

2003

Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming.

Paul A. Luce

University at Buffalo, The State University of New York, luce@buffalo.edu

Conor T. McLennan

Cleveland State University, c.mclennan@csuohio.edu

Jan Chance-Luce

university at buffalo

Follow this and additional works at: https://engagedscholarship.csuohio.edu/clpsych_facpub

 Part of the [Cognition and Perception Commons](#)

How does access to this work benefit you? Let us know!

Recommended Citation

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 Bowers, J. & Marsolek, C. (Eds.), *Rethinking implicit memory*, pp. 197-214. Oxford: Oxford University Press.

This Contribution to Books is brought to you for free and open access by the Psychology Department at EngagedScholarship@CSU. It has been accepted for inclusion in Psychology Faculty Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

ABSTRACTNESS AND SPECIFICITY IN SPOKEN WORD RECOGNITION: INDEXICAL AND ALLOPHONIC VARIABILITY IN LONG-TERM REPETITION PRIMING

PAUL A. LUCE, CONOR M^CLENNAN, AND JAN CHARLES-LUCE

Introduction

Speech scientists frequently make a subtle, and often unspoken, distinction between two closely allied areas of research: speech perception and spoken word recognition (Juszyk and Luce, 2002). Research on *speech perception* is typically concerned with consonants, vowels, and syllables, and focuses on such fundamental problems as *invariance* (are there acoustic events that consistently signal a given speech sound?) and *variability* (how does the listener cope with the diverse acoustic manifestations of phonetic segments?). Research on *spoken word recognition* focuses primarily on issues concerning the *activation of* and *discrimination among* form-based (i.e., sound-based) lexical representations. For the sake of tractability, researchers examining spoken word recognition have often side-stepped the fundamental problems that have been the provenance of speech perception. Indeed, most current models of spoken word recognition have little to say about the still unresolved problems of invariance and variability that have occupied speech perception research for over 60 years (Luce and Pisoni, 1998; M^Clelland and Elman, 1986; Norris, 1994).

The distinction between research on speech perception and spoken word recognition, while never neat, has recently become less clear, owing in part to the realization that the longstanding problem of variability has deep and important implications for models of lexical representation and process. Because of variability in the sex, age, and dialect of talkers, as well as differences in speaking rates, emotional states, and a host of other variables, listeners must cope with the myriad ways in which spoken words are uttered. Understanding how the perceptual system maintains perceptual constancy in the face of this pervasive variability may lead to fundamental insights in the nature and organization

of the form-based lexicon, as well as the processes by which memory representations of spoken words are contacted (Martin *et al.*, 1989; Mullenix *et al.*, 1989).

Recent advances in our understanding of variability and its implications for lexical representation and process have come in large part from research using the long-term repetition priming paradigm (Church and Schacter, 1994; Goldinger, 1996; Luce and Lyons, 1998; Schacter and Church, 1992; Sheffert, 1998a, 1998b). This paradigm has enabled investigators to examine the degree of specificity and abstractness of form-based representations, which in turn has provided insights into the architecture of the word recognition system. Briefly, investigators have used the phenomenon of long-term, form-based repetition priming to determine the degree to which lexical representations encode the variability inherent in spoken words.

The logic of the repetition paradigm is simple: Processing of a spoken word is typically facilitated when it is repeated exactly (as measured by accuracy, processing time, or both). However, if the first and second presentations (*prime* and *target*, respectively) mismatch on some dimension—for example, one is spoken by a male and the other by a female—the priming effect is often attenuated. We may infer from this reduction in priming that the prime and target activate somewhat different *specific* form-based lexical representations. If, on the other hand, the priming effect is *not* attenuated by any differences between the prime and target, we might conclude that the prime and target activate the same underlying *abstract* representations.

Our intent in this chapter is to explore the nature of form-based representations of spoken words, as revealed by variability (or specificity) effects in long-term repetition priming. We focus on two sources of variability—indexical and allophonic—each of which provides different clues about the specificity and abstractness of sound-based lexical representations. Based on a number of recent findings, we argue for the existence of both specific and abstract codes in spoken word recognition. Moreover, we argue that the data in favor of abstract lexical codes also implicate the activation of intermediate representations in the mapping of the acoustic-phonetic waveform onto lexical items in memory. Finally, we hypothesize that there is a time-course to specificity effects in spoken word recognition, such that immediate processing is dominated by abstract codes, whereas specific information takes time to percolate through the system and have its effects on perception.

We begin by drawing distinctions between indexical and allophonic variability and between episodic and abstractionist theories of lexical form. As we will argue, evidence for episodic theories comes primarily—although not exclusively—from research on indexical variability, whereas research on allophonic variability suggests the operation of more abstract codes. We conclude by arguing for a mixed representational model in which differential effects of abstract and episodic codes are predictable based on the processing time considerations.

Indexical and allophonic variability

Throughout the course of our discussion, we will distinguish between two sources of variability in spoken word recognition (see Remez *et al.*, 1997). *Indexical variability* refers to

variations in a spoken word that arise from differences among talkers, differences in speaking rates, differences in affective states, and so on (Pisoni, 1997). Typically, the 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 slow rate of speech, or in a happy or sad emotional state has no implication for the fact that the word refers to a device used for communication over distances. The indexical property of speech receiving the most attention in research examining long-term priming has been talker identity (typically male versus female).

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*). Each of these different versions are referred to as allophones of the phoneme /t/.

