How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words

Mark A. Pitt *

Department of Psychology, 1835 Neil Avenue, Ohio State University, Columbus, OH 43210-1222, United States

ARTICLE INFO

ABSTRACT

One account of how pronunciation variants of spoken words (*center*-> "senner" or "sennah") are recognized is that sublexical processes use information about variation in the same phonological environments to recover the intended segments [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]. The present study tests the limits of this phonological inference account by examining how listeners process for the first time a pronunciation variant of a newly learned word. Recognition of such a variant should occur as long as it possesses the phonological structure that legitimizes the variation. Experiments 1 and 2 identify a phonological environment that satisfies the conditions necessary for a phonological inference mechanism to be operational. Using a word-learning paradigm, Experiments 3 through 5 show that inference alone is not sufficient for generalization but could facilitate it, and that one condition that leads to generalization is meaningful exposure to the variant in an overheard conversation, demonstrating that lexical processing is necessary for variant recognition.

© 2009 Elsevier Inc. All rights reserved.

Introduction

Variability in the acoustic realization of spoken words is a major challenge for understanding how verbal communication succeeds. One form this variability takes is in the pronunciation of words in different ways. Factors such as speech rate and phonological context can turn a word like *probably* into [prabli] or [prali]. Even the highly reduced form [prai] occurs 4% of the time in the Buckeye corpus of conversational speech (Pitt et al., 2007). That communication does not break down when these pronunciation variants are heard demonstrates how well equipped listeners are to process such variation.

Views on how listeners recognize pronunciation variants can be divided into proposals that emphasize a representational solution and proposals that emphasize a processing solution. Representation-based accounts differ in terms of the amount of detail encoded in lexical entries. On the abstractionist side, Lahiri & Reetz (2003; Lahiri & Marslen-Wilson, 1991) postulate that words are stored in memory as bundles of phonetic features, but only marked features are represented. Lexical representations are therefore underspecified, possibly containing less detail than their citation forms. The implications of this for variant recognition is that processing will be insensitive to, and thus unaffected by, variation in unmarked features, regardless of the type or degree of variation.

The majority of representational accounts take an opposing view, that pronunciation variation is in fact encoded in lexical entries. Although they are in agreement that a listener's exposure to a variant plus the frequency with which it is encountered are important determiners of the structure of a word's representation, they differ in the level of detail that is encoded. Ranbom and Connine (2007) suggest that a separate lexical representation is formed for each pronunciation variant of a word. Exemplar

* Fax: +1 614 688 3984.
   E-mail address: pitt.2@osu.edu.
   URL: http://lpl.psy.ohio-state.edu.

models (Johnson, 2006; Pierrehumbert, 2001) differ from this view in two ways. One is that they postulate only one lexical entry for all variants of a word. The other is that, like their counterparts in the categorization literature (Nosofsky, 1986; Nosofsky & Zaki, 2002), they also maintain that every instance of a word ever encountered, including pronunciation variants, is encoded in the same representation, making it truly opposite of abstractionist accounts because extremely fine-grained phonetic detail is stored in memory.

Processing accounts of variant recognition postulate that sublexical mechanisms intervene prior to lexical access to rectify variation in pronunciation to the citation form of the word. They can be organized according to the nature of the input to which the process is applied. At the most basic level, there are proposals that invoke primitive auditory processes to explain perceptual effects found when studying variant recognition and related phenomena, such as compensation for coarticulation (Lotto & Holt, 2006; Mitterer & Blomert, 2003; Mitterer, Csepe, & Blomert, 2006). Slightly more abstract than this is the proposal by Gow (2003), Gow & Im, 2004) that principles of perceptual organization facilitate variant processing by sorting out which phonetic features in the signal belong to which segments or words when speech is reduced. The most abstract processing account is phonological inference (Gaskell & Marslen-Wilson, 1998), in which phonological rules undo the surface variation created by speech reduction into the underlying citation form. Phonological inference is the most well-studied processing account, and is the focus of the present investigation.

