Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Memory and Language 52 (2005) 322–338
Journal of Memory and Language
www.elsevier.com/locate/jml

Perception and representation of regular variation: The case of final /t/☆

Meghan Sumner, Arthur G. Samuel*

Department of Psychology, SUNY Stony Brook, Stony Brook, NY 11794-2500, USA

Received 18 June 2004; revision received 12 November 2004
Available online 2 February 2005

Abstract

Spoken words exhibit considerable variation from their hypothesized canonical forms. Much of the variation is regular, occurring often in language. The present work examines the immediate and long-term processing consequences for rule-governed final-/t/ variation in English. Two semantic priming experiments demonstrate that variation does not hinder short-term semantic processing, as long as variation is not arbitrary. Two long-term priming experiments with different tasks show that form processing over time is not as lenient as immediate semantic processing: strong priming is found only for the canonical, unchanged form of /t/. Our results suggest that surface information is used in immediate processing and exemplar representations for regular variants are not stored in long-term memory.

Keywords: Variation; Representation; Speech perception; Spoken word recognition

Listeners are confronted by a remarkably variable signal when they understand spoken language. Variation in the speech signal may be caused by many factors, such as speaker variation, speech rate, speech style, or phonotactic variation. A central issue in the perception of spoken language is this substantial variation found in the speech signal, and more specifically, the ways in which this variation is accommodated. Research in this area has been divided between studies of phonologically regular variation, such as assimilation (e.g., Coenen, Zwitserlood, & Bölte, 2001; Gaskell & Marslen-Wilson, 1996; Gow, 2001, 2002), and investigations of arbitrary variation caused by mispronunciation (e.g., Connine, Blasko, & Titone, 1993; Frauenfelder, Scholten, & Content, 2001) or by splicing together segments with mismatching formant information due to cross-splicing (Marslen-Wilson & Warren, 1994). For both types of variation, a critical theoretical question is how listeners cope with such variation.

The current study is concerned with two major issues regarding variation: (1) What effect (if any) does regular variation have on immediate word recognition? and (2) How do listeners represent regular variants in the long term? Understanding variation requires specification of both the recognition process and the longer-term representation of variable forms. The first issue focuses on whether regular phonetic variation has particular costs (or benefits) for word recognition, and the extent to

☆ This material is based upon work supported by the National Science Foundation under Grant No. 0325188 and by NIMH under Grants No. F32MH68204-01A and R0151663. We thank Carol Fowler, Stephen Goldinger, and two anonymous reviewers for their invaluable comments and suggestions. We also thank Donna Kat for help with this project.
* Corresponding author. Fax: +1 632 632 7876.
E-mail addresses: meghan.sumner@gmx.net (M. Sumner), asamuel@ms.cc.sunysb.edu (A.G. Samuel).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2004.11.004

which a surface form may vary from a base form without disrupting processing. The second question is concerned with the type of information about a word that is stored. For example, do we have a single abstract representation for variable words, or do we store each variant of a word independently? Research in this area has concentrated mainly on characteristics such as talker, speech rate, and amplitude (Bradlow, Nygaard, & Pisoni, 1999; Church & Schacter, 1994; Goldinger, 1996, 1998; Luce & Lyons, 1998; Nygaard, Sommers, & Pisoni, 1995; Schacter, 1987; Schacter & Church, 1992). Our research extends the issue of representation to regular variation. By examining both immediate and long-term effects of variation, we can tease apart the recognition *process* and the representations that participate in this process. This enables us to see how the representations affect the recognition process.

There are two types of competing models of representation in the literature. One includes abstract representations in which perception is generally assumed to involve the decoding of particular items into basic, or canonical representations. The second type consists of episodic theories of representation which suggest that representations are fine-grained and that subtle details about surface forms are stored. One commonly used paradigm to examine surface details involves having participants do an initial task (e.g., lexical decision, shadowing, old–new recognition, etc.) on a block of items. Participants then receive a second block of items in which test items are repeated. In this paradigm, items from the first block that are repeated in the second block are generally recognized more quickly than non-repeated items. Goldinger (1996) used this approach to examine the effect of hearing the same speaker versus different speakers across blocks, and manipulated both the delay in repetition and the initial level of processing. In a recognition memory task in which listeners identified words in the clear at study and were later given a surprise recognition test, he found that voice details are stored (but not as long as in implicit memory) and that the effects of voice are stronger with shallow processing. In an implicit memory task in which listeners identified words in noise for both study and test, effects of voice were strong for a week. These results show that details about voice are encoded into relatively long-lasting representations (see also Bradlow et al., 1999; Nygaard et al., 1995).

These studies of talker variation show that surface information is in fact stored (although this does not necessarily rule out the possibility that abstract representations also exist; see, for example, Luce & Lyons, 1998). However, there is an inherent difference between regular variation that listeners encounter and variation due to talker characteristics: one type of variation is based on the phonotactics of a particular language, and the other type is not (see Luce, McLennan, & Charles-Luce, 2003; for discussion). An unresolved issue, then, is whether regular surface variation may also be encoded into representations, or whether this type of information is abstracted to a basic canonical form over time. The current project addresses the effects of regular variation on both immediate and long-term processing and examines the relationship between a signal and its representation, either exemplar or canonical, for regular variation.

In addition to regular variation, there is a body of research on mismatching features, where the goal is to explicitly look for conditions that do affect word recognition. A typical approach is to manipulate one or more features in a base word in order to determine if this change affects access to the word (i.e., service–*gervice*). It has been shown that although mismatch by a single feature does not prevent access, word recognition is slowed down significantly (Connine, 1994; Connine et al., 1993; Deelman & Connine, 2001; Marslen-Wilson, Moss, & van Halen, 1996). For example, Connine et al. (1993) conducted a series of cross-modal priming experiments comparing semantic priming for pseudowords that are minimally different (1 feature changed) from a base word (e.g., battern vs. pattern) and maximally different (3 features changed, e.g., rattern). They found that minimally different pseudowords produced significant priming of a related target, independent of the position of the changed phoneme, while there was no priming for targets preceded by maximally different pseudowords.

While mismatching information is arbitrary, phonotactic variation is not. For example, /t/ and /d/ are produced as a voiced alveolar tap [ɾ] intervocalically after a stressed syllable in English, but the same sound is phonemic in other languages (e.g., Spanish) (Ladefoged, 1993). As another example, intervocalic voiced stops /b/, /d/, and /g/ in most dialects of Spanish are produced as the voiced fricatives [*β*], [ð], [ɣ], respectively (e.g., /nada/ → [naða] 'nothing') (Barrutia & Schwegler, 1993).

The type of variation that has received the most attention in the literature is assimilation. Assimilation is the process by which a segment becomes more like a neighboring segment in one or more features. For example, in the phrase *green beans*, the [n] in the word *green* assimilates its place of articulation to match that of the following consonant, making the final consonant more similar to the initial consonant of the following word, yielding something that sounds more like *greem beans* (Gow, 2001). There is a large body of evidence suggesting that this type of variation does not disrupt word recognition (Gaskell & Marslen-Wilson, 1996; Gow, 2001, 2002; Marslen-Wilson, Nix, & Gaskell, 1995). Word recognition may not be disrupted in the case of assimilation for two reasons. First, there is strong evidence that the assimilated segment in such forms is coarticulated, meaning that the sound change of the assimilated segment is not a complete featural change. So, the [m]-like sound in *greem beans* has both coronal and labial place information in the acoustic signal (Gow, 2001). Second, there is an explicit conditioning environment (the adjacent segment) that licenses this change. This is important, since the conditioning segment appears to be required in order to avoid disrupting processing. For example, there is evidence that coarticulated *gree[m]* in isolation does not access the lexical representation *green* (Marslen-Wilson et al., 1995).

Four major theories that have been proposed to account for variation are underspecification (Lahiri & Marslen-Wilson, 1991), phonological inference (Gaskell & Marslen-Wilson, 1996, 1998), feature parsing (Gow, 2003), and articulation-based models (Fowler, 1986, 1996; Fowler, Brown, & Mann, 2000).

