency but also potentially qualitatively, in terms of distinct types of representations. Finally, it is also feasible that representations capturing linguistic and indexical information are stored in distinct areas of the brain. We discuss this possibility in greater detail in the General Discussion.

---

[1] The proposed hypothesis makes specific, testable, and falsifiable predictions regarding the time course of indexical specificity effects. The resonance framework, although argued to account for the time course effects, is independent of the more direct and falsifiable hypothesis under examination.

In the present research, we directly tested our time-course hypothesis in a series of perceptual experiments designed to investigate the time course of indexical specificity effects in spoken word recognition. Furthermore, despite the previous demonstrations that variability has representational and processing consequences, this topic has received little theoretical attention or modeling efforts in the literature. In fact, no current major processing model of spoken word recognition (e.g., TRACE, McClelland & Elman, 1986; Shortlist, Norris, 1994; PARSYN, Luce, Goldinger, Auer, & Vitevitch, 2000; distributed cohort model [DCM], Gaskell & Marslen-Wilson, 1999) has been able to account for the representational and processing consequences associated with indexical variability (at least not without substantial modification). Consequently, we have recently proposed (McLennan et al., 2003) that a resonance framework (Grossberg & Myers, 2000; Grossberg & Stone, 1986), discussed earlier, most naturally accounts for our work on allophonic specificity.

In all three experiments, we examined the long-term representational consequences of indexical variability. Moreover, we assessed our time-course hypothesis by manipulating the ease of discrimination in a lexical-decision task (Experiments 1 and 2) and the response format in a shadowing task (Experiment 3). Table 1 summarizes the experiments and the ways in which we investigated the time-course hypothesis in each.

In Experiments 1 and 2, we examined time course effects by manipulating the ease of discrimination in a lexical-decision task. Both Experiments 1A and 2A included easy discrimination tasks that used unwordlike nonwords, whereas in Experiments 1B and 2B, we included difficult discrimination tasks that used wordlike nonwords. Manipulating the ease of discrimination in a lexical-decision task should result in different rates of responding, with faster response times in the easy discrimination task than in the difficult discrimination task. Consequently, this manipulation allowed us to examine the time course of indexical specificity effects. In Experiment 1, we built on our previous work by manipulating speaking rate. In Experiment 2, we manipulated talker identity.

In Experiment 3, we examined time course effects by manipulating the response format in a shadowing task. In Experiment 3A, the response format in the shadowing task was the typical speeded response whereby participants were instructed to repeat, or shadow, each stimulus item as quickly as possible. However, in Experiment 3B, the response format was implemented with a delayed-response paradigm in order to maximize the likelihood of observing specificity effects. In this experiment, participants were instructed to wait for a response cue before shadowing each stimulus item.

The present research had the following two major goals: The first goal was to examine the time course of indexical specificity effects. Our hypothesis was that indexical specificity effects take time to develop and influence spoken word recognition processes and thus these effects would be attenuated when processing is fast and would be more robust when processing is slow. The second goal was to examine how the results might be accounted for in a resonance framework. We proposed that chunks are composed of features representing acoustic-phonetic input that vary in frequency. Phonetic or phonological features are abstract, whereas the features that encode indexical variability in speech are less so. Furthermore, more abstract and general features are normally higher in frequency than are less abstract and more specific features. The feature frequency hypothesis states that more frequent features should resonate with the input before less frequent features. As a result, we predicted that abstract features would initially resonate with the input and indexical features would only do so after some delay.

Experiment 1: Speaking Rate

We used the long-term repetition priming paradigm and the lexical-decision task to examine indexical specificity effects associated with speaking rate. Moreover, we used different sets of nonwords in Experiments 1A and 1B. This manipulation was designed to affect the speed with which participants processed the spoken stimuli. In Experiment 1A, the nonwords were unwordlike, which should make the discrimination between words and nonwords relatively easy. Thus, processing of all items in the experiment, including the target stimuli, should be relatively fast. Therefore, our prediction on the basis of our time-course hypothesis was that indexical specificity effects would be attenuated. Alternatively, in Experiment 1B, the nonwords were wordlike, which should make the discrimination between words and nonwords relatively difficult. Thus, processing of all items in the experiment, including the target stimuli, should be relatively slow. Therefore,

Table 1
Experiment Manipulations and Time-Course Predictions

Experiment	Indexical manipulation	Time-course manipulation	Time-course predictions
1			
Speaking rate and ease of discrimination	Rate (fast, slow)	Discrimination (easy vs. difficult)	SE when discrimination is difficult (Experiment 1B) but not easy (Experiment 1A)
2			
Talker identity and ease of discrimination	Talker (same, different)	Discrimination (easy vs. difficult)	SE when discrimination is difficult (Experiment 2B) but not easy (Experiment 2A)
3			
Talker identity and response format	Talker (same, different)	Response format (speeded response vs. delayed response)	SE in delayed response (Experiment 3B) but not in speeded response (Experiment 3A)

Note. SE = specificity effects.

our prediction based on our time-course hypothesis was that larger indexical specificity effects would emerge.

Experiment 1A: Easy Discrimination

Method

Participants. Seventy-two participants were recruited from the University at Buffalo, State University of New York (UB) community. They were paid $5 or 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 (a) 12 casually and 12 carefully produced bisyllabic spoken target words; (b) 12 casually and 12 carefully produced bisyllabic nonwords; and (c) 8 bisyllabic control items. Casually and carefully produced stimuli differed primarily in speed of articulation. Casually articulated stimuli were recorded with a fast rate of articulation; carefully articulated stimuli, on the other hand, were recorded with a slow rate of articulation.[2] See the Appendix for a complete list of the stimuli used in all experiments.

To make word–nonword discrimination easy, the nonwords were unwordlike and were created by using sequences with low phonotactic probability (e.g., *thushthudge*). Phonotactic probability was determined both by positional segment frequency (i.e., how often a particular segment occurs in a position in a word) and biphone frequency (i.e., segment-to-segment co-occurrence probability). Low phonotactic probability nonwords were defined as nonwords with low segment and biphone frequencies.

The mean log frequency of occurrence for the target stimuli was .79 (Kučera & Francis, 1967). The mean duration for the target stimuli with a slow speaking rate was 622 ms. The mean duration for the target stimuli with a fast speaking rate was 461 ms. This difference in duration between target stimuli with slow and fast speaking rates reflects articulation style; no attempt was made to equate the durations for these two types of stimuli.

The stimuli were recorded in a sound-attenuated room by a phonetically sophisticated male speaker of a Midwestern dialect, low-pass filtered at 10 kHz, and digitized at a sampling rate of 20 kHz by using a 16-bit analog-to-digital converter. All words were edited into individual files and stored on computer disk for later playback.