What makes a phonological inference process plausible is that pronunciation variation is often systematic. When /t/ occurs intervocalically immediately after a stressed syllable (e.g., *better*), it is almost always flapped in American English (Patterson & Connine, 2001). Reduced vowels show a tendency to be deleted in post-stress syllables in three-syllable words (e.g., *cabinet -> cabnet*; Patterson, LoCasto, & Connine, 2003). Word-final /t/s and /d/s tend to be deleted with regularity, especially after /s/ and /n/ (Deelman & Connine, 2001; Mitterer & Ernestus, 2006; Neu, 1980), and reduction increases with word frequency (Jurafsky, Bell, Fosler-Lussier, Girand, & Raymond, 1998). In English, word-final coronals (/t, d, n/) can undergo regressive place assimilation when followed by a bilabial or velar consonant (*green ball -> greem ball*; Dilley & Pitt, 2007; Gow, 2003; Nolan, 1992). The same is found in Dutch for voicing instead of place of articulation (e.g., *pit bull -> pid bull*; Ernestus, Lahey, Verhees, & Baayen, 2006).

As an account of variant recognition, phonological inference is attractive because it takes advantage of these systematicities across identical phonological environments and compresses them into a single rule that can then be applied to all words that exhibit that same form of variation, including new words. Although the inference mechanism has received the most attention in the literature, Gaskell and Marslen-Wilson (1998) maintain that lexical processes work in conjunction with phonological processes during variant recognition. Convincing evidence supporting this claim comes from an experiment on processing words that undergo regressive place assimilation. They found that lexical status had an effect on performance independent of phonological context. When monitoring for /t/ in assimilated words (*freighp bearer*), participants misdetected /t/ more often in words than in nonwords, clear evidence of a lexical bias.

If phonological and lexical processes work together to ensure variant recognition, then in principle, it should be possible to find evidence of each operating independently of the other. Defining such boundary conditions is a necessary step toward understanding the role of each process. The purpose of the present investigation was to test the limits of phonological inference. Simply put, how much work can phonological inference do alone, independently of lexical involvement?

Some of the most convincing experimental evidence of the existence of an inference mechanism comes from studies in which similar effects are found with words and nonwords. Because nonwords do not have representations in lexical memory, comparable results suggest not only that the locus of the effect is sublexical, but that the rectification process itself is phonological. Gaskell and Marslen-Wilson (1998) used phoneme monitoring to learn which phoneme listeners perceived at the end of an utterance that occurred in a context in which regressive place assimilation was viable (*freighp bearer*) or not (*freighp carrier*). Even though in both cases the surface form, /p/, was specified in the speech signal, listeners misdetected /t/ more often in the viable than unviable context. Importantly, the effect was found for nonwords as well (e.g., *prayp bearer*), suggesting that phonological inference, triggered by the following context, restored the /p/ to its underlying form (i.e., /t/).

Mitterer and Blomert (2003) present data that leads to a similar conclusion. They manipulated lexicality across languages instead of stimuli. The stimuli were Dutch equivalents of the two-word sequences above, and listeners had to decide whether the first word was one of two alternatives (e.g., *freight* vs. *freighp* in English). Dutch listeners made a large number of errors only in the viable context (e.g., hearing *freighp* as *freight*, when followed by *bearer*). When the experiment was repeated with German listeners, for whom the sequences are nonwords, a highly similar pattern of results was obtained. Why? Regressive place assimilation occurs in German, so the same phonological process could have been applied in the appropriate context.