Underspecification is a model that stems from phonological theory (Steriade, 1995). In this model, variable segments are unspecified for particular features, depending on the type of variation and the language. Lahiri and Marslen-Wilson (1991) found that the representation of vowel nasalization differs in English and in Hindi. While nasalization is specified for vowels in Hindi, where it is contrastive, it is unspecified in English since it is not a contrastive feature of English. Underspecification can account for assimilation by assuming that the place feature for nasals is not specified and therefore free to take on the place of a following consonant.

In an inference-based view, listeners use phonological context to infer whether a particular surface form may be derived. In this model, activation of a form (e.g., *green* when hearing *greem*) either continues when a viable context is encountered (e.g., *beans*), or is deactivated when the necessary context is absent (e.g., *cucumber*). Gaskell and Marslen-Wilson (1998) found that when listeners monitored for underlying coronal segments which surfaced as non-coronals, they were faster and more accurate when the following context licensed the surface variation than when it did not. In addition, Gaskell and Marslen-Wilson (1996) found stronger priming of underlying coronals by modified surface non-coronals when the modified items appeared in a licensed context. These results have been used to promote the inferential view, in which listeners rely on post-segmental information to compensate for coarticulated sounds. More recently, Gaskell (2003) has proposed a model in which variation is learned to be acceptable through experience. Gaskell shows that a coronal nasal that has been assimilated to a labial resulting in an [m]-like sound may be recognized as [m] at first, but once a context for assimilation is identified, the same sound may later be recognized as [n]. So, it is the exposure to particular variants and context that allows listeners to process variation without difficulty.

Gow (2003) has proposed that, instead of inference, listeners parse feature cues present in the surface form to map and align features with segments. This is based on research showing that assimilated segments encode two places of articulation (one from the underlying segment and one from the following segment (in regressive assimilation)), and that processing of assimilated segments is not due to experience, since listeners are able to process types of assimilation that are not present in their native language. In his model, listeners use the wealth of cues available in the speech signal to map features appropriately to discrete segments.

Fowler (1986, 1996) proposes a theory in which listeners perceive gestures. Therefore, in the case of coarticulation where overlapping gestures occur, a listener is able to establish the relationship between gestures and the speech signal. Fowler et al. (2000) have shown that listeners use phonetic gestural information in order to compensate for coarticulation. When exposed to ambiguous phonemes dubbed onto unambiguously produced phonemes in a video, listeners use the gestural information in the video to identify the ambiguous phonemes. In this view, then, there is no abstract or underlying code, but a strong relationship between the production and perception of speech.

Theory development and our understanding of the effects regular variation has on spoken word recognition would benefit greatly from deeper investigation of a range of different types of variation. After all, assimilation is only one type of variation that occurs in natural language, and we do not have much information about how other types of variation affect processing. Whether or not the available theories are able to be generalized to different types of variation remains to be seen. Not all types of variation result in coarticulated segments, which favor both a feature-parsing approach and an articulatory-based approach, and not all types of variation have explicit contextual cues, which have been used to support an inference-based view. Many licensed surface variants may, in fact, be altered leaving no cue about the underlying segment other than context, phonotactic patterns, and phonotactic probability. In addition, contexts in which many alternations occur do not always involve the presence of adjacent segments, as in assimilation. For example, many alternations across languages are dependent on the more abstract notion of syllable structure (see, for example, Kenstowicz, 1994).

Recently, there have been a few studies focusing on other types of regular variation. For example, Deelman and Connine (2001) examined the importance of release information (or lack thereof) in voiced and voiceless stops in English, and McLennan, Luce, and Charles-Luce (2003) examined access and representational issues related to the tapping of intervocalic /t/ and /d/ in American English. Both of these studies offer important insights into the recognition and representation of regular variants in English.

In English, fully articulated final stops are produced with a release burst. Identification of these stops has been shown to be reduced when this release burst is not present (Malecót, 1958). In addition, when the release information does not match the prerelease information, reaction times increase (Marslen-Wilson & Warren, 1994). Deelman and Connine (2001) compared two regular variants of voiced and voiceless stops in English: released and unreleased stops. They investigated the interplay of voicing status and the presence or absence of a release to see whether a release burst is more important to either voiced or voiceless coronal stops. In a cross-modal semantic priming task, the primes were released variants of a word (e.g., *comba[t], vivi[d]*) and unreleased variants of the same words (e.g., *comba* _, *vivi* _). In the latter, the release bursts were spliced out of the original stimuli. Released and unreleased stops (both voiced and voiceless) showed comparable priming of a semantically related visually presented target.

Deelman and Connine also conducted a phoneme-monitoring experiment examining the importance of release in a task which forces subjects to make a decision on the segment in question. In this experiment, subjects were monitoring for [t] or [d] in both released and unreleased form. The authors found a difference between the effect of a release on voiced stops and its effect on voiceless stops. For voiced stops, there was no difference in reaction times for released and unreleased variants. For [t], however, targets with a release were consistently detected faster than unreleased variants. Deelman and Connine argued that the semantic priming results suggest that variation in form activation is neutralized at the semantic level, whereas the phoneme monitoring results suggest that phonetic cues, such as release, are influential in form activation.

This research provides a useful contrast between activation at the semantic level and activation at the form level when a regular variant of a word is missing information in the speech signal. In natural language, though, not every phonologically regular variation involves the lack of information. Many times, additional information may be included in the speech signal when a variant occurs. For example, word final /t/ in American English has a number of possible regular variable productions. While the context for these variants is predictable (i.e., coda /t/ can vary), the actual variants are in free variation. For example, word-final /t/ can be produced in citation form with a fully released coronal stop, as a glottalized coronal stop that is coarticulated with both coronal and glottal place with no audible release, and as a glottal stop (e.g., the medial sound in the English word *uh-oh*). While it is generally accepted by linguists that these variants are common in American English, there is little research supporting this claim. There is evidence that these three variants are regularly occurring in the Long Island dialect of American English (Huffman, personal communication, September 7, 2004), the population examined in our study. Huffman (1998) found that all three variants occur regularly in the Long Island dialect of American English. Furthermore, at the end of an intonation phrase (i.e., word-final position before a pause), 70% of final-/t/ words are produced as glottalized stops, with both coronal and glottal articulation. The remaining 30% of final-/t/ words are split between the canonical [t] and the glottal stop. A natural question is whether the information in the acoustic signal of these variants benefits semantic or form processing (i.e., whether glottalization cues an underlying /t/), or whether the lack of a singly articulated place of articulation instead impairs processing.

In addition, while glottal stops are not contrastive in English, they do differ from canonical [t] by the feature PLACE. There is an abundance of data suggesting that feature mismatch, while not blocking access to a particular word, does in fact slow down access and/or reduce activation (e.g., Bölte & Coenen, 2002; Connine et al., 1993). One issue surrounding the glottal stop variant of /t/, then, is whether this is true even for legitimate variants of a word, as opposed to an arbitrarily mismatched item. If so, then there is a general cost to feature variation, independent of its regular occurrence in a language. If not, then there is a clear distinction to draw between regular variation and arbitrary mismatch.

In the current study, we examine the effect phonologically regular variation has on spoken word recognition, and the ways in which this variation is manifested representationally. The present experiments take advantage of the fact that both the ideal or canonical form of each phoneme and its variants generally have several usable cues that have the potential to facilitate perception. As discussed above, final /t/ in English has at least three possible acoustic forms: a fully articulated [t], a coarticulated, glottalized [t] (represented by the symbol [ʔt˺]), and a singly articulated glottal stop [ʔ]. Thus, the word *flute* has three regular variants that the word recognition system must be able to recognize (i.e., [flut], [fluʔt˺], and [fluʔ]).