Design. Two blocks of stimuli were presented. The first consisted of the primes and the second the targets. The stimuli with slow and fast speaking rates served as both primes and targets. For both the primes and targets, half of the stimuli were spoken with a fast speaking rate and half were spoken with a slow speaking rate. Primes were matched, mismatched, or unrelated to the targets. Speaking rate of matched primes and targets was identical (e.g., $bacon_{fast}$, $bacon_{fast}$; $bacon_{slow}$, $bacon_{slow}$). Speaking rate of mismatched primes and targets differed (e.g., $bacon_{fast}$, $bacon_{slow}$; $bacon_{slow}$, $bacon_{fast}$). Both the prime and target blocks consisted of 24 stimuli, 12 words and 12 nonwords. The composition of the prime block was as follows: 8 target words, 8 nonwords, and 8 control stimuli (4 of the control stimuli were words, 4 were nonwords). The composition of the target block was as follows: 12 target words and 12 nonwords. In the target block, 8 stimuli matched, 8 mismatched, and 8 were controls. Note that all nonwords and unrelated control stimuli (words and nonwords) were simply fillers. The focus of the experimental manipulations and later statistical analyses is limited to the target words.

Orthogonal combination of the three levels of prime (match, mismatch, and control) and two levels of target (fast, slow) resulted in six conditions, which are shown in Table 2. Across participants, each slow and fast 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 *bacon* in one of the blocks, he or she did not hear another version of that word again in the same block.

Table 2

Experimental Conditions and Examples of Primes and Targets in Experiments 1A and 1B

Condition	Prime	Target
Match		
Slow prime → Slow target	$bacon_{slow}$	$bacon_{slow}$
Fast prime → Fast target	$bacon_{fast}$	$bacon_{fast}$
Mismatch		
Fast prime → Slow target	$bacon_{fast}$	$bacon_{slow}$
Slow prime → Fast target	$bacon_{slow}$	$bacon_{fast}$
Control		
Unrelated prime → Slow target	*jagged*	$bacon_{slow}$
Unrelated prime → Fast target	*jagged*	$bacon_{fast}$

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 as 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 target blocks, the stimuli were presented binaurally over headphones. A Macintosh Centris 650 computer controlled stimulus presentation and recorded participants' times to make lexical decisions. Stimulus presentation within each block was random for each participant.

A given trial proceeded as follows: A light at the top of the response box illuminated to indicate 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 lexical decision as quickly and as accurately as possible. Reaction times (RTs) were measured from the onset of the presentation of the stimulus word to the onset of the participant's button press response. 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.

Results

RTs less than 500 ms or greater than 2,500 ms were excluded from the analyses, resulting in the elimination of six RTs. Moreover, any participant whose overall mean RT fell two standard deviations beyond the grand mean was excluded, resulting in the elimination of 2 participants.

Prime (match, mismatch, control) × Target (slow, fast) participant and item analyses of variance (ANOVAs) were performed on mean RTs for correct responses and percentages correct for the target stimuli. Accuracy was greater than 94% overall and produced no significant effects. Mean percentage correct as a function of prime type for all experiments is reported in Table 3. For all experiments, all effects were significant at the .05 level, unless otherwise indicated. Although responses to nonwords were not the focus of the present study, the overall mean RT and mean percentage correct for the nonword stimuli were 1,206 and 99, re-

---

[2] Although we refer to our manipulation in terms of speaking rate, our stimuli also differ in articulation style, casual and careful. Consequently, we cannot distinguish between speaking rate and articulation style in the present research.

Table 3

Mean Percentage Correct as a Function of Prime Type for Experiments 1A–3B

Experiment and task	Match	Mismatch	Control
1			
Rate and ease of discrimination			
EDLD	97	98	94
HDLD	92	92	88
2			
Talker identity and ease of discrimination			
EDLD	94	96	93
HDLD	96	96	89
3			
Talker identity and response format			
Speeded	97	98	98
Delayed	92	92	89

Note. EDLD = easy discrimination lexical decision; HDLD = hard discrimination lexical decision; speeded = speeded-response shadowing; delayed = delayed-response shadowing.

spectively (*SE*s = 29.6 and .30), indicating that participants were both accurate and relatively fast in responding to the nonwords.

Note that traditional item analyses are not appropriate for the current experiments. First, we carefully selected our stimuli on the basis of many variables known to affect the dependent variables under scrutiny, thus calling into question the suitability of performing traditional ANOVAs with items as random factors (see Raaijmakers, Schrijnemakers, & Gremmen, 1999). Moreover, the design of our experiments included counterbalanced lists, such that each item appeared in every condition. Raaijmakers (2003; see also, Raaijmakers et al., 1999) has argued that conducting separate item analyses in designs that use counterbalanced lists is unfounded. Despite these concerns, we nonetheless report item analyses, more because of convention than because of their appropriateness. The reader should bear in mind these caveats in interpreting the significance levels of all item tests reported for the current studies.[3]

Mean RTs as a function of condition and magnitudes of specificity (MOS) and magnitude of priming (MOP) are shown in Table 4. MOS is the difference in RT between the match and mismatch conditions. MOP is the difference in RT between the match and control conditions.

Fast items were responded to more quickly than were slow items, $F_1(1, 69) = 42.48$, $MSE = 13,489.18$; $F_2(1, 11) = 10.38$, $MSE = 8,112.33$, presumably because of the differences in duration. The main effect of prime was also significant, $F_1(2, 138) = 8.13$, $MSE = 18,499.91$; $F_2(2, 22) = 5.50$, $MSE = 5,197.57$. Prime and target did not interact (both $Fs < 1$).

Planned comparisons based on the main effect of prime revealed significant differences between the match and control conditions, $F_1(1, 138) = 11.72$; $F_2(1, 22) = 7.89$, and between the mismatch and control conditions, $F_1(1, 138) = 12.64$; $F_2(1, 22) = 8.58$. There was no difference between the match and mismatch conditions (both $Fs < 1$).

Discussion

Both matched and mismatched primes produced significant facilitative effects on lexical decision responses. Furthermore,

matched primes facilitated responses to targets as much as mismatched primes. These results are consistent with our time-course predictions stated earlier: When processing was fast (as a result of the easy discrimination allowed by the unwordlike nonwords), we did not obtain indexical specificity effects of rate (i.e., no effects related to speaking rate differences).

Experiment 1B was conducted to test the hypothesis that when processing is slowed down by the use of wordlike nonwords, we should obtain specificity effects with the same target stimuli that we used in Experiment 1A.

Experiment 1B: Difficult Discrimination

Method

Participants. Seventy-two participants were recruited from the UB community. They were paid $5 or 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 (a) the same 12 casually and 12 carefully produced bisyllabic spoken target words used in Experiment 1A, (b) 12 new casually and 12 new carefully produced bisyllabic nonwords, and (c) 8 bisyllabic control items.

To make word–nonword discrimination more difficult, the nonwords were wordlike and were created by changing the endings of real words so that they became nonwords (e.g., *bacon, bacov*).[4]

The stimuli were recorded in a sound-attenuated room by a phonetically sophisticated male speaker of a Midwestern dialect, low-pass filtered at 10 kHz, and digitized at a sampling rate of 20 kHz by using a 16-bit analog-to-digital converter. All words were edited into individual files and stored on computer disk for later playback.

Design and procedure. The design and procedure were identical to those used earlier in Experiment 1A.

Results

Three RTs and 3 participants were excluded from the analyses on the basis of the same criteria as in Experiment 1A. Accuracy was greater than 87% overall and produced no significant outcomes.

Mean RTs as a function of condition and MOS and MOP are shown in Table 4. The overall mean RT and mean percentage correct for the nonword stimuli were 1,208 and 93, respectively (*SE*s = 24.4 and .9).