The logic of this cross-language manipulation was taken a step further by Mitterer et al. (2006a, 2006b; Experiment 1) who conducted a similar experiment using Hungarian words. Not only are such stimuli nonwords to Dutch listeners, but the type of assimilation used is not present in Dutch, so there should be no bias to hear the assimilated phoneme as its underlying form. Surprisingly, Dutch listeners responded like Hungarian listeners, only the magnitude of the effect was much smaller (approximately half the size), indicating that something other than language-specific phonological inference altered Dutch listeners' perception of the assimilated phoneme. Using French and English listeners and stimuli from both languages, Darcy, Ramus, Christophe, Kinzler, and Dupoux (in press) also found that segment rectifica-

tion generalizes to unfamiliar languages. Although English and French listeners readily restored assimilated phonemes to their intended form in their native language, they did so as well in the unfamiliar language (French or English), although to a less degree.

Most recently, Mitterer and McQueen (2009) evaluated phonological inference as an explanation for recognizing words in Dutch whose word-final ts are reduced (i.e., deleted or not released). The visual-world paradigm was used in which eye movements to objects were tracked when hearing pairs of words, embedded in phrases, that differed in the frequency with which word-final /t/ deleted (based on statistics from a Dutch speech corpus). The first word of the pair differed in the presence of a final /t/ (e.g., *kast/kas*), and the following word began with a segment that led to /t/ reduction very frequently (e.g., *b*) or less frequently (e.g., *n*). Looking behavior matched the biases of deletion frequency found in the speech corpus, suggesting that participants' knowledge of production in these environments assisted in perceiving /t/ when it was not present. As in the literature on assimilation, perceptual effects that span a word boundary are exactly the type of environment that could be exploited by an inference mechanism which learns regularities in speech reduction, in this case that /t/ is reduced more before /b/ than /n/.

What the preceding studies have in common is that they are tests of generalization and its limits. They demonstrate generalization of a phonological rule's application to a novel context, whether it be to nonwords, words in a foreign language, or another type of reduction phenomenon. The present study follows in this line of work by testing generalization to a variant of a newly learned word. After learning the citation pronunciation of a new word, will listeners process a pronunciation variant similarly to its citation form? As long as the variation is licensed by the phonological environment, a phonological inference account predicts generalization should occur. Indeed, rule formation is assumed to occur for precisely this reason, to generalize so that recognition succeeds. Failure to generalize would suggest a limiting condition for an inference mechanism, and identify at least one circumstance in which lexical involvement is necessary.

The focus on learning new words and the paradigms chosen to test generalization made it preferable to use a word-internal phonological process in place of the more popular word-final processes, even though this choice extends the theory beyond the environment in which it was originally developed. The choice was based on a consideration of the conditions that are likely to promote the development of a phonological inference rule. A constant phonological environment and frequent and predictable phonological variation in that environment are necessary. The former facilitates rule formation and the latter is a prerequisite of rule formation. One type of variation that meets both criteria is medial /t/ deletion (Neu, 1980). In their study of word-medial /t/ and /d/ variation in the Buckeye corpus of conversational speech, Raymond, Dautricourt, and Hume (2006) found that phonological properties of the surrounding environment greatly influenced segment realization. In particular, when /t/ occurs at the onset of a post-stress (reduced vowel) syllable when preceded by /n/ (e.g., *counter, cantaloupe*), /t/ deletion (i.e., production of a nasal flap), occurs frequently. The specificity of the phonological environment that licenses /t/ deletion is evident when words are considered that differ only in the quality of the post-/t/ vowel, being full instead of reduced (e.g., *context*, *contact*).

Data on /t/ realization in both types of words are shown in Table 1. The data are a subset of those from the corpus, for two-syllable words only, which are the primary stimuli used in the present study. In the first row are words with a reduced second vowel. Note the contrast in frequency between words labeled with a fully realized versus deleted /t/ (.05 vs. .75). The proportions change minimally (deletion frequency actually increases) when words with three or more syllables are included. In the second row are the data from words with a full second vowel. /t/ deletion drops dramatically in this context to .11. Across the two classes of words, deletion-biased (DB) and citation-biased (CB), the proportions reverse in the first two columns. /t/ is most often realized as [t] in the citation-biased words, making their canonical pronunciations what listeners are most likely to hear. Just as importantly, /t/ deletion is very infrequent, making it rare that listeners hear such variants of these words.