This particular variation is interesting for a number of reasons. First, the ideal or prototypical form (e.g., [t]) is not the most frequently produced variant. Second, the coarticulated variant contains information about the canonical place of articulation (coronal) *and* additional information such as vowel glottalization which may or may not be used by listeners for processing at the semantic level and the form level. Additionally, while this variant is similar to an assimilated segment in that it carries information about the canonical form, the context is more abstract than that of an assimilated segment. For example, as we have discussed, when assimilated, the [n] in *green beans* has cues to the place of articulation of both the /n/ in the word *green* and the labial information matching the initial segment of the following word *beans*. In the variants of word-final /t/, this context is not present. The context for the change is simply the end of a syllable. No additional segment is present to provide a cue about the variant to the listener. In this sense, this type of variation is similar to out-of-context assimilation in which a coarticulated segment is present without the triggering context, which has been shown to incur a cost in processing (Coenen et al., 2001). However, the /t/ variation is common and legitimate in isolation. There are thus three possibilities in the processing of the variants: (a) The supplementary information (vowel glottalization or coarticulation) may enable a speaker to process these variants as well as the canonical form at all levels of processing; (b) the lack of a singly articulated [t], the absence of an explicit conditioning context, and the lack of a release burst will preclude processing at one or more levels, if these cues cannot be used by listeners, or (c) individual representations for each variant are stored and accessed.

Final /t/ variation in American English provides an additional interesting complication: the glottal variant, which is the most deviant form, allows us to (a) consider variation of a single canonical form in different degrees ([t] vs. [ʔt] vs. [ʔ]) and (b) examine featural mismatch when a variant deviates from the canonical form by a feature in a regular, rule-governed manner. These three variants make it possible to ask whether more or less information in the acoustic signal, where variation is regular and predictable, has different consequences for spoken word recognition.

Our first experiment tests whether types of variation (regular or arbitrary) affect how well listeners can recognize a spoken word: Does variation impair perception? Experiments 2 and 3 address representational issues: Do listeners encode the legal variants differently, or are all forms ultimately mapped to a canonical version?

Experiment 1a

It is well established that hearing a word primes the processing of an immediately following related word (e.g., flute–music). In our first experiment, we examine whether the strength of such priming is influenced by phonetic variation in the prime. For example, do all three variants of 'flute' (e.g., [flut], [fluʔt˺], and [fluʔ]) prime 'music' equally? To test this, we use the semantic priming paradigm with final /t/ variants serving as primes for semantically related targets. We expect significant facilitation of the target when primes are legally changed forms, and either less priming or no priming at all when they are arbitrarily changed forms (e.g., [flus]). If legal variants do not hinder access, then we should find no difference in the activation of target words. However, if a feature change does have a cost on processing regardless of the phonological status of the deviation, the glottal variant of /t/ should facilitate target processing less than the basic and coarticulated variants do.

Method

Participants

A total of 96 undergraduate students participated in this experiment for course credit. All participants were native speakers of American English, and none reported any hearing deficiencies. The participants were split into four groups of 24, corresponding to four experimental conditions.

Materials and design

A large-scale norming pretest was used to select semantically related word pairs. Ninety monosyllabic words with final-/t/ endings were selected as possible primes. To select semantically related target words, we presented four printed lists of 45 words to a group of 500 participants (each word occurred on two lists, and each list was given to 125 participants) who were instructed to write down a related word for each item. Thirty-six items from the original list of 90 were chosen as primes based on the number of participants who produced a given target response. The average prime–target pair was the response of 41% of the participants.

All stimuli were recorded by a female phonetician in a sound-dampened room. The stimuli were digitized at 16 kHz. Three versions of each word were produced, corresponding to three experimental conditions. The first version included a fully articulated production of final [t] with a single place of articulation, no glottalization on the vowel, and a full release burst. This version corresponds to the BASIC condition (the canonical form of words ending in [t], e.g., [flut]). Items in the second condition were produced with final glottalized voiceless alvealor stops ([ʔt˺]). All items in the second condition were produced with glottalization on the vowel, coronal and glottal coarticulation, and no release burst (e.g., [fluʔt˺]). This set of stimuli corresponds to the COARTICULATION condition. Items in the third experimental condition (GLOTTAL) were produced with singly articulated glottal stops ([ʔ]) as the final segment. All glottal-final words were produced with the tongue body and tip at the floor of the mouth with no gesture toward the coronal place of articulation. These items were produced without coronal articulation, and with vowel glottalization and a final release burst (e.g., [fluʔ]).

A fourth type of prime (FEATURE MISMATCH) was also recorded. Pseudowords were created by changing the final /t/ in each of the 36 primes to a sound differing from [t] by a single feature (e.g., [flus]). To control for possible featural differences, 12 pseudowords were formed by changing the feature VOICE (i.e., [d]), 12 by changing the feature PLACE (i.e., [p] or [k]), and 12 by changing the feature MANNER (i.e., [s]). Table 1 provides

Table 1
Example stimuli for Experiment 1a: Auditory–Auditory semantic priming

Condition	Prime	Related target	Unrelated target
	'flute'		
Basic-t	[flut]		
Coarticulated	[fluʔt˺]	Music	Money
Glottal	[fluʔ]		
Feature MM	[flus]		
	'debt'		
Basic-t	[dεt]		
Coarticulated	[dεʔt˺]	Money	Music
Glottal	[dε˺]		
Feature MM	[dεs]		

an example of two stimulus sets, including the four primes, and the two targets (related test items and unrelated controls).

To ensure that the stimuli all had the features we desired, the spectrogram of each Basic [t] word was analyzed using the Praat software package (http://www.praat.org) to ensure that it had a strong alveolar release and a lack of glottalization on the vowel. Each Coarticulated [ʔt˺] item was confirmed to have vowel glottalization and no release burst. Finally, each Glottal [ʔ] item was checked to ensure that vowel glottalization and a clear glottal release burst were both present. Appendix B lists the 36 prime–target pairs, and the feature mismatched pseudoword for each prime.

To create a set of semantically unrelated items as controls, we first created two lists, each with the 36 final-/t/ primes. For one list, 18 final-/t/ primes were paired with their related targets (e.g., sleet–snow, kite–fly) and the remaining 18 primes were paired with the remaining 18 targets which were rotated to form semantically unrelated pairs (e.g., eat–rich, jet–graph). The second list contained all 36 primes as well, but the 18 primes that were paired with unrelated targets in List 1 were paired instead with their corresponding semantic targets (e.g., eat–food, jet–plane). The remaining 18 primes were rotated through the list of remaining targets to form 18 semantically unrelated prime–target pairs (e.g., sleet–blood, kite–love). The two lists together, then, accounted for all 36 semantically related prime–target pairs. Each subject heard all 36 items without repetition. We replicated this set of two lists to account for all four experimental conditions. One version included all Basic [t] final primes, the second version included all Coarticulated [ʔt˺] final primes, the third version included Glottal [ʔ] final primes, and the fourth version included Feature Mismatch primes. In total, then, there were eight experimental lists. To each list, 164 filler trials were added (64 with real word targets and 100 with pseudoword targets) to avoid the development of strategies in the responses. The sound [t] was absent from all fillers, as was any semantic relationship between primes and targets. Each subject received one experimental list. Therefore, each subject heard each prime and target only once.

Procedure

Participants completed the experiment individually or in groups of two or three in a sound-dampened booth. All stimuli were presented binaurally over headphones at a comfortable listening level. An ISI of 500 ms was used in Experiment 1a to ensure that semantic activation of the prime would be complete by the time the target was heard (see Experiment 1b for a comparison to a shorter ISI).

On each trial, participants were presented with an auditory prime, followed by a 500 ms ISI, followed by an auditory target. Participants were instructed to make a lexical decision for the second word of a pair. Examples of stimuli were provided and each participant completed a practice session with 20 trials. The experiment lasted approximately 15 min.

Results and discussion

Response times faster than 500 ms and slower than 2500 ms (measured from the onset of the target word) were excluded from all analyses. Based on this criterion, 3.9% of the responses were excluded. Twelve participants were replaced because of high error rates (above 15% incorrect).