We will refer to both indexical and allophonic variability as *surface variation*. Surface variation, whether indexical or allophonic, does *not* result in different lexical items. For example, the word *telephone* spoken by a male or female (indexical variability) still refers to a communicative device (or the act of using such a device). Likewise, allophonic variation cannot, by definition, result in different words, although it may produce lexical ambiguities (see below). In contrast to surface variation, *underlying variation* has direct consequences for lexical discrimination. For example, featural changes in the putatively underlying abstract phoneme /t/ in *top* will result in a change in phonemic category, and hence a new lexical item (e.g., *pop*). In short, surface variants are *tokens*; underlying variants are *types*.

Although indexical and allophonic variation are both examples of surface variation, allophonic variation is in large part constrained by the phonological system of a given language. In general, allophonic variation is highly predictable based on purely linguistic considerations, such as syllable position and phonetic context. On the other hand, the effects of indexical variability are constrained by extra-linguistic factors and are generally predictable only by reference to vocal tract size, emotional disposition, and so on. Consider a monolingual English speaker listening to various speakers of Thai. Our English speaker may be able to distinguish among male and female, fast and slow, and even angry and happy Thai speakers. However, it is unlikely that our English speaker would be able to distinguish among allophonic variants of Thai phonemes. Coping with allophonic variation is a uniquely linguistic skill acquired through extensive exposure to a given language.

The observation that allophonic variability is linguistic, whereas indexical variability is not, raises the possibility that the spoken word recognition system may treat these two sources of variability differently. Perhaps the more predictable effects of allophonic variation are represented more generally or abstractly in memory, whereas the less predictable, highly variable indexical properties are represented in a more specific format. Perhaps both abstractionist *and* episodic accounts of form-based lexical representations may be needed to account for how listeners cope with pervasive variability in the speech signal.

Abstractionist and episodic theories of form-based lexical representation

The conventional assumption among theories of spoken word recognition is that lexical items are represented in memory by *abstract* phonological codes that only preserve information relevant for lexical discrimination (e.g., Jackson and Morton, 1984; Luce and Pisoni, 1998; McClelland and Elman, 1986; Norris, 1994). In most current models of word recognition, stimulus variation—both indexical and allophonic—is treated as irrelevant information that is discarded early in the encoding process (see, however, Luce *et al.*, 2000). The extraction of information that is solely relevant for identification is referred to as *normalization*, and it is during the normalization phase that representations that vary in physical detail but fall within a given perceptual category are equated by processes that abstract defining information.

For example, feature-based accounts of spoken word recognition (Klatt, 1989; Pisoni and Luce, 1987; Stevens, 1998) have proposed that speech sounds and words are processed using the elemental features of linguistic description (e.g., [vocalic], [consonantal], [sonorant]). Because spoken words may differ on many physical dimensions not captured by these features, the normalization process is responsible for winnowing the information in the speech signal and extracting only the featural information that is necessary and sufficient for unique identification of a word. This process thereby serves a substantial data reduction function that may ultimately result in considerable economy of process and representation.

Despite the arguments that have been made for abstract lexical representations in memory, recent research (Church and Schacter, 1994; Goldinger, 1996, 1998; Luce and Lyons, 1998; Schacter and Church, 1992; Sheffert, 1998a, 1998b) has suggested that putatively irrelevant surface details of words—such as information specific to a given talker—are preserved in some form in memory. These findings regarding specificity effects, which we review below, have led to the proposal that lexical representations are episodic (Goldinger, 1992, 1998). One episodic account proposes that form-based lexical representations are exemplars that preserve, rather than discard, much of the physical detail of the stimulus. Indeed, according to the most extreme episodic models (Hintzman, 1986), a unique *instance* is stored in memory each time a new token is encountered.

One advantage of episodic models is that they have the potential for solving the longstanding problem of normalization by dispelling the notion that the ultimate goal of the perceptual process is to map acoustic-phonetic information onto *abstract* form-based representations of words in memory. In episodic models, the representational currency of the perceptual encoding process is more-or-less true to the details of the stimulus itself. In an application of this general theoretical approach to spoken word recognition, Goldinger (1992, 1998) has proposed an exemplar-based lexicon in which the individual memory traces themselves may encode both abstract and surface information, with the degree of stimulus specificity depending on attentional factors during encoding.

Note that episodic theories of the lexicon need *not* be exemplar- or instance-based. For example, distributed models are capable of representing specificity without laying down a separate trace for each encounter with a spoken word. Instead, specificity is encoded in

minor changes in convolved feature vectors (Weber and Murdock, 1989) or in connection weights among nodes in the network that makes up the distributed representation (Gaskell and Marslen-Wilson, 1997). Furthermore, episodic models may behave *as if* they contain abstract representations, via emergent generalizations over episodes (Hintzman, 1986; Tenpenny, 1995).

There has been considerable recent debate over the distinction between episodic and abstractionist theories (Bowers, 2000; Tenpenny, 1995). The crux of this debate concerns whether abstract representations emerge from averaging over instances, or whether they have their own status and are different in kind from episodic representations (Tenpenny, 1995). Despite the potential importance of resolving this issue, we will remain agnostic regarding the form and origin of abstract form-based lexical representations, although we will argue throughout the remainder of our discussion for some type of mixed representational model (with both abstract and episodic formats).