Fast items were again responded to more quickly than were slow items, $F_1(1, 68) = 48.27$, $MSE = 20,772.59$; $F_2(1, 11) = 11.13$,

---

[3] Furthermore, the nature of the long-term repetition priming paradigm necessarily limits the number of items that can be used in a within-participants manipulation. On the basis of our experience with this paradigm, we have found that increasing the number of items tends to reduce the likelihood of obtaining long-term repetition priming effects. Presumably, the high degree of sublexical overlap among the items makes it difficult to obtain differential effects of priming. Consequently, the low number of items used also reduces the statistical power of our tests.

[4] We conducted a control experiment in order to ensure that the phonological overlap between our words and nonwords (e.g., *bacon, bacov*) did not contaminate our results. We selected a set of wordlike nonwords without phonological overlap for use in the control experiment (e.g., *bacon, albug*). We observed the same pattern of results as that reported for Experiment 1B.

Table 4
Reaction Times, Standard Errors, and Magnitudes of Specificity and Priming for Experiments 1 and 2

Experiment and task	Match		Mismatch		Control		MOS	MOP
	RT	SE	RT	SE	RT	SE		
1								
Rate and ease of discrimination								
EDLD	855	13	853	15	911	16	2	−56[a,b]
HDLD	901	16	942	14	985	16	−41[a]	−84[a,b]
2								
Talker identity and ease of discrimination								
EDLD	755	12	763	12	800	15	−8	−45[a,b]
HDLD	773	11	808	13	837	14	−35[a,b]	−64[a,b]

Note. RT = reaction time; MOS = magnitude of specificity (match–mismatch); MOP = magnitude of priming (match–control); EDLD = easy discrimination lexical decision; HDLD = hard discrimination lexical decision.
[a] Effects were significant by participants.
[b] Effects were significant by items.

$MSE = 11,071.12$. Once again, we obtained a main effect of prime, $F_1(2, 136) = 11.07$, $MSE = 22,455.20$; $F_2(2, 22) = 10.02$, $MSE = 4,794.73$. Prime and target did not interact (both $Fs < 1$).

Planned comparisons based on the main effect of prime revealed significant differences between the match and control conditions, $F_1(1, 136) = 22.14$; $F_2(1, 22) = 18.60$, between the mismatch and control conditions, $F_1(1, 136) = 5.94$; $F_2(1, 22) = 10.21$, and, crucially, between the match and mismatch (by participants) conditions, $F_1(1, 136) = 5.15$; $F_2(1, 22) = 1.25$, $p = .28$.

Discussion

Both matched and mismatched primes produced significant facilitative effects on lexical decision responses in Experiment 1B. However, the significant difference between the matched and mismatched primes demonstrates that matched primes served as more effective primes than did mismatched primes. These findings are consistent with our time-course hypothesis: When processing was relatively slow (because of the more difficult discrimination caused by the presence of wordlike nonwords), indexical specificity effects of speaking rate emerged.

In Experiment 1A, when processing was fast, matched and mismatched primes were equally effective. On the other hand, in Experiment 1B, when processing was slow, primes matched in speaking rate were more effective than were mismatched primes. Although the current results do not reveal absolute specificity (the mismatched condition was faster than the control condition), faster responses in the matched condition than in the mismatched condition are consistent with the predicted pattern for specificity.

To preview, at the end of Experiment 2 we directly compare the MOS in the easy and difficult discrimination tasks by conducting a combined analysis of Experiments 1 and 2. We carried out this overall analysis to determine (a) that our time-course hypothesis generalizes across two different sources of indexical variability (i.e., speaking rate and talker identity) and (b) that there is no interaction between discrimination difficulty and source of indexical variability.

Experiment 2: Talker Identity

Experiments 2A and 2B are identical to Experiments 1A and 1B, with one exception: Instead of manipulating speaking rate, we manipulated talker identity. Our goal in this experiment was to provide a further test of our time-course hypothesis by examining a different indexical property in speech. Although we manipulated different indexical properties in Experiments 1 and 2, we proposed that both indexical properties operate on the same time course. In other words, we hypothesized that indexical specificity effects associated with talker identity should also take time to influence processing. Assuming that our hypothesis is correct, we predicted the same pattern of results as Experiments 1A and 1B: When processing is fast, specificity effects associated with talker identity should be absent or attenuated. In contrast, when processing is slow, clear effects of specificity should emerge.

Experiment 2A: Easy Discrimination

We again used the long-term repetition priming paradigm and the lexical-decision task. As in Experiment 1A, the nonwords were unwordlike. Therefore, we predicted that specificity effects would be attenuated.

Method

Participants. Seventy-two participants were recruited from the UB community. They were paid $5 or 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 (a) 12 monosyllabic spoken target words, (b) 12 monosyllabic spoken nonwords, and (c) 8 monosyllabic spoken control items. To make word–nonword discrimination easy, the nonwords were unwordlike and low in phonotactic probability (e.g., *thaz*).

The stimuli were recorded in a sound-attenuated room by both a male (PL) and a female (TA) talker, low-pass filtered at 10 kHz, and digitized at a sampling rate of 20 kHz using a 16-bit analog-to-digital converter. All words were edited into individual files and stored on computer disk for later playback.

The mean log frequency of occurrence for the target stimuli was 1.54 (Kučera & Francis, 1967). The mean durations for the target stimuli produced by talkers PL and TA were 409 ms and 337 ms, respectively. This difference in duration reflects the difference in the talkers' natural

speaking rates; no attempt was made to equate the durations of the stimuli produced by talkers PL and TA.[5]

Design. The same design used in Experiments 1A and 1B was used in this experiment. Two blocks of stimuli were presented. The first consisted of the primes and the second the targets. The stimuli spoken by talkers PL and TA served as both primes and targets. For both the primes and targets, half of the stimuli were spoken by talker PL and half were spoken by talker TA. Primes were either matched, mismatched, or unrelated to the targets. Matched primes and targets were identical (e.g., bee_{PL}, bee_{PL}; bee_{TA}, bee_{TA}). Mismatched primes and targets differed on the talker dimension (e.g., bee_{PL}, bee_{TA}; bee_{TA}, bee_{PL}). Both the prime and target blocks consisted of 24 stimuli, 12 words and 12 nonwords. The composition of the prime block was as follows: 8 target words, 8 nonwords, and 8 control stimuli (4 of the control stimuli were words, 4 were nonwords). The composition of the target block was as follows: 12 target words and 12 nonwords. In the target block, 8 stimuli matched, 8 mismatched, and 8 were controls. As in Experiment 1, all nonwords and unrelated control stimuli (words and nonwords) were simply fillers. The focus of the experimental manipulations and later statistical analyses is limited to the target words.

Orthogonal combination of the three levels of prime (match, mismatch, and control) and two levels of target (talker PL, talker TA) resulted in six conditions, shown in Table 5. Across participants, each item was assigned to 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 *leg* in one of the blocks, he or she did not hear that word again in the same block.

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

Results

RTs less than 500 ms or greater than 2,500 ms were excluded from the analyses, resulting in the elimination of 14 RTs. Moreover, any participant whose overall mean RT fell two standard deviations beyond the grand mean was excluded, resulting in the elimination of 4 participants. Accuracy to target stimuli was greater than 92% overall and produced no significant outcomes.

Prime (match, mismatch, control) × Target (talker PL, talker TA) participant and item ANOVAs were performed on mean RTs for correct responses and percentages correct for the target stimuli.