Table 2 shows lexical decision reaction times for the four experimental conditions and their controls. Two-factor analyses of variance (Prime type (basic, coarticulated, glottal, and mismatch) × Target type (related or unrelated)) of correct responses were performed for subjects ($F1$) and items ($F2$). Overall, reaction times were significantly faster for related targets than for unrelated targets ($F1(1,92) = 51.09$, $p < .001$; $F2(1,35) = 9.295$, $p < .01$). There was also a main effect of prime type ($F1(3,92) = 7.480$, $p < .01$; $F2(3,105) = 8.460$, $p < .01$). This effect is indicative of the difference between the priming ability of regular variants of final /t/ and the lack of priming seen in the Mismatch condition. An interaction between prime type and target type was also

Table 2
Mean lexical decision reaction times in milliseconds for Experiment 1a: Auditory–Auditory semantic priming with 500 ms ISI

	Related	Unrelated	Priming effect
Basic	831 (2.1)	914 (1.7)	83
Coarticulated	884 (2.8)	944 (2.1)	60
Glottal	866 (1.4)	936 (1.4)	70
Feature MM	928 (2.8)	926 (2.3)	−2

Note. Error percentage for each condition is in parentheses.

found ($F1(3,92) = 6.399$, $p < .01$; $F2(1,35) = 4.116$, $p < .01$). This interaction reflects the large priming effect for the three legal variants of /t/, together with a complete lack of priming for the unlicensed variation.

Planned comparisons support this interpretation. Targets preceded by regular variants were identified more quickly than unrelated items (Basic $F1(1,23) = 33.293$, $p < .001$, $F2(1,35) = 14.404$, $p < .01$; Coarticulated $F1(1,23) = 25.468$, $p < .001$, $F2(1,35) = 8.763$, $p < .01$; Glottal $F1(1,23) = 23.224$, $p < .01$, $F2(1,35) = 5.487$, $p < .05$). This was not the case for the arbitrary mismatch condition; targets preceded by mismatched items were not identified faster than unrelated items ($F1(1,23) < 1$, $F2(1,35) < 1$). The RTs for targets preceded by regular variants of /t/ were not significantly different from each other (Basic–Coarticulated, $F1(1,46) = 1.418$, $F2(1,35) < 1$; Basic–Glottal, $F1(1,46) < 1$, $F2(1,35) < 1$; Coarticulated–Glottal, $F1(1,46) < 1$, $F2(1,35) < 1$). An analysis of the priming effect for each of the Basic, Coarticulated, and Glottal variants also showed no difference among the three variant types ($F1(2,69) < 1$, $F2(2,70) = 1.012$, n.s.).

These results demonstrate that all regular variants of /t/ (basic, coarticulated, and glottal) equally and effectively prime a semantically related target. Regular variation, even variation by a phonological feature, does not hinder semantic activation during spoken word recognition. All three variants provide access to a related word, and there is no apparent benefit of being the basic or canonical form (e.g., [t]) or the most frequent form (e.g., [ʔt]). The contrast between the glottal case (e.g., [fluʔ]) and the arbitrary feature mismatch situation (e.g., [flus]) highlights the critical difference between regular and arbitrary single-feature variation. When activating the lexical representation of spoken words, it is not the degree of variation that matters, but the legitimacy of the variation.

These results contrast with predictions made by exemplar- or experience-based models. Theoretically, in an experience-based model, each time a variant is heard, that particular variant is stored. Therefore, we might expect to see some overall benefit to being the most frequent form. In the case of final /t/, the most common form is the coarticulated variant, which fares no better or worse than the other two regular variants examined. The data are compatible with the remaining theories (i.e., feature parsing, gestural, underspecification), since all predict equivalent recognition for all three variants.

Given that one of the central conclusions depends on a null effect (the equivalence of the regular variants), we conducted a second version of the experiment, using a 100 ms prime–target ISI. This version was run to test whether any immediate effects of deviating from a basic or canonical form might have faded away over the course of 500 ms. If any effects of variation on word recognition did not show up in Experiment 1a because of the relatively long ISI, a shorter ISI should enable us to observe these effects.

Experiment 1b

Method

Participants

Ninety-six undergraduate students from the same population as before participated in this experiment for course credit. None had participated in Experiment 1a.

Materials and design

The design and all materials were identical to Experiment 1a.

Procedure

The procedure of this experiment was identical to Experiment 1a, except that the prime–target ISI was 100 ms.

Results and discussion

Reaction times faster than 500 ms and slower than 2500 ms were excluded from all analyses (4.2% of total responses). Nine subjects were replaced because of error rates (higher than 15% incorrect). Reaction times for Experiment 1b are provided in Table 3. Response times were significantly faster for related targets than for unrelated targets ($F1(1,92) = 22.421$, $p < .001$; $F2(1,35) = 7.076$, $p < .01$). A main effect of prime type was present by item ($F2(3,105) = 13.584$, $p < .01$), but not by subject ($F1(3,92) = 1.579$, n.s.). Critically, the analysis of variance showed an interaction between prime type (Basic, Coarticulated, Glottal, and Mismatch) and target type (Related, Unrelated) ($F1(3,92) = 6.424$, $p < .01$; $F2(3,105) = 7.400$, $p < .01$).

Overall, the results from Experiment 1a were clearly replicated. Targets preceded by semantically related regular variants were identified more quickly than unrelated items (Basic $F1(1,23) = 12.365$, $p < .01$, $F2(1,35) = 9.101$, $p < .01$; Coarticulated $F1(1,23) = 11.640$, $p < .005$, $F2(1,35) = 11.044$, $p < .01$); (Glottal $F1$(1,

Table 3
Mean lexical decision reaction times in milliseconds for Experiment 1b: Auditory–Auditory semantic priming with 100 ms ISI

	Related	Unrelated	Priming effect
Basic	880 (1.8)	945 (1.5)	65
Coarticulated	908 (2.2)	988 (1.3)	80
Glottal	860 (1.2)	909 (1.8)	49
Featural MM	950 (1.4)	960 (1.6)	−10

Note. Error percentages are provided in parentheses.

23) = 11.100, $p < .01$; $F2(1,35) = 6.699$, $p < .05$). Again, this pattern was not found for items in the arbitrary mismatch condition, with a trend in the wrong direction ($F1(1,23) = 2.731$, n.s.; $F2(1,35) < 1$). As in Experiment 1a, the RTs for targets preceded by regular variants of /t/ were not significantly different from each other (Basic–Coarticulated, $F1(1,46) < 1$, $F2(1,35) < 1$; Basic–Glottal, $F1(1,46) < 1$, $F2(1,35) = 1.229$; Coarticulated–Glottal, $F1(1,46) = 1.362$, $F2(1,35) = 1.451$). There was also no difference found in the priming effect among the three common variants ($F1(2,69) < 1$, $F2(2,70) < 1$).

Given the extremely short prime–target ISI, if there are any costs of varying the surface form, they should have shown up in this experiment. Instead, the results for this version fully replicate our findings from the version with a longer ISI. We again found no difference in RTs for targets preceded by regular variants of /t/, with all three variants contrasting sharply with items that arbitrarily vary from basic [t]. Thus, regardless of ISI, our results support the same general conclusion: In activating lexical representations, variation does not matter as long as it is regular. However, any deviation from a canonical form that is *not* governed by the phonotactics of the language, as in the arbitrary mismatch case, appears to impair lexical access, as it should.

So, in immediate processing, regular variation is very well tolerated for all variants. The equivalence of the three variants suggests that there is no direct benefit to being the most commonly produced form, thus lending little support to experience-based models of recognition.

To gain a better understanding of how variation is handled, it is important not only to address the process of recognition, but representation as well. Therefore, the next question we address is how these variants are accessed and stored in the long-term.

Experiment 2

As discussed earlier, recent studies focusing on talker variation show strong evidence in favor of specific information about the form of words being encoded into representations in both implicit and explicit memory (Bradlow et al., 1999; Church & Schacter, 1994; Goldinger, 1996; Nygaard et al., 1995; Schacter & Church, 1992).

Considering phonologically regular variation, McLennan et al. (2003) conducted a series of long-term repetition experiments examining the intervocalic tap in English (the medial sound in *butter*; []), which is a phonologically regular variant of both /t/ and /d/. In the long-term repetition priming paradigm, the focus is on the effects items presented in an initial block of stimuli have on target items presented later, in a second block. It has been consistently shown that listeners are better at recognizing repeated items compared to non-repeated items in a speeded word reading test (Scarborough, Cortese, & Scarborough, 1977). The paradigm has been used to examine the relationship between surface forms and stored representations, as discussed earlier. This task has also been shown to tap into the form level of representation (Luce et al., 2003; McLennan et al., 2003).