Indexical variability and representational specificity

Evidence for representational specificity in spoken word recognition comes from a variety of long-term repetition priming experiments that have examined indexical variability. In general, these studies have demonstrated that under certain circumstances, changes in the surface details of stimuli between study (prime) and test (target) attenuate the repetition priming effect. For example, Church and Schacter (1994) and Schacter and Church (1992) observed effects of talker variability in *implicit* tasks such as fragment completion and identification of low-pass filtered stimuli.¹ Subjects were more likely to complete a fragment of a target word if the fragment was repeated in the same voice as the prime. Subjects were also more accurate at identifying low-pass filtered target words that were primed by stimuli sharing similar surface characteristics. However, these researchers failed to observe effects of stimulus specificity in *explicit* tasks. When subjects performed cued recall or old/new recognition of previously presented items, changing surface characteristics between study and test had no detectable effects on performance.

Goldinger (1996) also performed a series of experiments examining the effects of voice on memory for spoken words. In one of his experiments, he presented words in explicit (old/new recognition) and implicit (perceptual identification in noise) tasks with varying delays between study and test. He found significant effects of voice in both the recognition memory and identification tasks, demonstrating that voice effects are not, in fact, restricted to implicit tasks. In another experiment, Goldinger manipulated levels of processing and voice in the study-test implicit-explicit format. His results demonstrated that effects of voice varied with level of processing, such that strongest effects of stimulus

¹ Schacter (1987) defines implicit memory as improved performance on a task following previous experience when the task itself does not depend upon conscious recall or recollection of the previous experience. In contrast, explicit memory is defined as performance on a task requiring conscious recollection of a previous event. We will focus more heavily on results from *implicit* tasks because they tend to be less reflective of conscious, strategy-based processing (although there is probably no pure process; see Bowers, 2000; Jacoby, 1991).

specificity were observed in the shallower processing conditions, especially for recognition memory.

Despite some differences in the circumstances under which specificity effects were observed, the Church and Schacter and Goldinger results demonstrate that implicit long-term repetition priming is attenuated by surface variation between prime and target, strongly suggesting that form-based representations of spoken words are indeed specific. However, the specificity effects observed in these studies may have been amplified by processing difficulties introduced by the use of degraded stimuli (in the form of noise-masked, low-pass filtered, or gated stimuli), which may have encouraged subjects to interrogate memory representations of previously encountered items when they might not have normally done so. Moreover, the use of degraded stimuli in untimed tasks raises the possibility that *speeded* processing of *non-degraded* spoken stimuli may be less susceptible to effects of stimulus variability.

Luce and Lyons (1998) attempted to establish more clearly the degree to which surface variation of *non-degraded* spoken stimuli plays a role in the identification process itself. In particular, they examined the effects of changing voice on stimulus repetition in both auditory lexical decision and recognition memory tasks (Biederman and Cooper, 1992; Cooper *et al.*, 1992). They first presented subjects with a list of stimuli spoken by two talkers in a lexical decision task. They then followed this first block of lexical decision trials with either (1) another block of lexical decision trials (implicit task) or (2) a block of old/new recognition trials (explicit task). The stimuli in the second block of the experiment were either repeated in the same voice, a new voice, or were new items that had not appeared in the first block. They hypothesized that if memory representations of spoken words are abstract, repetition effects for words repeated in the same voice should be equivalent to repetitions of words repeated in a different voice. That is, changing the voice of the speaker for an item between blocks should have no effect on performance in the second block. On the other hand, if the representations of spoken words used in on-line processing are episodic, words repeated in the same voice should show greater advantages in terms of processing speed than words repeated in a different voice.

Luce and Lyons' results demonstrated that repetition priming for spoken words might not always be sensitive to changes in the surface characteristics of the stimuli (see Figure 9.1). In the implicit lexical decision task, response times to repetitions in the same voice were not statistically different from response times to repetitions in the different voice, although overall effects of repetition priming were robust. However, in the explicit recognition memory experiment, Luce and Lyons obtained significant effects of voice: Subjects responded *old* more quickly to words repeated in the same voice than to words repeated in the different voice. Thus, they observed no detectable effects of voice on response time in the implicit task using lexical decision but large effects on old/new recognition time in the explicit task.

The failure to observe specificity effects in lexical decision is actually consistent with certain episodic models (Hintzman, 1986). According to Hintzman's model, each stimulus activates an aggregate of memory traces referred to as an *echo*. Because the echo is comprised of a weighted average of multiple traces, the influence of features that *differ* among the stored exemplars will be minimized, whereas features that are *shared* will be strength-

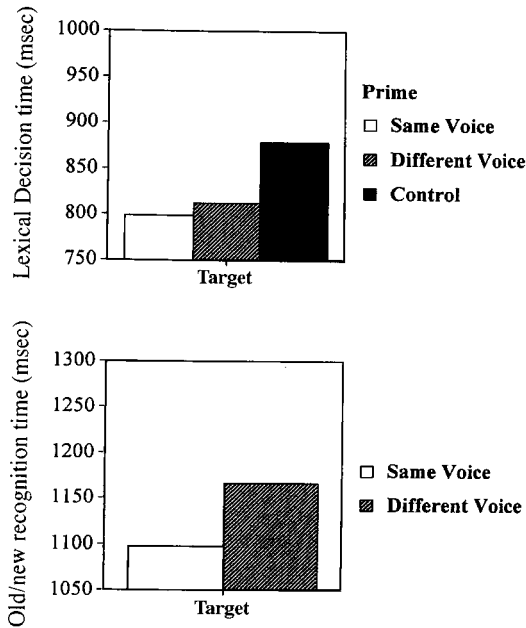


Figure 9.1 Lexical decision times (upper panel) and old/new recognition times (lower panel) to target words as a function of prime type. Target words in the second block of trials were repeated in the same voice, repeated in a different voice, or were not repeated (control). In the implicit task (lexical decision) equivalent effects of long-term repetition priming were obtained for repeated words regardless of voice, whereas in the explicit task (old/new recognition), pronounced effects of specificity were observed. (From Luce and Lyons, 1998.)

ened. With decay of individual features over time, information regarding the central tendency of a category will tend to be more robust. Such a memory system will give the appearance of abstraction, although the representations from which this abstraction emerges are themselves specific (see, however, Bowers and Kouider, this volume). Thus, even within strictly episodic models of memory, both abstract and specific information may coexist (at least functionally).