Mean RTs as a function of condition and MOS and MOP are shown in Table 4. The overall mean RT and mean percentage

Table 5
Experimental Conditions and Examples of Primes and Targets in Experiment 2

Condition	Prime	Target
Match		
PL prime → PL target	bee_{PL}	bee_{PL}
TA prime → TA target	bee_{TA}	bee_{TA}
Mismatch		
PL prime → TA target	bee_{PL}	bee_{TA}
TA prime → PL target	bee_{TA}	bee_{PL}
Control		
Unrelated prime → PL target	*hat*	bee_{PL}
Unrelated prime → TA target	*hat*	bee_{TA}

Note. PL = male talker; TA = female talker.

correct for the nonword stimuli were 1,006 and 83, respectively (SEs = 20.1 and 1.0).

Target stimuli produced by talker TA were responded to more quickly than were target stimuli produced by talker PL. The main effect of target was significant, $F_1(1, 67) = 20.72$, $MSE = 15,826.40$; $F_2(1, 11) = 14.61$, $MSE = 3,381.48$. The main effect of prime was also significant by participants and marginally significant by items, $F_1(2, 134) = 4.65$, $MSE = 16,574.51$; $F_2(2, 22) = 2.97$, $MSE = 6,152.32$, $p = .07$. Prime and target did not interact (both $Fs < 1$).

Planned comparisons based on the main effect of prime revealed significant differences between the match and control conditions, $F_1(1, 134) = 8.12$; $F_2(1, 22) = 5.77$, and between the mismatch and control conditions, $F_1(1, 134) = 5.60$; $F_2(1, 22) = 2.45$, $p = .13$. There was no difference between the match and mismatch conditions (all $Fs < 1$).

Discussion

As in Experiment 1A, both matched and mismatched primes produced significant facilitative effects on lexical decision responses. Furthermore, matched primes facilitated responses to targets as much as did mismatched primes. These results are consistent with our time-course predictions stated earlier: When processing is fast, we should fail to observe indexical specificity effects of talker identity.

Experiment 2B was conducted to test the hypothesis that when processing is slowed, we should obtain specificity effects associated with talker identity with the same target stimuli used in Experiment 2A.

Experiment 2B: Difficult Discrimination

This experiment is essentially a replication of Experiment 2A, with one important exception. Instead of using unwordlike nonwords, we used wordlike nonwords in the present experiment. This change was expected to slow participants' processing of target stimuli, as occurred in Experiment 1B. Therefore, we now predicted that we would obtain indexical specificity effects for talker variability.

Method

Participants. Seventy-two participants were recruited from the UB community. They were paid $5 or 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 (a) the same 12 monosyllabic spoken target words used in Experiment 2A, (b) 12 new spoken monosyllabic nonwords, and (c) 8 monosyllabic spoken control items.

To make the word–nonword discrimination task more difficult, the nonwords were created as they were in Experiment 1B, by changing the endings of real words so that they became nonwords (e.g., *book*, *boop*).

The stimuli were recorded in a sound-attenuated room by both a male and female talker, low-pass filtered at 10 kHz, and digitized at a sampling

---

[5] A trained speech scientist judged the stimuli produced by both talkers to be at a normal speaking rate.

rate of 20 kHz by using a 16-bit analog-to-digital converter. All words were edited into individual files and stored on computer disk for later playback.

Design. The design was identical to that used in Experiment 2A.

Procedure. The procedure was identical to that in Experiments 1A, 1B, and 2A.

Results

Six RTs and 5 participants were excluded from the analyses on the basis of the same criteria used in Experiment 2A. Mean RTs as a function of condition and MOS and MOP are shown in Table 4. The overall mean RT and mean percentage correct for the nonword stimuli were 1,016 and 78, respectively (SEs = 20.5 and 1.5).

Accuracy to target stimuli was greater than 88% overall. We observed a significant main effect of prime, $F_1(2, 132) = 7.95$, $MSE = 253.56$; $F_2(2, 22) = 3.49$, $MSE = 97.35$. This effect was driven entirely by lower accuracy in the control condition.

Again, target stimuli produced by talker TA were responded to more quickly than target stimuli produced by talker PL. This main effect of target was significant by participants, $F_1(1, 66) = 5.05$, $MSE = 20,674.30$; $F_2(1, 11) = 1.21$, $MSE = 11,998.51$, $p = .30$. Once again, we obtained a main effect of prime, $F_1(2, 132) = 6.63$, $MSE = 20,664.54$; $F_2(2, 22) = 6.53$, $MSE = 4,025.03$. Prime and target did not interact, $F_1(2, 132) = 1.58$, $MSE = 15,837.96$, $p = .22$; $F_2(2, 22) = 1.60$, $MSE = 2,982.92$, $p = .22$.

Planned comparisons based on the main effect of prime revealed significant differences between the match and control conditions, $F_1(1, 132) = 13.22$; $F_2(1, 22) = 12.94$, and, crucially, between the match and mismatch conditions, $F_1(1, 132) = 3.96$; $F_2(1, 22) = 4.38$. The difference between the mismatch and control conditions was not significant, $F_1(1, 132) = 2.71$, $p = .10$; $F_2(1, 22) = 2.26$, $p = .15$.

Discussion

Both matched and mismatched primes produced facilitative effects on lexical decision responses. However, the difference between the matched and mismatched conditions demonstrates that matched primes served as more effective primes than did mismatched primes. The pattern is consistent with our time-course predictions: When processing was relatively slow, indexical specificity effects of talker identity emerged. In contrast, when processing was fast, we did not obtain indexical specificity effects of talker identity. Consequently, Experiments 2A and 2B provide further support for our general hypothesis that time course is an important factor in determining the role that indexical variability plays in spoken word recognition.

In Experiment 2A, when processing was fast, matched and mismatched primes were equally effective, whereas in Experiment 2B, when processing was slow, primes matched on talker identity were more effective than mismatched primes.

Combined Analysis of Experiments 1 and 2

We conducted a final overall analysis comparing the MOS in the easy (Experiments 1A and 2A) and difficult (Experiments 1B and 2B) discrimination lexical decision experiments. Recall that MOS is the difference in RT between the match and mismatch conditions. We performed the combined analysis, rather than analyzing the MOS by task difficulty interaction in Experiments 1 and 2 separately, to increase the statistical power of our test.

First, RTs to make lexical decisions in the difficult discrimination experiments were significantly longer than in the easy discrimination experiments, $F_1(1, 272) = 12.55$, $MSE = 86,679.58$; $F_2(1, 46) = 4.21$, $MSE = 56,969.44$, indicating that across Experiments 1 and 2, the manipulation of ease of discrimination was indeed successful. Furthermore, we observed a main effect of discrimination difficulty that was significant by participants, $F_1(1, 270) = 4.31$, $MSE = 39,431.39$; $F_2(1, 44) = 1.58$, $MSE = 5,323.40$, $p = .15$, indicating more specificity in the difficult than in the easy discrimination experiments, as predicted. Finally, the interaction between discrimination difficulty and source of indexical variability was not significant (both $Fs < 1$).