The central question for McLennan et al. was whether a surface tap activates a canonical representation. For example, does hearing the word [siɾɪŋ] (seating) activate an abstract representation /sitɪŋ/ and/or a more specific, or detailed representation matching the surface form [siɾɪŋ]? In a single word shadowing task, they found that casually articulated (words produced with a tap, e.g., [æɾəm] or [siɾɪŋ]) and carefully articulated (words produced with [t] or [d], e.g., [ætəm] ‘atom’ vs. [ædəm] ‘Adam’ or [sitɪŋ] ‘seating’ vs. [sidɪŋ] ‘seeding’) items were equally effective at priming both casually and carefully produced targets. So, the casually produced word [siɾɪŋ] primed both ‘seating’ and ‘seeding’ equally, and carefully produced words primed both casual and careful variants, as well. In addition, they conducted an easy discrimination lexical decision task and a hard discrimination lexical decision task with the same stimuli. In the easy discrimination task, listeners were presented with the critical stimuli along with pseudowords that have a low phonotactic probability in English (e.g., *thushshlug*). The hard discrimination task included pseudowords that differ from real words by only one sound (e.g., *bacov* (from *bacon*)). In the easy discrimination task, only items identical on the surface exhibited a repetition effect (e.g., [siɾɪŋ] primed the identical [siɾɪŋ], and [sitɪŋ] primed [sitɪŋ], but [siɾɪŋ] did not prime [sitɪŋ], or vice versa). This was not the case in a hard discrimination lexical decision task, though, where a repetition effect was found between mismatching items (careful–casual, and vice versa). These results suggest that (1) abstract underlying representations are accessed by legal variants when a deeper level of processing is used and (2) both abstract and detailed representations exist.

Although McLennan et al. examined tapping in a repetition paradigm, their results do not necessarily reflect long-term representations, with only 24 trials per block. We have already seen that there is no cost associated with regular variation *in the short term*. It may be the case, then, that their conditions tap immediate processing, similar to the results we found in our semantic priming experiments. In addition, the tap is a predictable allophone of both /t/ and /d/, whereas all variants of final-/t/ are free phonetic variants. The predictable nature of tapping may also reinforce the abstract representation. One question that remains, then, is how legal variants of /t/ are processed over a longer period of time, and how these variants are stored in the lexicon. To address this issue, we again consider the regular variants of /t/ (basic, coarticulated, and glottal), but this time we examine these variants within a long-term repetition paradigm.

In Experiment 2, we use a hard-discrimination lexical decision task, like that of McLennan et al., in a long-term repetition priming paradigm to address the issue of storage and encoding of all variants of final /t/. Experiment 1 demonstrated strong and equal immediate activation of words with all three legal variants of final /t/. One possible interpretation of this result is that all variants are immediately mapped onto one common, abstract representation that could then be maintained in memory. If this is the case, then we should find long-term priming by all variants, for all target variants. Alternatively, consistent with research that shows memory for surface details in implicit memory tasks (Schacter, 1987; Tenpenny, 1995), if exemplar representations of all variants of /t/ are stored in memory, then we expect to find identity effects: only identical items (e.g., [frut]–[frut]; [fruʔt˺]–[fruʔt˺]; [fruʔ]–[fruʔ]) may produce long-term priming effects.

Method

Participants

Sixty students from the same population as in Experiment 1 participated in this experiment for course credit. All participants were native speakers of American English and reported normal hearing.

Design and materials

The critical stimuli consisted of 60 monosyllabic words with final-/t/ endings (the 36 final-/t/ items from Experiment 1 plus an additional 24 words). The stimuli are listed in Appendix B. The stimuli were recorded in a sound-dampened booth by the same phonetician as in Experiment 1, and digitized at 16 kHz. Once again, we used three versions of each word and checked for release burst (or lack thereof) and the absence or presence of glottalization. The three versions correspond to the three experimental conditions: Basic, Coarticulated, and Glottal.

Two blocks of stimuli were formed. Items in the first block served as *primes*, and those in the second block served as *targets*. Each block contained critical items, controls, and fillers. The mapping of the critical items is shown in the top part of Table 4. The first block contained 60 critical items which ended in *either* Basic [t], Coarticulated [ʔt˺], or Glottal [ʔ]. These items served as primes for the critical items in the second block. The three variants presented in Block 1 were varied between subjects. The variants in the second block (containing the *targets*) were manipulated within-subject with 15 items from each of the three conditions (basic, coarticulated, and glottal). We used this design to test each variant's processing as a function of an earlier presentation of each variant. This yielded nine combinations of the critical stimuli.

This design has a number of benefits. First, there is an identity condition for each variant of final /t/ (*Basic*–Basic, *Coarticulated*–Coarticulated, and *Glottal*–Glottal). Therefore, within subjects, we can see whether identical items are being recognized more quickly than their variant counterparts, which would support a view that each variant has its own stored representation. We can also compare the recognition of a particular variant with respect to the variant presented in Block 1 (i.e., *Basic*, *Coarticulated*, and *Glottal*–Basic). Therefore, for a particular variant, we are able to examine

Table 4
Experiment 2 design and sample stimuli

	Block 1: Auditory lexical decision 360 trials		Block 2: Auditory lexical decision 360 trials	
	60 Critical items:			
Between subject	60 Basic [t]	*fru* [t]	15 Basic [t]	*fru* [t]
	60 Coartic.[ʔt]	*fru* [ʔt]	15 Coarticulated [ʔt]	*fru* [ʔt]
	60 Glottal [ʔ]	*fru* [ʔ]	15 Glottal	*fru* [ʔ]
			15 Arbitrary mismatch (pseudowords)	*fru* [p]
	60 Controls:			
	30 Real words	*cave*	15 New real words	*share*
		lash	15 Repeated real words	*lash*
	30 Pseudowds	*spime*	15 New pseudowords	*feace*
		pleak	15 Repeated pseudowords	*pleak*
	240 Fillers		240 Fillers	

the effects of hearing each of the three variants earlier, helping to discriminate among theories of encoding/representation.

To understand how non-critical repeated and new words are treated in these testing conditions, we added 30 real word controls in Block 1. In Block 2, each participant received 15 of the control items that were presented in Block 1, and 15 words that were not previously presented. This resulted in two control conditions: repeated and new. The frequency of repeated control items (77.42) was matched with the frequency of the final-/t/ critical items (74.43). The repeated and new controls established a baseline for non-repeated (new) real words and a repetition effect for those non-critical items that were repeated. These two controls also allowed us to evaluate whether non-identical variants of the critical words were treated as new or as repeated items.

In addition to the real word controls, we included the same control design for pseudowords. There were 180 pseudoword trials, half of which were repeated in Block 2 (e.g., *pleak*–pleak), and half of which were not repeated in Block 2 (e.g., *spime*–feace). An additional 135 real word fillers were added to reduce the possibility of strategic effects (none of the fillers had a final [t]).

Overall, Block 2 (like Block 1) included 180 real words and 180 pseudowords. Half of the items (for both real words and pseudowords) were repeated from Block 1, whereas the other half were new. So, in Block 2, there were 90 repeated real words (45 ending in final /t/), 90 new real words, 90 repeated pseudowords, and 90 new pseudowords.

Procedure

Participants were tested individually or in groups of two or three and were not told at the beginning of the experiment that there would be two blocks of trials. Participants performed the lexical decision task for both blocks of trials. All stimuli were presented binaurally over headphones. Each trial consisted of the auditory presentation of a word or pseudoword followed by a lexical decision. Each decision was followed by 1000 ms of silence before the next trial began. Subjects were instructed to respond to each item as quickly and accurately as possible.