Time-course of processing and representational specificity

The results of Luce and Lyon's explicit old/new recognition task are consistent with a number of previous demonstrations that voice matters in recognition memory (Goldinger, 1996; Palmeri *et al.*, 1993). However, the failure to observe specificity effects in the implicit priming task does *not* replicate previous work. The reasons for this discrepancy may lie in either stimulus or task differences among the studies, or both. In Church and Schacter's and Goldinger's work, voice effects were typically observed in implicit priming tasks using degraded (filtered or noise-masked) stimuli, which may have amplified the effects of voice by either slowing processing or encouraging activation

of specific previous memory traces to aid in identification. In addition, compared to off-line identification, responses in the lexical decision task may be so rapid as to precede potentially slower acting effects of stimulus specificity in processing (Hintzman and Caulton, 1997; Hintzman and Curran, 1997). We refer to the proposal that the rapidity of responding may mediate the presence or absence of specificity effects as the *time-course hypothesis*.

Goldinger (1996) reports one of the only other spoken word recognition studies that has examined response latencies in an implicit task in which voice was manipulated. Response latencies to classify stimuli in his fastest condition were almost 100 ms longer than the latencies that Luce and Lyons observed in their implicit priming task. Thus, it may be that if subjects are capable of making an identification decision quickly enough, effects of stimulus specificity will be small. Conversely, when responses are slower, as in Luce and Lyons' old/new recognition experiment or in Church and Schacter's and Goldinger's studies, effects of voice emerge. Despite the rapidity of the response, however, Luce and Lyons obtained significant overall effects of repetition priming, suggesting that although specific effects of voice failed to develop strongly in the time window within which responses were made, effects of prior lexical activation were still in evidence.

Thus, task and stimulus differences may have enabled faster responding in the Luce and Lyons study, potentially circumventing effects of specificity. Further support for this hypothesis comes from a study by M'Lennan *et al.* (2002). In contrast to the stimuli used by Luce and Lyons, which were short consonant–vowel–consonant words with a fairly high average frequency, M'Lennan *et al.* examined specificity effects for longer, lower-frequency bisyllabic spoken words. All things considered, Luce and Lyons' short, higher-frequency stimuli should have been recognized faster than M'Lennan *et al.*'s longer, lower-frequency stimuli. If our time-course hypothesis is viable, specificity effects should have emerged for those stimuli requiring longer processing times.

M'Lennan *et al.* presented primes and targets in an implicit long-term repetition-priming paradigm. Items in the first and second blocks either matched or differed on the manner in which they were articulated. (The same talker produced all stimuli.) Half of the stimuli were *hypoarticulated*, or produced in a casual manner, and half were *hyperarticulated*, or carefully produced. Subjects performed a speeded single-word shadowing task in both the first and second blocks.

The average shadowing times to target stimuli in M'Lennan *et al.*'s study was 65 ms longer than the average lexical decision times reported by Luce and Lyons, despite the fact that single-word shadowing typically produces appreciably faster response times than lexical decision (see, for example, Luce and Pisoni, 1998; Vitevitch *et al.*, 1999). Also, even though the differences in surface variation (i.e., casual versus careful articulation) in M'Lennan *et al.*'s study were more subtle than those examined by Luce and Lyons, large effects of specificity were observed: Relative to the appropriate baseline, carefully articulated spoken words only primed other carefully articulated spoken words and casually articulated words only primed other casually articulate words (see Figure 9.2). Thus, when stimuli mismatched on the manner in which they were articulated, no long-term priming was observed.

Taken together, the Luce and Lyons and M'Lennan *et al.* results suggest that specificity effects may take time to develop. If we are able to tap the perceptual process early, by

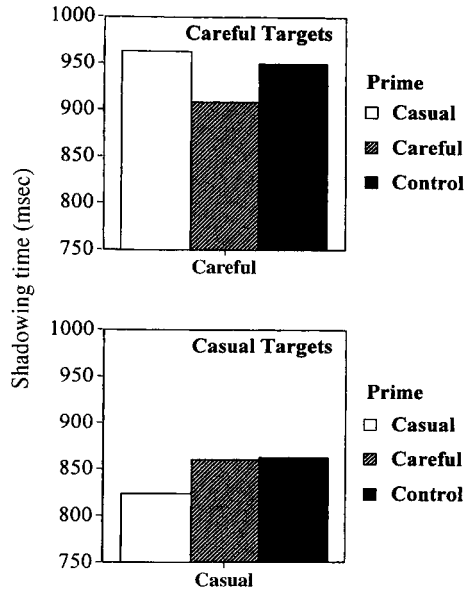


Figure 9.2 Single-word shadowing times to carefully and casually produced target words (upper and lower panels, respectively) as a function of prime type. Target words in the second block of trials were repeated in the same style of articulation (e.g., casual-casual; careful-careful), repeated in a different style of articulation (e.g., casual-careful; careful-casual), or were not repeated (control). Long-term repetition priming effects were obtained only for target words repeated in the same style of articulation. (From McLennan *et al.*, 2001.)