These statistics suggest that any phonological process that evolves for /t/ deletion will be very environment specific, to the point of restoring the medial /t/ when *counter* is pronounced as [kaune] but not when *content* is pronounced as [kanent]. That is, application should be restricted to DB words. In Experiments 1 and 2, direct and indirect measures of lexical activation were used to demonstrate that this environment satisfies the conditions necessary for a phonological inference mechanism to explain the recognition of /t/-deleted variants. Such a demonstration is a necessary prerequisite for testing generalization, which was done in Experiments 3–5.

Table 1
Proportions and frequency counts of /t/ realization in bisyllabic words in the Buckeye corpus that have a medial /nt/ cluster, primary stress on the first syllable, but vary in the quality of the second vowel.

Second vowel		Realization of /t/					Tokens	Types
		t	Deleted	Glottal	*d*	Other		
Reduced (deletion-biased)	Proportion	.05	.75	.06	.07	.07		
	Frequency	36	508	39	45	50	678	51
Full (citation-biased)	Proportion	.65	.11	.01	.17	.06		
	Frequency	294	49	3	80	25	451	13

Experiment 1

In Experiment 1, listeners made lexical decisions to citation and deleted pronunciations of DB and CB words. All stimuli should be categorized as words except the deleted pronunciations of the CB words, which are likely to be heard as nonwords. Classification response times should show similar selectivity, with "word" RTs being slowest to the deleted version of the CB words. For the DB words, RTs should be slower to deleted than citation pronunciations. This latter prediction comes from a growing literature showing that pronunciation variants generate weaker lexical activation than their citation forms (LoCasto & Connine, 2002; Ranbom & Connine, 2007; Sumner & Samuel, 2005). These outcomes across conditions would demonstrate the selectivity of lexical activation necessary to hypothesize the operation of an inference mechanism in this phonological environment.

The comparisons in this experiment can be thought of as word-internal equivalents of conditions used in experiments examining regressive place assimilation (Gaskell & Marslen-Wilson, 1998; Gow, 2003), where the phonological structure of the DB words corresponds to the viable context, which licenses /t/ rectification, and the phonological structure of the CB words is the unviable context, which does not.

Method

Stimuli

Two variables were crossed in this experiment, whether the target word was a member of the class of words in which medial /t/ frequently or rarely deletes (DB vs. CB) and how the medial /t/ word was pronounced ([t] vs. nasal flap). Half of the 36 target words were citation-biased and half deletion-biased. Their status was determined from counts in the Buckeye speech corpus (Pitt et al., 2007) and from choosing words with similar phonological structure. The 18 words of each type were split equally between two- and three-syllables in length. All are listed in the Appendix A.

Fillers (78 words and 96 pseudowords) were chosen such that when combined with the 36 target words, there were approximately equal proportions of one-, two-, and three-syllable words. Nonwords were created by altering consonants and vowels of words not already used in the experiment. Because it was impossible to know whether listeners would hear the deleted pronunciations as words or pseudowords, the number of words and pseudowords that were presented could not be balanced perfectly.

Stimuli were recorded by a native speaker of American English onto DAT tape in a sound-dampened room, digitally transferred to a PC where they were downsampled to 16 kHz (7.8 kHz low-pass filtered), and then stored as individual sound files. Care was taken to pronounce the deleted pronunciations of the DB and CB words in a natural style, with pronunciations of the CB words being practiced a number of times prior to recording.