Results and discussion

Response times faster than 500 ms and slower than 2500 ms (measured from the onset of the target word) were excluded from all analyses. One concern about this task is that some participants may not be as accepting of glottal-final items as words as others. To account for this, we used a contingent scoring procedure for critical trials and we used responses to only non-critical items as exclusion criteria for participants.[1] In the contingent scoring procedure, we only included critical trials for which the participant had accepted the Block 1 variant as a word (i.e., we eliminated cases in which the subject called a /t/-word a non-word in Block 1). A total of 4.8% of the responses were excluded from all analyses based upon reaction time exclusion and contingent scoring. The participant exclusion criteria were based only on responses to the 315 non-critical fillers for both Blocks. This yielded a total of 630 non-critical trials. Five participants were replaced because of high error rates (above 22% incorrect). A preliminary analysis showed that error rates across Block 1 and 2 conditions were stable, with no significant effects (Block 1 (Basic, Coarticulated, and Glottal): $F1(2,57) = 1.123$, n.s., $F2(2,84) = 2.545$, n.s.; Block 2 (Basic, Coarticulated, Glottal): $F1(2,42) = 1.044$, n.s., $F2(2,114) = 2.653$, n.s.). Therefore, we were able to consider the reaction times without any concern of speed-accuracy tradeoffs.

In general, the reaction times from Block 1 showed that coarticulated and glottal variants are responded to more slowly than basic variants (Basic, 1009 ms; Coarticulated, 1029 ms; and Glottal, 1076 ms ($F1(2,57) = 3.56$, $p < .05$), ($F2(2,118) = 25.808$, $p < .01$)). This pattern suggests that the Basic form is indeed a better match to the underlying representation of /t/, and that there is thus some cost associated with even regular variation despite the equivalent semantic activation produced by the variants. The reaction times to Block 2 conditions showed the same pattern; in general, varying from the canonical form slows processing of a word. This is independent of which variant was heard in Block 1. Reaction times for three Block 1 conditions crossed with the five Block 2 conditions are provided in Table 5.

Table 5
Reaction times for all Block 2 conditions following each Block 1 variant in Experiment 2

Block 2	Block 1		
	Basic	Coarticulated	Glottal
Basic	902	947	999
Coarticulated	1001	958	1007
Glottal	1071	1013	1060
Repeated control	956	918	973
New control	1012	989	1023

[1] In both this experiment and in Experiment 3, we have analyzed the data several different ways (e.g., with both contingent and non-contingent scoring; with both conservative and liberal criteria for eliminating poorly performing participants). The pattern of results and their significances were unaffected by these variations in analysis; the pattern is robust.

To establish whether a priming effect exists in this design, we conducted a two-factor ANOVA for Repeated and New Controls (Block 1 condition: Basic, Coarticulated, Glottal × Block 2 condition: Repeated, New). Overall, there was a significant difference between Repeated (960 ms) and New (1016 ms) conditions ($F1(1,57) = 14.685$, $p < .01$, $F2(1,28) = 4.151$, $p < .05$). No difference was found for Block 1 condition ($F1(2,57) < 1$, $F2(2,57) = 2.593$, n.s.), and there was no interaction ($F1(2,57) < 1$, $F2(2,57) < 1$, n.s.). This result establishes a strong effect for repeated items in this paradigm, and shows that the three participant groups did not differ in their overall susceptibility to long-term priming.

Having established a clear repetition priming effect, our main interest is the extent to which variation from a canonical form affects form activation. Even though there was no significant effect of Block 1 group, there were certainly differences in individuals' susceptibility to priming. Therefore, for each participant, we subtracted out the Repeated Control mean from each variant mean in Block 2. This value provides us with the most specific index of the priming effect of a particular variant on a particular variant, over and above (or below) the general tendency for that subject to show priming. These values are provided in Table 6.

We conducted a two-factor ANOVA (Block 1: Basic, Coarticulated, Glottal × Block 2: Basic Coarticulated, Glottal) for the facilitation effect values presented in Table 6. There was no overall effect of Block 1 by subject or item ($F1(2,57) = 1.077$, n.s., $F2(2,84) = 1.988$, n.s.). Therefore, there is no overall benefit or cost to hearing a particular variant in the first Block. The ANOVA did yield a main effect of Block 2 variant ($F1(2,114) = 23.45$, $p < .01$, $F2(2,42) = 7.897$, $p < .01$). This result supports the finding from Block 1 that listeners are generally slower to respond to Glottals than to other variants. An interaction was also found ($F1(4,114) = 3.785$, $p < .01$, $F2(4,84) = 5.411$, $p < .01$). To examine whether the Basic–Basic cell was driving the interaction, we conducted a post hoc test comparing the Basic–Basic cell to the other eight cells. There was a significant difference ($F(1,57) = 18.98$, $p < .01$), suggesting that only the Basic variant may be stored.

Table 6
Difference scores (Repeated control—Block 2 reaction time) for the three Block 2 variants of final /t/ following each Block 1 variant

Block 2	Block 1			
	Basic	Coarticulated	Glottal	Mean
Basic	54	−29	−26	−0.33
Coarticulated	−45	−40	−34	−39.67
Glottal	−115	−95	−87	−99
Mean	−35.33	−54.67	−49	

To investigate the pattern of identity priming, we conducted single-factor ANOVAs for each Block 2 variant, dependent on the Block 1 variant. The overall facilitation effect for Basic [t] in Block 2 (Basic–Basic identity was 61 ms faster than Coarticulated–Basic and 110 ms faster than Glottal–Basic) was found to be significant ($F1(2,57) = 5.12$, $p < .05$, $F2(2,28) = 13.877$, $p < .01$). The same was not true for the other variants. (Coarticulated: $F1(2,57) = 1.261$, n.s., $F2(2,28) = 1.272$, n.s.); (Glottal: $F1(2,57) = 1.202$, n.s., $F2(2,28) = 2.204$, n.s.).

Together, these results suggest that there is a strong advantage to being the canonical form (e.g., Basic [t]). Of all the possible combinations of variants, the only case that shows a strong priming effect was the Basic–Basic combination. These results differ from those found by McLennan et al. (2003), who found that a surface variant was able to activate both a specific form, *and* an abstract form for medial taps in English. One interesting point about these results is that the strong priming effect found was not for the most frequent variant (the coarticulated form), but for the ideal form (Basic [t]).

The results of Experiment 2 do not support a model in which exemplar representations are stored in long-term memory for regular variants of final /t/. If so, we should have seen identity priming for the coarticulated variant. In addition, it also appears that the variants are not being recoded into a single abstract form. If this were the case, we should have seen, at the very least, faster reaction times for all Basic [t] items in Block 2 independent of the variant presented in Block 1. We appear to have a result that is not consistent with available models of variation and representation. We see a strong effect for the canonical form with no evidence that other variants are being recoded into this ideal form.

Before drawing any broad theoretical conclusions, it is reasonable to consider whether this effect might be specific to the particular task that we used. As discussed in the literature, many differences have been found between implicit memory tasks, such as lexical decision in a long-term repetition paradigm, and explicit tasks, such as old–new recognition. It is possible that a different task may be more sensitive to form-based variation. Within the speaker variation literature, listeners have been found to be better at recognizing an old, or repeated, word when that word was repeated in the same voice than when it was repeated in a different voice (Palmeri, Goldinger, & Pisoni, 1993). Bradlow et al. (1999) found the same effect for words repeated at the same speaking rate (see also Nygaard et al., 1995). It is possible then, that the same might be true with surface variation examined in the current study.

One possibility is that the nature of the lexical decision task bolsters a preference for the canonical form, independent of its relatively low frequency compared to the coarticulated form. If regular variation patterns like other types of surface variation (e.g., voice, speaking rate), then we would expect to find identity effects when the items share the same variant of final /t/. To address this possibility, we conducted a second long-term repetition experiment in which the task in the second block was an old–new recognition task.

Experiment 3

In an old–new recognition task, participants are presented with a set of items in the context of some task, such as lexical decision. They are not told at the beginning of the experiment that they will need to recognize these items later. Participants are then presented with a second block of items in which they must decide whether each word or pseudoword they hear is *new* (not repeated from the first block) or *old* (repeated from the first block).

Method

Participants

Sixty undergraduate students participated in this experiment for course credit. All were native speakers of American English with normal hearing.

Materials and design

The materials and design were the same as in Experiment 2.