examining processing of short, high-frequency words in a speeded task, no effects of indexical variability are observed. However, specificity effects on long-term priming are clearly in evidence when perception is slowed, even in a speeded perceptual task.²

We speculate that either weakly abstractionist or episodic models may be able to account for the effects of time-course on specificity. Abstractionist theories need only assume that the abstract codes respond or resonate more strongly, and hence more quickly, to stimulus input. This is not an unreasonable or untenable assumption: Abstract representations may code the most frequent features of objects. Thus, the time-course of specificity effects in spoken word recognition may simply be yet another example of frequency effects at a sublexical or featural level of processing (see Vitevitch and Luce, 1999). Episodic theories may likewise be able to account for temporal specificity effects. For example, although Hintzman (1986) makes no explicit claims about time-course, he does note that the most frequent features will be most strongly represented in the echo. It is a small leap to assume

² The absence of priming between mismatching tokens of the same words suggests that specific form-based representations may actually dominate (or, perhaps, inhibit) abstract codes later in processing. We have proposed elsewhere (Gaygen and Luce, 2002; Newman and Luce, 2002) that abstract codes may be immediately inhibited once recognition has been accomplished (in a manner similar to Grossberg *et al.*'s 1997 mismatch reset principle). If correct, this hypothesis predicts that only slower-acting specific codes may be in play during later phases of processing.

that the response of the echo to stimulus input may take time, with more robust or frequent features responding more rapidly.

To summarize, our working hypothesis is that a mixed model best characterizes representation of lexical form in spoken word recognition, be it either episodic or weakly abstractionist. Research on indexical variability has demonstrated that form-based representations preserve much more detail than traditional models of speech perception had previously postulated or current models of spoken word recognition have acknowledged. However, we argue that understanding the nature of abstract and specific form-based lexical representations requires an appreciation of the time-course of processing. In particular, we propose that under optimal processing conditions when identification is easy and fast, abstract codes will dominate spoken word processing, or at least dominate the initial contact of sensory input with memory representations. On the other hand, more specific information may exert its influence on processing over a somewhat longer time frame. We furthermore propose that time-course differences in the processing of abstract and specific form-based information derive directly from the frequency with which the system has encountered features of the input.

Allophonic variability and representational abstractness

Thus far, we have devoted our attention to one kind of variability and its representation, namely indexical variability. However, there is reason to believe that not all variability is equivalent, at least from a processing standpoint. As we discussed, allophonic variability, unlike its indexical counterpart, is tightly constrained by the linguistic system. Consequently, variation at the allophonic level is generally highly predictable. Indeed, some might argue that it is rule-governed (although the rules may be optional, Kentowicz and Kisseberth, 1979). In short, given a constrained and predictable set of allophones, a limited number of highly frequent features in memory may be able to robustly capture and encode this form of systematic phonetic variation.

If our speculation regarding feature frequency is correct, we expect that allophonic variation would be handled under optimal processing conditions by abstract codes. To evaluate this hypothesis, Luce *et al.* (1999; see also M^cLennan *et al.*, 2001) examined the phenomenon of flapping in American English. A flap is an allophone of /t/ and /d/. When /t/ or /d/ occurs between two vowels in casual speech, as in *writer* or *rider*, the speaker often produces a segment that has acoustic properties that do not correspond exactly to either /t/ or a /d/ (see Charles-Luce, 1997; Fox and Terbeek, 1977). The resulting intermediate articulation is referred to as a flap (/V) and has acoustic properties that are ambiguously /t/ or /d/. Under a traditional phonological analysis, flaps are neutralized surface realizations of the underlying phonemic forms /t/ and /d/.³ According to abstractionist linguistic theories of representation (Kentowicz and Kisseberth, 1979), *only* forms containing fully specified /t/ and /d/ are represented in memory. However, according to an episodic view of

³ The phonemes /t/ and /d/ differ on the dimension of voicing, /t/ being voiceless and /d/ being voiced. A flap is said to *neutralize* this voicing distinction by blurring the distinction between the two voicing categories.

lexical representation, flapped tokens of spoken words should have full representational status as well.

In a long-term repetition priming experiment, Luce *et al.* presented bisyllabic words (e.g., *writer*, *rider*) containing both carefully articulated /t/s and /d/s and flaps.⁴ (See Table 9.1 for a description of the experimental conditions.) Subjects shadowed the primes and targets as quickly as possible. Luce *et al.* attempted to determine if flapped primes facilitate processing of carefully articulated targets, and vice versa. Failure to observe equivalent levels of priming between flapped and carefully articulated stimuli would indicate that the carefully articulated and flapped primes and targets do *not* share a common underlying representation. This result would be consistent with episodic models of lexical representation. However, equivalent facilitation of target processing by both flapped and careful primes would indicate the existence of a shared, and presumably abstract, representation.