This combined analysis of Experiments 1 and 2 indicates that our time-course hypothesis generalizes across two different sources of indexical variability, namely speaking rate and talker identity. Furthermore, the lack of an interaction between discrimination difficulty and source of indexical variability demonstrates that the MOS in the easy and difficult discrimination experiments was the same for both speaking rate and talker identity sources of indexical variability.

Experiment 3: Talker Identity

In Experiments 3A and 3B, we further examined the representational consequences of indexical variability by using the single-word shadowing task. By extending the scope of our investigation beyond the lexical decision paradigm, we attempted to ensure that our results were not peculiar to, or dependent on, this particular task. However, the change in task necessitated a change in the manipulation we used to examine the time course of processing. Because there is no explicit lexical discrimination component in the shadowing task that can be directly speeded up or slowed down, as in the previous experiments, we needed a method for manipulating processing speed other than varying the wordlikeness of nonwords. Thus, we used a delayed shadowing manipulation (see Balota & Chumbley, 1985). Delayed shadowing should provide participants more time to process the stimuli, which should increase the likelihood of obtaining specificity effects.

Experiment 3A: Speeded Shadowing

We examined the degree of indexical specificity associated with talker identity by using the typical speeded-response shadowing task. Because the speeded-response shadowing task could potentially tap processing relatively early, our prediction was that we would not obtain specificity effects.

Method

Participants. Forty-eight participants were recruited from the UB community. They were paid $5 or 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 (a) 12 bisyllabic spoken target words, (b) 12 bisyllabic spoken filler words, and (c) 8 bisyllabic control words. The stimuli were recorded in a sound-attenuated room by both a male talker (PL) and a female talker (TA), low-pass filtered at 10 kHz, and

Table 6
Reaction Times, Standard Errors, and Magnitudes of Specificity and Priming for Experiment 3

	Match		Mismatch		Control			
Task	RT	SE	RT	SE	RT	SE	MOS	MOP
Speeded-response shadowing	814	14	808	11	855	12	6[b]	−41[a,b]
Delayed-response shadowing	350	11	378	16	388	11	−28[a,b]	−38[a,b]

Note. RT = reaction time; MOS = magnitude of specificity (match–mismatch); MOP = magnitude of priming (match–control).
[a] Effects were significant by participants.
[b] Effects were significant by items.

digitized at a sampling rate of 20 kHz by using a 16-bit analog-to-digital converter. All words were edited into individual files and stored on computer disk for later playback.

The mean log frequency of occurrence for the target stimuli was .46 (Kučera & Francis, 1967). The mean durations for target stimuli produced by talkers TA and PL were 569 ms and 560 ms, respectively. This difference was not significant, $t(11) = .283$, $p = .78$.

Design. The design was identical to that used in Experiment 2.

Procedure. The procedure was identical to the procedures used in Experiments 1A, 1B, 2A, and 2B, with one exception. Rather than a lexical-decision task, we used a single-word speeded-response shadowing task in this experiment.

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 single-word shadowing task in which they attempted to repeat (or shadow) the stimulus word as quickly and as accurately as possible. In both the prime and target blocks, the stimuli were presented binaurally over headphones. The headphones had an attached microphone that was placed approximately 1 in. from the participant's lips. A Centris 650 computer controlled stimulus presentation and recorded shadowing times. Stimulus presentation within each block was random for each participant.

A given trial proceeded as follows: A light at the top of the response box was illuminated to indicate the beginning of the trial. The participant was then presented with a stimulus word binaurally over the headphones. The participant was instructed to shadow the stimulus word as quickly and as accurately as possible. Reaction times (RTs) were measured from the onset of the presentation of the stimulus word to the onset of the participant's shadowing response. 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.

Results

RTs less than 200 ms or greater than 2,000 ms were excluded from the analyses, resulting in the elimination of five RTs.[6] Moreover, any participant whose overall mean RT fell two standard deviations beyond the grand mean was excluded, resulting in the elimination of 1 participant.

Prime (match, mismatch, control) × Target (talker PL, talker TA) participant and item ANOVAs were performed on mean RTs for correct responses and percentages correct for the target stimuli. Accuracy was greater than 97% overall and produced no significant effects.

Mean RTs as a function of condition and MOS and MOP are shown in Table 6. We observed a significant main effect of prime type, $F_1(2, 92) = 10.90$, $MSE = 5,602.22$; $F_2(2, 22) = 32.48$,

$MSE = 63.34$. Prime and target did not interact, $F_1(2, 92) = 1.28$, $MSE = 8,359.10$ and $F_2 < 1$.

Planned comparisons based on the main effect of prime revealed a significant difference between the match and control conditions, $F_1(1, 92) = 13.80$; $F_2(1, 22) = 63.10$, and between the mismatch and control conditions, $F_1(1, 92) = 18.56$; $F_2(1, 22) = 26.53$. Crucially, the difference between the match and mismatch conditions was not significant by participants, $F_1 < 1$ and $F_2(1, 22) = 7.8$.

Discussion

Both matched and mismatched primes produced significant facilitative effects on shadowing times. Furthermore, matched primes facilitated target shadowing as much as mismatched primes.

Contrary to our previous work (see McLennan, 2003), we failed to obtain specificity effects in this experiment. We suspected that participants might have been responding too quickly (i.e., before indexical information could play a role). Consequently, in an attempt to allow additional processing time, we reran the experiment with a delayed-response procedure. Our expectation was that this additional processing time would increase the likelihood that we would obtain specificity effects.[7]

Experiment 3B: Delayed Shadowing

This experiment is a replication of Experiment 3A, with one important exception: We used a delayed-response paradigm in

---

[6] Different upper and lower cutoffs were used for the two types of tasks (shadowing and lexical decision), consistent with McLennan et al. (2003).

[7] Nonetheless, the question arises as to why we obtained indexical specificity effects in a speeded-response shadowing task in our previous study (McLennan et al., 2003) but not in the current study. Several factors, including stimulus characteristics, could potentially affect processing speed and thus the likelihood of obtaining specificity effects (for a related discussion, see Luce et al., 2003). The stimuli in these two studies differed on the dimension of concreteness. In the current study, all of our stimuli were concrete nouns; in our previous study this was not the case. Previous research has demonstrated that voice effects are more difficult to obtain with concrete words than with abstract words (Sheffert, 1998). Moreover, concrete words are recognized more quickly than abstract words (Tyler, Voice, & Moss, 2000). Consequently, this difference in concreteness could explain why we were able to obtain specificity effects with the speeded-response shadowing task in our previous study.

order to maximize the likelihood of observing specificity effects. Unlike the typical speeded-response shadowing tasks (such as that used in Experiment 3A), the delayed-response task cues the participant when to respond. As a result, participants have the opportunity to spend additional time processing and rehearsing each stimulus. We hypothesized that this additional time would allow indexical information to influence processing, thus maximizing the likelihood of our obtaining specificity effects for talker identity.

Method

Participants. Forty-eight participants were recruited from the UB community. They were paid $5 or 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 the same (a) 12 bisyllabic spoken target words, (b) 12 bisyllabic spoken filler words, and (c) 8 bisyllabic control words used in Experiment 3A.

The stimuli were recorded in a sound-attenuated room by both a male talker (PL) and a female talker (TA), low-pass filtered at 10 kHz, and digitized at a sampling rate of 20 kHz by using a 16-bit analog-to-digital converter. All words were edited into individual files and stored on computer disk.