Procedure

The procedure was the same as in Experiment 2, except for the task in Block 2. In this experiment, participants made lexical decisions for each item in the first block, but the task for the second block was an old–new recognition task. Participants were instructed to respond "OLD" if they had heard a word or pseudoword that had been presented in the first block, and "NEW" if they heard a word that was *completely new* and had not been presented in Block 1. We emphasized the fact that participants should only respond "NEW" if the item was an entirely new word or pseudoword. This, we hoped, would not draw the participants' attention to the critical items, but at the same time lessened the possibility that they would respond "NEW" to a non-identical variant. Responses were made by pushing buttons labeled "OLD" or "NEW."

Results and discussion

Of the 60 original participants, 8 were replaced due to high mean error rates (above 35%). As in Experiment 2, we used only the responses to non-critical items to exclude participants. Higher error rates were expected and observed in this experiment because remembering an item among hundreds that were presented is much more difficult than deciding if an item is a real word. Thus, we expected that error rates would be the primary index of effects. A preliminary analysis of reaction times was conducted, to determine if the planned analyses of errors needed to be qualified by any speed-accuracy issues. Reaction times faster than 500 ms and slower than 2500 ms were discarded and not used in any analyses. We also adopted the contingent scoring procedure used in Experiment 2 (recall (see Footnote 1) that the data pattern was unaffected by scoring details). Based on these criteria, 4.1% of responses were excluded. A two-way analysis of variance showed that there was no overall effect on reaction times of Block 1 condition ($F1(2,57) = < 1$, n.s., $F2(2,140) = 1.619$, n.s.) or Block 2 condition ($F1(4,228) = 1.727$, n.s., $F2(4,70) = 1.317$, n.s.). No interaction between the two conditions was found ($F1(8,228) = 1.423$, n.s., $F2(8,140) < 1$).

We therefore concentrated on the error rates (as is typically done in these studies, e.g., Bradlow et al., 1999; Nygaard et al., 1995; Palmeri et al., 1993) for the three variants of final /t/, with respect to what was heard in the first block and how a particular

Table 7
Mean error rates for the old–new recognition task in Experiment 3

Error type	Block 2	Block 1		
		Basic % incorrect	Coarticulated % incorrect	Glottal % incorrect
Miss	Basic	17 (1159)	31 (1180)	30 (1152)
Miss	Coarticulated	26 (1106)	29 (1123)	22 (1138)
Miss	Glottal	27 (1163)	30 (1171)	27 (1148)
Miss	Repeated	27 (1183)	32 (1201)	31 (1123)
False alarm	New control	32 (1159)	33 (1165)	35 (1160)

Note. Reaction times for each condition are provided in parentheses.

variant was recognized in the second block overall. Table 7 provides the error rates for the three final-/t/ conditions.

Based on the instructions that only entirely new items should be called "NEW," we counted all "NEW" responses to all final /t/ variants as errors.

Similar to Experiment 2, we conducted two-factor analyses of variance (Block 1: Basic, Coarticulated, Glottal × Block 2: Basic, Coarticulated, Glottal) by subject and item. The results showed an effect of Block 1 condition by item and by subject ($F1(2,57) = 3.599$, $p < .05$, $F2(2,84) = 8.670$, $p < .01$) and no effect for Block 2 condition ($F1(2,114) = 1.003$, n.s., $F2(2,42) < 1$). An interaction was found between the variant types presented in Blocks 1 and 2 ($F(4,114) = 3.786$, $p < .01$, $F2(4,84) = 4.243$, $p < .05$). A planned analysis on the Basic–Basic cell showed that it is different from the other cells ($F(1,57) = 18.44$).

Once again, we are interested in how a particular Block 2 variant fares with respect to the three Block 1 conditions. For Basic [t] items in Block 2, there is a main effect of Block 1 condition ($F1(2,57) = 12.590$, $p < .01$; $F2(2,28) = 15.344$, $p < .01$). This effect is clearly driven by the low error rate in the Basic–Basic cell. Listeners make fewer errors in recalling Basic [t] items in Block 2 when they followed Basic [t] words in Block 1 than when they were preceded by Coarticulated or Glottal variants in Block 1. This result is a replication of the Basic–Basic effect found in the reaction times for Experiment 2. It is striking that this result is task independent, in contrast to many voice specificity effects which show a distinction between implicit and explicit memory tasks.

We also see clearly in this experiment that this result is not a general identity effect, as there is no effect of Block 1 condition for Coarticulated Block 2 items ($F1(2,57) = < 1$; $F2(2,28) = 1.589$, n.s.) or for Glottal Block 2 items ($F1(2,57) < 1$, $F2(2,28) < 1$). Error rates are generally stable across all critical conditions, with a range from 24% incorrect to 35% incorrect, except for the Basic–Basic cell which is very low at 17% incorrect. It is also clear from the stable reaction times that this result is not due to a speed-accuracy tradeoff.

Overall, the results here replicate those from Experiment 2 and confirm that we are not seeing an identity effect. If that were the case, at the very least, we would have seen lower error rates for Coarticulated–Coarticulated, as the coarticulated form is the most frequently produced variant of the three /t/ variants examined; no such Coarticulated–Coarticulated advantage was found. What we are seeing is a strong benefit for the canonical form across two very different tasks, along with a weaker ability of variants to activate the stored representation *in the long term*. This provides evidence that the basic [t], not a different variant and not an underspecified form, is in fact stored in these words.

General discussion

Overall, there were two main findings: regular variants of the canonical form are equally effective in activating a semantic associate in the short term, but they are not as effective in accessing exemplar or abstract representations in the long term. While these two effects may appear to be independent of each other, together they help explain both the influence of a representation on the recognition process, and the factors involved in developing and maintaining a representation. Experiment 1a demonstrated that the three legal variants of final /t/ (Basic [t], Coarticulated [ʔt̚], and Glottal [ʔ]) all prime semantically related targets equally and effectively with a 500 ms ISI between prime and target. Both the coarticulated and glottal variants, showed levels of activation of targets similar to that produced by the basic, or canonical, [t]. Mismatched variants (pseudowords with a final segment that differs from /t/ by a single feature, like the glottal variant, but one that is not phonotactically productive) were unable to prime semantically related targets. Experiment 1b replicated these findings with a shorter ISI of 100 ms. These results provide evidence that variation does not hinder immediate lexical access, as long as the variation is regular.

In contrast, the three variants do not appear to be equivalent in their ability to promote activation of a lexical representation over the long term. In Experiment 2, the Basic [t] variant yielded the strongest effects of form activation, while the Coarticulated and Glottal variants were not as effective. There was little evidence that regular variants other than Basic [t] were able to be recoded into a durable form. It appears as though Basic /t/ is the long-term code, and regular variants of /t/ are less good at invoking this code than the identical surface form. Furthermore, there was also little evidence in support of an overall identity priming effect. While there was a trend in the data for Coarticulated and Glottal identity effects in Experiment 2, this trend was not found in the old–new task of Experiment 3. In contrast, the Basic [t] identity priming was robust across the implicit (Experiment 2) and explicit (Experiment 3) measures. The overall similarity of the results across a change from lexical decision to old–new recognition provides strong evidence that in the long-term, even legal variants do not support the laying down of episodic lexical representations.

Taken together, the immediate processing results and the long-term results suggest that the recognition process is largely based on the use of feature or gestural cues. The results of all experiments lend little support to experience or inference-based models like that of Gaskell and Marslen-Wilson (1996, 1998) or Gaskell (2003). It appears that we are able to cope with surface variation during the recognition process without much influence of the stored representation. Therefore, we must be able to extract relevant information from the surface form during immediate processing, making all variants equivalent in the short term. In addition, the reverse can be said for issues of representation. There is clearly a long-term advantage to being a basic [t]. The results suggest that the ideal form is stored in the representations of these forms, with no additional information based on frequency or experience. This does not mean that idiosyncratic information is not stored, as proposed by Goldinger (1996, 1998), but it does mean that there is a strong benefit for the ideal form in the representation of regular variants.