Luce *et al.*'s results revealed that flapped and carefully articulated words were equally effective primes for flapped and carefully articulated targets (see Figure 9.3). That is, both types of primes produced equivalent levels of facilitation for each type of target. If effects of priming were restricted to veridical, surface representations of primes and targets—as predicted by a strict exemplar, or strongly episodic, model—markedly different patterns of priming between flapped and carefully articulated words should have been observed. In particular, facilitative priming between *identical* items (i.e., flap to flap and careful to careful) should have been greater than facilitative priming between *different* versions of the stimuli (i.e., flap to careful and vice versa). Instead, Luce *et al.* observed statistically equivalent degrees of facilitative priming for flapped and careful primes and targets. It appears that flapped stimuli activate representations containing both /t/ and /d/, which in turn serve to facilitate subsequent processing of flapped and carefully articulated targets.⁵ These results suggest that both

Table 9.1 Stimulus conditions for flap priming experiment (Luce *et al.*, 1999)

Condition	Block 1: Primes	Block 2: Targets
1	Flap (jaɪrə)	Flap (jaɪrə)
2	Flap (jaɪrə)	Careful /t/ (jaɪtə)
3	Flap (jaɪrə)	Careful /d/ (jaɪdə)
4	Careful /t/ and /d/ (jaɪtə)/(jaɪdə)	Flap (jaɪrə)
5	Careful /t/ (jaɪtə)	Careful /t/ (jaɪtə)
6	Careful /d/ (jaɪdə)	Careful /d/ (jaɪdə)
7	None (control)	Flap (jaɪrə)
8	None (control)	Careful /t/ (jaɪtə)
9	None (control)	Careful /d/ (jaɪdə)

⁴ In a screening experiment, Luce *et al.* determined that the flapped stimuli were perceived to be ambiguous.

⁵ An alternative interpretation of these results is that no intermediate *form-based* representations are involved. Instead, the ambiguous flapped stimuli activate both lemmas of the word, which in turn mediate the priming effect. However, no mediating semantic representations were in evidence in the M^cLennan *et al.* (2001) study of unflapped, bisyllabic words that were also carefully and casually articulated, demonstrating that lemma activation is not responsible for the repetition effect in this task.

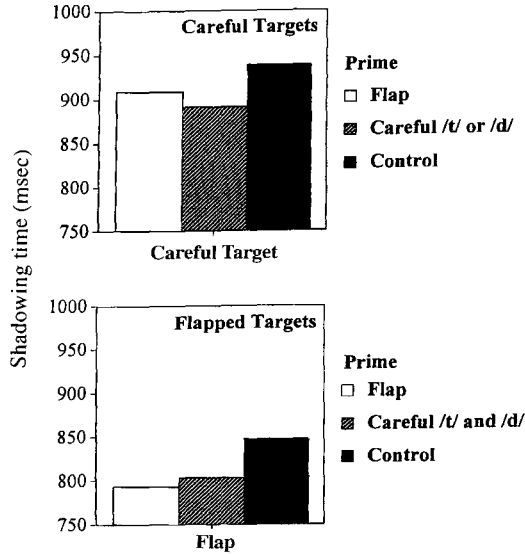


Figure 9.3 Single-word shadowing times to carefully produced and flapped target words (upper and lower panels, respectively) as a function of prime type. Target words in the second block of trials were repeated in the same style of articulation (e.g., flap–flap; careful–careful), repeated in a different style of articulation (e.g., flap–careful; careful–flap), or were not repeated (control). Equivalent effects of long-term repetition priming were obtained for repeated words regardless of the presence or absence of intervocalic flapped consonants. (From Luce *et al.*, 1999.)

underlying and surface forms are activated, at least for the type of allophonic variation examined in this study. To identify a flapped word uniquely, the listener must map the potentially ambiguous word (e.g., *jaɪrə*) onto one of two lexical representations (*writer* or *rider*). Because of this ambiguity, we propose that form-based representations corresponding to clearly delineated voicing categories (i.e., /t/ and /d/) are contacted during the perceptual process. In short, our results suggest that—at least in the perception of flapped items in American English—the surface manifestation of the flap is mapped onto representations in which underlying abstract voicing categories are well specified.

Indexical variability, allophonic variability, and the time-course of processing

We have now discussed a number of pieces of evidence concerning the representation of both indexical and allophonic variability that support the operation of both specific and abstract representations in spoken word recognition. To review:

1. Church and Schacter and Goldinger demonstrated that indexical variability attenuates long-term repetition priming effects.
2. Luce and Lyons demonstrated that indexical variability does *not* attenuate priming effects for short, high-frequency words in a lexical decision task. They hypothesized

that when targets are responded to quickly, specificity effects do not have time to develop (the *time-course hypothesis*).

3. McLennan *et al.* demonstrated that indexical variability attenuates priming effects for longer, lower-frequency words in a speeded shadowing task, providing further evidence for the time-course hypothesis.
4. Luce *et al.* demonstrated that allophonic variability does *not* attenuate priming effects in a speeded shadowing task: Flapped and carefully produced stimuli prime one another equivalently, providing evidence for underlying abstract representations.

Thus, recent research on both indexical and allophonic variability suggests the operation of both abstract and episodic codes in spoken word recognition. In general, indexical variability gives rise to specificity effects whereas allophonic variability results in activation of abstract codes.⁶ Although as we have seen, specificity effects arising from indexical variability may be circumvented under certain circumstances. Nonetheless, questions now arise as to why and under what circumstances one of the two codes dominate perception, and why indexical and allophonic variability tend to produce divergent effects of specificity in long-term priming. We propose two answers. The first concerns ambiguity in allophonic variation and attempts to account for the specific effects observed in the flap priming experiment. The second, more general account of the operation of abstract and episodic codes concerns the related issues of predictability, feature frequency, and the time-course of processing. We consider each in turn.