Design. The design was identical to that used in Experiment 2.

Procedure. The procedure was identical to that used in Experiment 3A, with one exception. Rather than a speeded-response shadowing task, we used a delayed-response shadowing task in this experiment. Participants were instructed to delay their vocal shadowing response until they saw the cue to respond (five large, red asterisks) in the center of the computer monitor, which appeared 150 ms after the offset of the spoken stimulus. A practice session ensured that participants understood and followed this instruction. Any response initiated before the response cue was treated as an error. RTs were measured from the onset of the presentation of the response cue to the onset of the participant's shadowing response. Psy-Scope software (Cohen, MacWhinney, Flatt, & Provost, 1993) running on a Centris 650 computer was used for stimulus presentation and response collection.

Results

RTs less than 100 ms or greater than 1,500 ms[8] were excluded from the analyses, resulting in the elimination of 36 RTs (less than 4% of the total RTs). No participant was excluded from the analyses.

Prime (match, mismatch, control) × Target (talker PL, talker TA) participant and item ANOVAs were performed on mean RTs for correct responses[9] and percentages correct for the target stimuli. Accuracy was greater than 90% overall and produced no significant effects. Mean RTs as a function of condition and MOS and MOP are shown in Table 6.

Absolute RTs were faster in Experiment 3B, in which we used the delayed-response shadowing task, than they were in Experiment 3A, in which we used the speed-response shadowing task, because of the difference in how RTs were measured in these experiments. RTs in Experiment 3A were measured from the onset of the presentation of the stimulus word, and RTs in Experiment 3B that were due to the nature of the delayed-response task, were measured from the onset of the response cue.

We observed a significant main effect of prime type, $F_1(2, 94) = 4.91$, $MSE = 7,635.79$; $F_2(2, 22) = 3.87$, $MSE = 3,145.97$. Prime and target did not interact (both $Fs < 1$).

Planned comparisons based on the main effect of prime revealed a significant difference between the match and control conditions, $F_1(1, 94) = 9.14$; $F_2(1, 22) = 4.81$, and, crucially, between the match and mismatch conditions, $F_1(1, 94) = 4.96$; $F_2(1, 22) = 6.66$. The difference between the mismatch and control conditions was not significant (both $Fs < 1$).

Although the only difference between Experiments 3A and 3B was the response format (speeded or delayed shadowing), the priming patterns clearly differed in these two experiments. More specifically, in Experiment 3B, but not in Experiment 3A, the difference between the match and mismatch prime conditions was significant (in fact, this difference was in the opposite direction in Experiment 3A, with faster RTs in the mismatch than in the match prime condition). Furthermore, we also observed a significant Prime Type × Experiment interaction, $F_1(2, 186) = 3.07$, $MSE = 6,629.94$; $F_2(2, 44) = 2.34$, $MSE = 1,604.66$, $p = .11$, providing statistical confirmation of the differential patterns of priming in the two experiments.

Discussion

Matched but not mismatched primes produced significant facilitative effects on shadowing times. Moreover, matched primes were more effective than were mismatched primes. These results are consistent with the predictions of our time-course hypothesis: When processing is probed relatively late (as a result of the nature of the delayed-response paradigm), indexical specificity effects of talker identity emerge.

We have examined the role of task variables on indexical specificity effects by manipulating ease of discrimination (easy vs. difficult) and response format (speeded vs. delayed). Our ease of discrimination manipulations (Experiments 1 and 2) demonstrated that indexical specificity effects are attenuated when processing is relatively fast and they are more pronounced when processing is slowed. Our response format manipulation (Experiment 3) demonstrated that when participants' responses are delayed in time (as in Experiment 3B), we obtain indexical specificity effects. Therefore, on the basis of the results of the current experiments, we know that task variables related to the amount of time spent generating a response or processing the stimuli and the point at which one taps the process for a response support the argument that indexical variability influences perception of spoken words relatively late.

General Discussion

The central hypothesis under investigation was that indexical information in speech takes time to influence spoken word pro-

---

[8] Different cutoffs were used here than were used in Experiment 3A because of the delayed-response paradigm.

[9] Two separate timers were used. The second timer began at the onset of the response cue and recorded the RTs of interest. The first timer began at the onset of the auditory stimulus and ended at the onset of the response cue. If the first timer recorded a value, then the participant did not, as instructed, wait for the response cue to begin his/her shadowing response. Trials in which the first timer recorded a value were counted as errors and thus are not part of the RT analyses.

cessing. Consequently, we predicted that we would observe in-dexical specificity effects when processing is slow but not when processing is fast.

In Experiments 1 and 2, we manipulated the ease of discrimination between words and nonwords in a lexical-decision task. We hypothesized that in the easy discrimination tasks (Experiments 1A and 2A), processing would be fast and we would not obtain indexical specificity effects. On the other hand, we hypothesized that in the difficult discrimination tasks (Experiments 1B and 2B), processing would be slow and indexical specificity would emerge (with the same targets used in the easy discrimination versions of the experiments). Our predictions were confirmed: The difference between the match and mismatch conditions was significant only in the difficult discrimination tasks. Moreover, subsequent analyses confirmed that we obtained greater indexical specificity effects in the difficult discrimination tasks than in the easy discrimination tasks.

Experiment 3 was designed to probe the processing system at different times by using different response formats, namely speeded shadowing and delayed shadowing. We hypothesized that because we would be probing the processing system relatively late in delayed shadowing (Experiment 3B), we should obtain indexical specificity effects. On the other hand, because we would be probing the processing system earlier in speeded shadowing (Experiment 3A), we hypothesized that we would only obtain attenuated effects of indexical specificity. Once again, our predictions were confirmed: The difference between the match and mismatch conditions was significant only in the delayed-response shadowing task. Overall, our data support the hypothesis that indexical specificity effects arise late in processing.

Two points of clarification are in order: First, as we have previously discussed (McLennan, et al., 2003), depth of form-based processing and time course of processing are typically coextensive: Deeper form-based processing likely occurs in difficult discrimination lexical decision experiments (and possibly in delayed shadowing), in which processing is also relatively slow. At this time, it is difficult to discriminate between a pure time-course account and one based on depth of processing. Indeed, adjudicating between these alternatives may ultimately prove impossible.

Second, in the current experiments, the tasks and response formats were the same in both the prime and target blocks, making it potentially difficult to discriminate between effects arising during prime as opposed to target processing. In our previous research on allophonic variability, we directly investigated the consequences of crossing tasks (shadowing and lexical decision) in the prime and target blocks (i.e., by using shadowing for the prime block and lexical decision for the target block and vice versa). Our findings from this previous work support two important conclusions: First, although it is possible for responses to both primes and targets to be initiated before all resonances have fully developed, processing continues (probably obligatorily) until all resonances are established, as is evidenced by the finding that primes were equally effective regardless of the speed or depth with which they were processed. Second, we demonstrated that the resonances themselves might serve as the vehicles of priming. That is, we demonstrated that if responses in the prime block are based on fully developed resonances, these resonances might develop more

quickly and easily during the target block. Consequently, resonances between relatively low frequent features and the input may be in evidence in the target block even in tasks that typically do not allow sufficient time for the development of such resonances. In the case of indexical specificity effects, we expect that if the difficult discrimination lexical-decision task were performed during the prime block, indexical specificity effects would be obtained even with the easy discrimination lexical-decision task (or speeded shadowing) during the test block.