This result pattern allows us to draw a distinction between immediate activation at the semantic level, and long-term form activation. In the short-term, legal variation does not disrupt semantic activation: even with a very short (100 ms) prime–target ISI, all variants yield equal semantic priming. We are able to use information, such as release and glottalization, to process regular variants immediately. The frequency of a particular variant, however, does not appear to be relevant to immediate processing. In the long-term, however, the inequality of the three variants in their ability to support long-term lexical activation suggests that the cues used for immediate processing are not stored in recognition memory. One important question is why the representational preference is for basic [t] when it is rarely produced in word-final position. One possibility is that the glottal stop (or a coarticulated variant) simply cannot exist in the representation of an English word since it is not a phoneme of the language. Glottal stops and glottalization are not contrastive in English. They may act purely as surface cues to signal a voiceless sound, and not as a regular phoneme of English. This rationale also helps explain how all three variants are equivalent in immediate processing. If glottalization serves as a cue, then listeners can extract that cue in order to recognize the final-/t/ words, without needing information available from the representation.

The substantial and equivalent short-term semantic priming that we found for the three /t/ variants fits nicely with the existing literature on regular variation. Several different types of variation have been examined in the literature, using a number of tasks, and results consistently show that phonological variation causes little or no problem in short-term processing. For example, Deelman and Connine (2001) found no difference in the activation of a semantically related visual target word for released and unreleased variants of prime words. Using a gating task, Marslen-Wilson et al. (1995) also showed this to be the case for English place assimilation, with legal variants of coronals identified as such. The additional in-depth studies of assimilation conducted by Gow (2001, 2002, 2003) provide similar results for legal variants of a word in sentential contexts.

In the short term, then, the preponderance of evidence supports the conclusion that phonological variation does not disrupt semantic processing. In contrast, a similar generalization cannot be made for long-term effects for two reasons. First, there have been few studies of the effect of regular variation in the long term; most longer-term studies have focused on more idiosyncratic variation (e.g., variation due to speaker differences). Second, for the research that does exist, there are apparently conflicting results. McLennan et al. (2003) studied long-term repetition priming for taps in English. They reported that deep processing (e.g., a hard discrimination lexical decision task, also used in our Experiments 2 and 3) enabled listeners to access the form representations equally for all variants. Using the same task, we found contrasting results for final /t/ variation: only the canonical form [t] was successful at activating the form level representation in the long term.

A crucial difference between the two studies, however, is how we define *long-term*. McLennan et al. presented listeners with two blocks of 24 items to examine representational issues for English taps, whereas in our experiments, listeners heard two blocks of 360 items each. We used these longer blocks both to avoid strategic responses (by primarily presenting subjects with non-final-/t/ items), and to make sure any short-term effects were gone. Therefore, one possible explanation for the differing results is that the costs associated with long-term processing of variants did not have sufficient time to surface in the McLennan et al. design. The equivalence of the tap variants in that study might reflect the pattern of variant equivalence that characterizes short-term semantic processing. A study of tap variation that employed the larger number of fillers used in our study could clarify whether the apparent discrepancy merely reflects a timing issue, or is instead a more theoretically interesting difference between the different types of variation.

More generally, the strikingly different results we have observed for variants over the short versus the long term suggest that examining other forms of regular variation, using different time frames, could prove very informative. In particular, although assimilation has been the subject of considerable research, no studies have been done that examine long-term effects of assimilation. By broadening both the range of types of regular variation, and the temporal window of investigation, we can make much greater progress in understanding how language users accommodate regular variation.

Appendix A

Experimental stimuli used in Experiments 1a and 1b

Prime	Mismatch	Related target	Unrelated target
mutt	[mʌp]	dog	dry
height	[hʌɪs]	tall	painting
freight	[fɹeɪp]	train	horse
neat	[nip]	clean	castle
chat	[tʃæs]	talk	electricity
dart	[daɹd]	board	bed
hate	[heɪd]	love	water
fruit	[fɹud]	apple	train
chart	[tʃaɹs]	graph	dog
debt	[dɛs]	money	music
eat	[ip]	food	dumb
out	[aʊs]	in	shoe
start	[staɹp]	finish	black
scout	[skaʊd]	boy	pepper
flute	[flus]	music	money
cut	[kzinvvd]	blood	snow
wilt	[wɪlp]	flower	bag
smart	[smaɹd]	dumb	finish
jet	[dɛp]	plane	graph
kite	[kaɪd]	fly	love
white	[wʌɪs]	black	apple
cot	[kas]	bed	board
volt	[voɫd]	electricity	talk
moat	[moʊs]	castle	fly
colt	[koɫp]	horse	clean
art	[aɹd]	painting	tall
wet	[wɛs]	dry	plane
bait	[beɪp]	fish	rich
tote	[toʊs]	bag	flower
sleet	[slid]	snow	blood
boat	[boʊs]	water	sun
salt	[saɫd]	pepper	boy
melt	[mɛɫp]	ice	fish
boot	[bud]	shoe	in
yacht	[jap]	rich	food
bright	[brʌɪp]	sun	money

Appendix B

Experimental stimuli used for Experiments 2 and 3

Final /t/	Mismatched pseudoword
art	ard
bait	baip
beet	beece
belt	belp
boat	boce
boot	bood
bright	brighp
chart	charce
cheat	chead
coat	coce
colt	colp
cot	coss
court	courp

Appendix B (*continued*)

Final /t/	Mismatched pseudoword
cult	culd
dart	dard
debt	dess
dirt	dird
drought	droughd
eat	eap
float	floce
flute	fluce
freight	freighp
fruit	fruid
goat	goce
hate	hade
hate	hace
heart	hearce
height	hice
jet	jep
kite	kide
light	lighp
melt	melp
moat	moce
mute	mjude
mutt	mupp
neat	neap
out	aus
plate	plape
port	porce
quilt	quilp
root	rooce
salt	sald
scout	scoud
shirt	shird
skate	skade
sleet	sleef
smart	smard
split	splid
start	starp
street	streep
throat	throap
tight	tice
tote	toce
treat	treap
volt	vold
vote	vode
wet	wess
white	wice
wilt	wilp
yacht	yachp

Three versions of each final-/t/ word were used: Final Basic [t], Coarticulated [ʔt˺], and Glottal [ʔ].

References

Barrutia, R., & Schwegler, A. (1993). *Fonética y fonologia españolas*. New York: Wiley.

Bölte, J., & Coenen, E. (2002). Is phonological information mapped onto semantic information in a one-to-one manner. *Brain and Language, 18*, 384–397.

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

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*, 521–533.

Coenen, E., Zwitserlood, P., & Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes, 16*, 535–564.

Connine, C. M. (1994). Vertical and horizontal similarity in spoken word recognition. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 107–120). Hillsdale, NJ: Erlbaum.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition?. *Journal of Memory and Language, 32*, 193–210.

Deelman, T., & Connine, C. M. (2001). Missing information in spoken word recognition: Nonreleased stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 656–663.

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

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America, 99*, 1730–1741.

Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human perception and performance, 26*, 877–888.

Frauenfelder, U. H., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes, 16*, 583–607.

Gaskell, G. (2003). Modelling regressive progressive effects of assimilation in speech perception. *Journal of Phonetics, 31*, 447–463.

Gaskell, G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144–158.

Gaskell, G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 380–396.

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

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

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*, 133–159.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance, 21*, 344–359.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics, 65*, 575–590.

Huffman, M. (1998). Segmental and contextual effects on coda glottalization. *Journal of the Acoustical Society of America, 103*, 1818.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

Ladefoged, P. (1993). *A course in phonetics* (3rd ed.). New York: Harcourt Brace.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon.

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

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). Oxford, England: Oxford University Press.

Malecót, A. (1958). The role of releases in the identification of released final stops. *Language, 34*, 370–380.

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

Marslen-Wilson, W. D., Nix, A., & Gaskell, G. (1995). Phonological variation in lexical access: Abstractness, inference, and English place assimilation. *Language and Cognitive Processes, 10*, 285–308.

Marslen-Wilson, W. D., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1376–1392.

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

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57*, 989–1001.

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*, 309–328.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 1–17.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.

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.

Steriade, D. (1995). Underspecification and markedness. In J. Goldsmith (Ed.), *The handbook of phonological theory*. Oxford: Blackwell.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2*, 339–363.