Ambiguity

One crucial difference between indexical and allophonic variability lies in the fact that the type of allophonic variability that we have examined produces ambiguity, which may require the system to generate multiple interpretations of the input. Intervocalic flapped consonants often produce a word that is lexically ambiguous (e.g., something halfway between *writer* and *rider*). Indexical variability rarely results in this kind of lexical ambiguity (see, however, Peterson and Barney, 1952). Luce *et al.* demonstrated that in dealing with this ambiguity, the system appears to activate the alternative underlying *form-based* representations corresponding to the two possible interpretations of the stimulus. Thus, lexical ambiguity may force the activation of abstract codes to enable the recognition system to identify a given lexical item uniquely.

⁶ Remez *et al.* (1997) dispute the notion that indexical properties alone are responsible for talker differences. Indeed, they provide compelling evidence that allophonic (or idiolectal) information also conveys information about the identity of a talker. Although we have adopted the more traditional view that indexical properties are primarily responsible for talker identification, the possibility that differences among talkers may be signaled at the allophonic level does not undermine our hypothesis that certain aspects of allophonic variation, as well as lexical discrimination, may be handled by abstract codes.

Predictability, frequency, and time-course of processing

The more general predictor of the differential roles of specific and abstract codes appears to involve the speed with which the two types of codes respond to the input, which may itself be intimately tied to feature frequency and predictability. As previously discussed, indexical variability may be far less predictable and constrained than allophonic variability. Flapping in American English is restricted to the consonants /t/ and /d/ and occurs (almost) exclusively intervocally. Although flapping of /t/ and /d/ is optional (albeit common in casual speech), the circumstances under which this type of phonetic neutralization occurs are highly predictable. The phonological system of American English specifies when and under what circumstances allophonic variation can occur. Aside from fairly general biomechanical constraints, indexical variability is typically not so predictable.

The systematicity and predictability of allophonic variability suggests that the recognition system should have robust and abstract representations of allophones. If we equate abstractness with generalization over consistently and frequently occurring feature patterns in the input, allophonic variation is a prime candidate for abstract representational coding. If we further assume that abstract codes represent the most frequent patterns of features encountered by the system, abstract codes should be the first to respond to the input, with specificity percolating through the system over time.

In short, according to our proposal, abstract codes in memory should respond first to sensory input, with specificity effects in evidence shortly thereafter. Thus, when listeners are able to respond quickly, specificity effects may be small (Luce and Lyons, 1998). However, when processing is slowed, because of stimulus or task factors, specificity effects should emerge.

Our time-course hypothesis also makes a specific prediction regarding the effects of flapping: Fast responses—as in single-word shadowing of high-frequency words—should produce attenuated specificity effects. On the other hand, specificity effects should be more strongly in evidence in situations associated with slowed processing times. To further test this time-course hypothesis, McLennan *et al.* examined long-term repetition priming in an auditory lexical decision task for the same flapped and unflapped stimuli used by Luce *et al.* As previously discussed, lexical decision typically produces appreciably longer reaction times than shadowing. If our time-course hypothesis is correct and specificity effects emerge *after* the response of the abstract allophonic codes, a task such as lexical decision should produce effects of indexical variability for the same stimuli that failed to produce specificity effects in shadowing. This is in fact the case. In lexical decision, which produced reaction times approximately 120 ms longer than single-word shadowing, effects of specificity were clearly in evidence, such that priming effects were attenuated when the stimuli mismatched on the dimension of careful-casual articulation.⁷ In short, we propose that abstract codes, by virtue of their higher frequency, tend to respond or resonate first to stimulus input. Episodic codes emerge later and may be in evidence when processing is delayed due to task demands or stimulus characteristics.

⁷ Of particular interest was finding that the specificity effects were almost entirely carried by those stimuli and subjects producing the longest reaction times.

Conclusion: a mixed model of representation of lexical form

The pattern of results reviewed in this chapter support a mixed model of form-based representations in spoken word recognition. Moreover, the available evidence suggests some fairly precise circumstances under which abstract and specific codes will dominate spoken word processing. In particular, if the process of mapping acoustic-phonetic input onto form-based representations is fast and unimpeded, abstract codes should initially dominate processing. More specific, idiosyncratic information—represented by the indexical properties of speech—may exert its influence later in the course of perception. We have also proposed that the distinction between abstract and specific representational codes should be coextensive with the distinction between highly frequent, systematic, and predictable feature patterns (allophonic variability) and less frequent, less predictable information (indexical variability).

The available evidence on the nature of form-based lexical representations in spoken word recognition fails to distinguish between weakly abstract and episodic models. Indeed, it is doubtful that any set of findings will enable this subtle dispute to be reconciled. Moreover, if episodic theories prevail, the evidence thus far does little to help us distinguish among episodic accounts based on exemplars or distributed representations. Although both sets of theories may account for specificity effects in spoken word recognition, they also may behave *as if* abstract codes are operative in perception.

Despite these as yet unanswered questions, it is clear that the next generation of spoken word recognition models must incorporate mechanisms for accounting for the growing body of research demonstrating representational specificity. Although existing models of spoken word identification (Luce *et al.*, 2000; McClelland and Elman, 1986; Norris, 1994) may be correct in their focus on immediate processing of spoken words as abstract entities, they are ultimately inadequate as representational theories of lexical form. Moreover, they may fail to explain how the recognition system behaves in less than ideal circumstances in which perception is impeded by limitations on data or process. In short, recent work on long-term repetition priming of spoken words has provided important new insights into the nature of form-based representations, insights that may ultimately usher in a new generation of models of spoken word recognition.