Taken together, the results of the present experiments have important implications for current theories and models of spoken word recognition, illustrating the representational and processing challenges posed by effects of indexical variation in speech (see also Luce & McLennan, in press). In particular, the current results provide evidence that early during perceptual processing, more abstract or underlying features dominate, whereas during later stages of processing, features corresponding to more specific, detailed surface information dominate. Furthermore, feature frequency appears to be one of the conditions determining which type of representation (abstract or specific) is likely to dominate at various points during processing. In our earlier work on allophonic variability, features corresponding to the more frequent flap were in evidence early during perceptual processing, whereas features corresponding to the less frequent underlying /t/s and /d/s affected only later processing. Similarly, in the current investigation, features corresponding to the more frequent abstract information were in evidence early during perceptual processing and features corresponding to the less frequent and more specific indexical information only affected later processing.

As mentioned in the introduction, no current major processing model of spoken word recognition to date (e.g., DCM; Gaskell & Marslen-Wilson, 1999; PARSYN; Luce et al., 2000; TRACE; McClelland & Elman, 1986; Shortlist; Norris, 1994) is able to account for both the representational and processing consequences associated with indexical variability in speech (at least not without substantial modification). One shortcoming of these traditional models is that they do not represent indexical information. However, work on the representation of indexical variability (including the current investigation; see also, Church & Schacter, 1994) demonstrates that indexical information is encoded and retained in memory and has long-term processing consequences for spoken word recognition.

Although in their present forms, current models of spoken word recognition are unable to account for our complete set of results, nothing in their architectures prohibits the necessary modifications: Models could add representations designed to capture indexical variability (e.g., representations associated with various rates of speech). However, even with the appropriate representations, they would need to account for our time-course effects. One possibility is to manipulate the weights of their representations, such that the more frequent abstract representations have higher weights than the less frequent representations corresponding to specific indexical information. Nonetheless, because our results also demonstrate that indexical information does not affect immediate online processing of spoken words, current models may be correct in their emphasis on abstract linguistic representations.

Overall, however, we believe that the adaptive resonance framework (Grossberg & Myers, 2000; Grossberg & Stone, 1986) most

naturally accounts for our time course effects of coexisting form-based lexical representations that are both abstract and specific (see also Luce et al., 1999; McLennan et al., 2003). Again, in this framework, acoustic–phonetic input activates a set of features that resonate with the input. The resonance between input and chunk is a composite representation consisting of features of the input and chunk that are mutually consistent and a feedback loop that selects and enhances these shared features. Furthermore, the resonance constitutes the percept and mediates priming and specificity effects. According to the feature frequency hypothesis, higher frequency features establish a resonance with the input before lower frequency features. Recall that features capturing abstract linguistic information (e.g., phonemes) are typically higher in frequency than are features capturing more highly variable surface information, including indexical information (e.g., talker identity). Therefore, abstract features should typically resonate with the input before more specific features. However, when features corresponding to specific information are higher in frequency than the relevant abstract features, the specific features should resonate with the input before the abstract features (see McLennan, in press).

This adaptive resonance framework accounts for our present set of results. Participants in the easy discrimination experiments made their decisions more quickly than did participants in the difficult discrimination experiments. Consequently, in the easy discrimination experiments, there was only enough time for high-frequency features to establish a resonance with the input. Low-frequency features representing the more specific relevant indexical information (i.e., speaking rate in Experiment 1 and talker identity in Experiment 2) had not yet established a resonance with the input. A similar explanation can be offered for the results obtained in Experiment 3: More processing time was available in the delayed-response shadowing task. This additional processing time increases the likelihood that features corresponding to indexical information (in this case, talker identity) will have had time to establish a resonance with the input. As a result, we observed greater indexical specificity effects in the delayed-response shadowing task.

The current data also support our earlier argument that allophonic and indexical specificity effects manifest themselves at different points during processing. The measure of specificity (i.e., RTs in the matching condition minus RTs in the mismatching condition) in the easy discrimination lexical decision in the present Experiment 1A investigating indexical variability was 8 ms. The measure of specificity in the easy discrimination lexical decision in the previous Experiment 3 investigating allophonic variability (McLennan et al., 2003) was −115 ms. In other words, indexical specificity effects are absent, but allophonic specificity effects are robust when processing is fast.

As we mentioned earlier, we believe that information associated with linguistic and indexical variability may potentially map onto qualitatively distinct types of representations. Whereas information associated with linguistic variability maps onto discrete, abstract, idealized segmental representations, information associated with indexical variability may map onto more continuous representations. Furthermore, these two types of representations may be stored and processed in distinct areas of the brain, with each area processing the input at a different rate. For example, McClelland,

McNaughton, and O'Reilly (1995) have argued that the hippocampus typically stores specific information, and the neocortex stores more abstract information (see also, Tulving & Schacter, 1990). Furthermore, these authors have argued that the function of the hippocampus is to allow for the retention of the specific aspects of particular episodes while avoiding interference with the structured knowledge held in the neocortex. McClelland et al. have also claimed that information initially stored in the hippocampus will eventually be consolidated into the neocortex and that highly detailed idiosyncratic aspects of the input will require more time to consolidate. In other words, information that is common across a variety of episodes is consolidated into the neocortex more quickly than information that is more variable. Qualitative differences, such as being represented in distinct brain regions, could also emerge if different aspects of the chunks consistently establish resonances with the input at different rates.

Marsolek (1999; see also, Squire, 1992) has also argued in favor of a dissociable subsystems theory, claiming that the recognition of abstract and specific aspects of input are carried out by distinct neural subsystems. According to Marsolek, the subsystem responsible for the recognition of abstract information operates more efficiently in the left cerebral hemisphere, whereas the subsystem responsible for the recognition of more specific information operates more efficiently in the right cerebral hemisphere. Evidence for qualitatively distinct patterns of priming in the two hemispheres supports this claim. For example, Marsolek, Kosslyn, and Squire (1992) obtained greater specificity effects when they presented objects to the right hemisphere, providing some support for the dissociable subsystems theory. Similarly, other researchers have obtained evidence consistent with the idea that linguistic information is represented and processed in the left hemisphere and paralinguistic information in the right hemisphere (see, e.g., Kreiman & Van Lancker, 1988; Van Lancker, Kreiman, & Cummings, 1989).

Regardless of whether abstract and specific information is stored in distinct brain regions, or hemispheres, our time-course hypothesis predicts that linguistic and indexical information is processed at different rates, which in turn affects the degree to which stimulus variability plays a role in normal spoken word recognition.

Although variability is a longstanding issue in the domain of language research, the present findings provide some new and potentially important insights into both the representations and processes necessary to account for this phenomenon. The next generation of theories and models must deal with both the representational and processing consequences of these findings on indexical variability in speech perception if they are to be considered psychologically viable models of spoken word recognition.

References

Abercrombie, D. (1967). *Elements of general phonetics.* Chicago: Aldine.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24,* 89–106.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics, 61, 2,* 206–219.

Charles-Luce, J. (1997). Cognitive factors involved in preserving a phonemic contrast. *Language & Speech, 40, 3,* 229–248.

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, 3,* 521–533.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instrumentation, and Computers, 25,* 257–271.