Acknowledgments

This work was supported (in part) by research grant number R01 DC 0265801 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

References

- Biederman, I. and Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 122–33.

- Bowers, J. S.** (2000). In defense of abstractionist theories of repetition priming and word identification. *Psychonomic Bulletin and Review*, 7 (1), 83–99.
- Bowers, J. S. and Kouider, S.** (In press). Developing theories of priming with an eye on function. In *Rethinking implicit memory* (ed. J. Bowers and C. Marsolek). Oxford: Oxford University Press.
- Charles-Luce, J.** (1997). Cognitive factors involved in preserving a phonemic contrast. *Language and Speech*, 40, 229–48.
- Church, B. A. and 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), 521–33.
- Cooper, L. A., Schacter, D. L., Ballesteros, S., and Moore, C.** (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 43–57.
- Fox, R. A. and Terbeek, D.** (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics*, 5, 27–34.
- Gaskell, M. G. and Marslen-Wilson, W. D.** (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12, 613–56.
- Gaygen, D. and Luce, P. A.** (2002). Troughs and bursts: Probabilistic phonotactics and lexical activation in the segmentation of spoken words in fluent speech. In C. T. McLennan, P. A. Luce, G. Maunder, and J. Charles-Luce (eds.), *University at Buffalo Working Papers on Language and Perception*, Vol. 1, 496–549.
- Goldinger, S. D.** (1992). *Words and voices: Implicit and explicit memory for spoken words* (Technical Report 7). Bloomington, IN: Indiana University Speech Research Laboratory, Department of Psychology.
- Goldinger, S. D.** (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–83.
- Goldinger, S. D.** (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105 (2), 251–79.
- Grossberg, S., Boardman, I. and Cohen, M.** (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 481–503.
- Hintzman, D. L.** (1986). ‘Schema abstraction’ in a multiple-trace memory model. *Psychological Review*, 93 (4), 411–28.
- Hintzman, D. L. and Caulton, D. A.** (1997). Recognition memory and modality judgments: A comparison of retrieval dynamics. *Journal of Memory and Language*, 37, 1–23.
- Hintzman, D. L. and Curran, T.** (1997). Comparing retrieval dynamics in recognition memory and lexical decision. *Journal of Experimental Psychology: General*, 126, 228–47.
- Jackson, A. and Morton, J.** (1984). Facilitation of auditory recognition. *Memory and Cognition*, 12, 568–74.
- Jacoby, L. L.** (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Experimental Psychology: General*, 30, 306–40.
- Jusczyk, P. W. and Luce, P. A.** (2002). Speech perception. In *Stevens’ handbook of experimental psychology* (ed. S. Yantis and H. E. Pashler), 3rd edn, Vol. 1, pp. 493–536. New York: John Wiley and Sons.
- Kentowicz, M. and Kisseberth, C.** (1979). *Generative phonology*. New York: Academic Press.
- Klatt, D. H.** (1989). Review of selected models of speech perception. In *Lexical representation and process* (ed. W. D. Marslen-Wilson), pp. 201–62. Cambridge, MA: MIT Press.
- Ladefoged, P.** (2000). *A course in phonetics*, 5th edn. San Diego: Harcourt, Brace, and Jovanovich.

- Luce, P. A. and Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory and Cognition*, **26**, 708–15.
- Luce, P. A. and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, **19**, 1–36.
- Luce, P. A., Charles-Luce, J., and M^cLennan, C. (1999). *Representational specificity of lexical form in the production and perception of spoken words*. Paper presented at the Proceedings of the 1999 International Congress of Phonetic Sciences.
- Luce, P. A., Goldinger, S. D., Auer, E. T. and Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, **62**, 615–25.
- Martin, C. S., Mullenix, J. W., Pisoni, D. B. and Summers, W. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 676–84.
- M^cClelland, J. L. and Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1–86.
- M^cLennan, C., Luce, P. A. and Charles-Luce, J. (2002). Representation of lexical form in spoken word recognition. *Submitted*.
- Mullenix, J. W., Pisoni, D. B. and Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, **85**, 365–78.
- Newman, R. N. and Luce, P. A. (2002). Radical activation in spoken word recognition. *Submitted*.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189–234.
- Palmeri, T. J., Goldinger, S. D. and Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 309–28.
- Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24**, 175–84.
- Pisoni, D. B. (1997). Some thoughts on ‘normalization’ in speech perception. In *Talker variability in speech processing* (ed. K. Johnson and J. W. Mullenix), pp. 9–32. San Diego, CA: Academic Press.
- Pisoni, D. B. and Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, **25**, 21–52.
- Remez, R. E., Fellowes, J. M. and Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experiment Psychology: Human Perception and Performance*, **23**, 651–66.
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **13**, 501–18.
- Schacter, D. L. and Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 915–30.
- Sheffert, S. M. (1998a). Contributions of surface and conceptual information on spoken word and voice recognition. *Perception and Psychophysics*, **60**, 1141–52.
- Sheffert, S. M. (1998b). Voice-specificity effects on auditory word priming. *Memory and Cognition*, **26** (3), 591–8.
- Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.
- Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review*, **2**, 339–63.
- Vitevitch, M. S. and Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, **40**, 374–408.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B. and Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, **68**, 306–11.

Weber, E. U. and Murdock, B. B. (1989). Priming in a distributed memory system: Implications for models of implicit memory. In *Implicit memory: Theoretical issues* (ed. S. Lewandowsky, J. C. Dunn, and K. Kirsner), pp. 87–98. Hillsdale, NJ: Lawrence Erlbaum Associates.