Connine, C. M. (in press). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review.*

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

Creelman, C. D. (1957). The case of unknown talker. *Journal of the Acoustical Society of America, 29,* 655.

Fourcin, A. J. (1968). Speech source interference. *IEEE Transactions on Audio Electroacoustics, 16,* 65–67.

Franks, J. J., Bilbrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer appropriate processing (TAP) and repetition priming. *Memory & Cognition, 28,* 1140–1151.

Fujimoto, M. (2003). The effect of voice variation on the nature of the representations of speech and recognition memory: Evidence from form-based priming. *University at Buffalo Working Papers on Language and Perception, 2,* 87–163.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science, 23, 4,* 439–462.

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, 5,* 1166–1183.

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, 1,* 152–162.

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Vol. 1. Speech perception* (pp. 187–294). New York: Academic Press.

Grossberg, S., & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review 107, 4,* 735–767.

Grossberg, S., & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review 93,* 46–74.

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 1143–1154.

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

Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research, 40,* 1395–1405.

Kreiman, J., & Van Lancker, D. (1988). Hemispheric specialization for voice recognition: Evidence from dichotic listening. *Brain and Language, 34,* 246–252.

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

Ladefoged, P. (2000). *A course in phonetics* (5th ed.). San Diego, CA: Harcourt.

Luce, P. A., Charles-Luce, J., & McLennan, C. T. (1999, August). Representational specificity of lexical form in the production and perception of spoken words. Paper presented at the proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, CA.

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. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition, 26, 4,* 708–715.

Luce, P. A., & McLennan, C. T. (in press). Spoken word recognition: The challenge of variation. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception.* Malden, MA: Blackwell.

Luce, P. A., McLennan, C. T., & Charles-Luce, J. (2003). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory* (pp. 197–214). England: Oxford University Press.

Marsolek, C. J. (1999). Dissociable neural subsystems underlie abstract and specific object recognition. *Psychological Science, 10,* 111–118.

Marsolek, C. J., Kosslyn, S. M., & Squire, L. R. (1992). Form-specific visual priming in the right cerebral hemisphere. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 3,* 492–508.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Sommers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 4,* 676–684.

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

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

McLennan, C. T. (2003). The time course of indexical specificity effects in spoken word recognition. *Dissertation Abstracts International, 64* (08), 4078.

McLennan, C. T. (in press). *The time course of variability effects in the perception of spoken language: Changes across the lifespan.*

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 4,* 539–553.

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., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.), *Speech perception, production, and linguistic structure* (pp. 113–134). Tokyo: Ohmsha Publishing.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60, 3,* 355–376.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1990). Episodic encoding of voice and recognition memory for spoken words. *Research on Speech Perception Progress Report, 16* (pp. 391–412). Bloomington: Indiana University.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 2,* 309–328.

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica, 58,* 254–275.

Peters, R. W. (1955). The relative intelligibility of single-voice and multiple-voice messages under various conditions of noise. *Joint Project*

Report, 56 (pp. 1–9). Pensacola, FL: U. S. Naval School of Aviation Medicine.

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: The Acoustical Society of Japan.

Pisoni, D. B. (1992). Talker normalization in speech perception. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds*.), Speech perception, production, and linguistic structure.* Tokyo: Ohmsha Press.

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

Raaijmakers, J. G. W. (2003). A further look at the "Language-as-Fixed-Effect Fallacy". *Canadian Journal of Experimental Psychology, 57,* 141–151.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How do deal with "The Language-as-Fixed-Effect Fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426.

Remez, R. E., Fellowes, J., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 651–666.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 33, 3,* 441–452.

Ryan, C., Chasaide, A. N., & Gobl, C. (2003). Voice quality variation and the perception of affect: Continuous or categorical? In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the International Congress of Phonetic Sciences* (pp. 2409–2412). Barcelona, Spain: Futuregraphic.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 915–930.

Sheffert, S. M. (1998). Contributions of surface and conceptual information to recognition memory. *Perception & Psychophysics, 60,* 1141–1152.

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging, 11,* 333–341.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99,* 195–231.

Tulving, E., & Schacter, D. L. (1990, January). Priming and human memory systems. *Science, 247,* 301–306.

Tyler, L. K., Voice, J. K., & Moss, H. E. (2000). The interaction of meaning and sound in spoken word recognition. *Psychonomic Bulletin & Review, 7,* 320–326.

Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology, 11,* 665–674.

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America, 60,* 198–212.

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older adults. *Psychology and Aging, 15,* 88–99.

Appendix

Target, Nonword, Filler, and Control Stimuli Used in
Experiments 1A–3B

		Experiment		
1A	1B	2A	2B	3A and 3B

Targets

1A	1B	2A	2B	3A and 3B
bacon	*bacon*	*bear*	*bear*	*accordion*
baggage	*baggage*	*bee*	*bee*	*alligator*
boycott	*boycott*	*book*	*book*	*donkey*
bucket	*bucket*	*bowl*	*bowl*	*flower*
bygone	*bygone*	*car*	*car*	*lobster*
bypass	*bypass*	*cat*	*cat*	*mountain*
cabbage	*cabbage*	*deer*	*deer*	*mushroom*
cabin	*cabin*	*fly*	*fly*	*needle*
caucus	*caucus*	*key*	*key*	*pliers*
circuit	*circuit*	*leg*	*leg*	*scissors*
circus	*circus*	*nail*	*nail*	*thimble*
coping	*coping*	*nut*	*nut*	*whistle*

Nonwords / **Fillers**

1A	1B	2A	2B	3A and 3B
jʌʃðʌtʃ	bekəv	ʤʊg	bæp	*barrel*
θʌsjʌʤ	bægənt	ʤɝg	biθ	*banana*
tʃʌʃθʌʤ	boɪkɔf	θaʃ	bʊp	*finger*
jʌtʃtʃʌʤ	bʌkəm	ðɔŋ	bog	*giraffe*
θʌtʃʃʌʤ	baɪgəps	jev	kaf	*gorilla*
ðaɪðʃaɪð	baɪpæb	ʤaʊm	kæg	*lemon*
ʃaɪðtʃaɪð	kæbəv	gɝp	dut	*leopard*
gaɪðdaɪz	kɔkəg	θaʒ	floɪ	*onion*
ðaɪbdʒaɪz	sɜkɚ	zeð	kaɪ	*orange*
ðaɪvʃaɪb	kopɚg	fʊp	lɛb	*pumpkin*
tʃaɪzwaɪð	ʤægʌp	naʊʤ	næv	*rabbit*
jiʃgiʃ	wɛpʌks	voɪz	nʌp	*turtle*

Controls

1A	1B	2A	2B	3A and 3B
luggage	*luggage*	*goat*	*goat*	*ostrich*
jagged	*jagged*	*hand*	*hand*	*lettuce*
nugget	*nugget*	*hat*	*hat*	*monkey*
ribbon	*ribbon*	*heart*	*heart*	*raccoon*
ʤeθʃeʒ	kɪkbæp	tʊz	gog	*rooster*
θeʤʃeð	mædkʌs	θaʊʤ	hæb	*ruler*
jɜzjɜθ	bamʃɛz	θaʊð	hæθ	*spider*
ʃɜθjɜg	kaɪfæp	taʊtʃ	haɪb	*squirrel*