Acoustic analyzes of the DB words showed no evidence (visual or auditory) of cues that are strongly associated with the perception of syllable-initial /t/ (e.g., closure plus burst) in the deleted forms, but VOT averaged 84 ms in the citation forms and there was a visible burst.[1] The difference in pronunciation resulted in the deleted variants being 55 ms shorter than the citation forms (655 ms vs. 710 ms). For the CB words, there were no cues for /t/ in the deleted variants, but VOT averaged 116 ms in the citation forms, which made the former shorter than the latter as well (809 vs. 872 ms). Acoustic analyses of both pronunciations showed that the stretches of speech altered by reduction were comparable across DB and CB word sets. For the deleted CB words, /n/ averaged .070 of the word's duration. For the DB words this value was .075. The relative duration of VOT in the citation forms was also similar across pronunciation bias, being .133 and .124 for the CB and DB words, respectively.

Design and procedure

With 36 target words, each spoken in a citation and deleted version, there were a total of 72 target stimuli. These were divided evenly between two stimulus lists with a different pronunciation in each. Targets were distributed across lists so that there were approximately an equal number of two- and three-syllable CB and DB words in their citation and deleted pronunciations. Along with the 174 fillers, this yielded 210 stimuli per list. Stimulus presentation was randomly ordered within each list and then hand-adjusted to avoid runs of /t/-deleted targets.

Listeners were tested up to four at a time, each in a separate sound-dampened cubicle. Stimuli were presented binaurally over headphones at a comfortable listening level. Listeners were instructed to make fast and accurate responses by pressing the appropriately labeled button ("word" or "nonsense") on a response box as soon as they had made a decision. There was a 2500 ms timeout after stimulus offset. A pause of 2000 ms preceded presentation of the next stimulus. The test session began after 14 practice trials.

Participants

Forty-eight undergraduates at Ohio State University served as listeners.

Results

The data were analyzed first by collapsing over word length. The proportion of pseudoword classifications was calculated for each listener and item. These were then averaged over items in the four conditions. The same was done with the word-response RT data after subtracting out word duration from response times, which was neces-

---

[1] The absence of these cues does not mean others (e.g., duration of the preceding vowel) were not present in the stimuli, but to the extent that they were, they failed to affect performance. A related issue is the question of whether pronunciation might be conditioned by how the word is used. For example, *counter* might refer most frequently to a surface on which to place things, with other meanings (e.g., numerical, a reversal) being less frequent. Flapping might be found only when the dominant meaning is used. Inspection of pronunciations of *counter* and *center* in the Buckeye corpus of conversational speech did not differ as a function of meaning. They were always pronounced with a nasal flap.
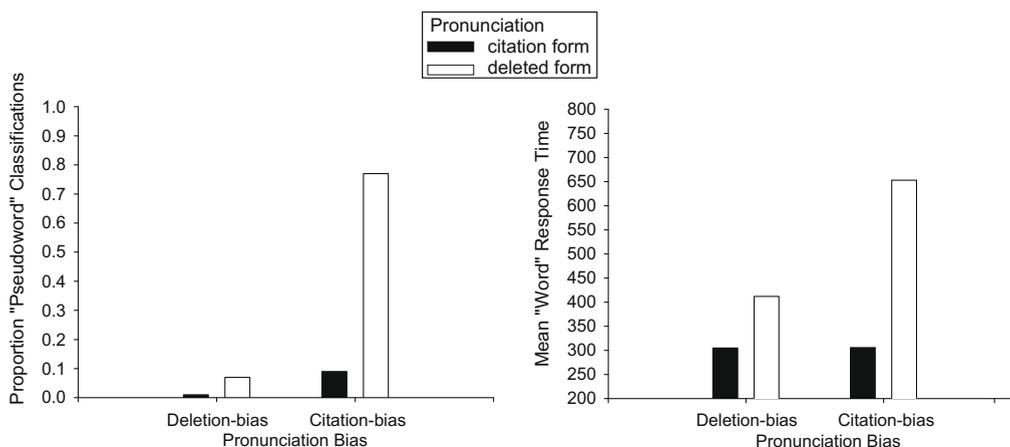
Fig. 1. Bar graphs of performance in Experiment 1. The pseudoword classification data are on the left and the mean "word" response times are on the right.

sary to control for word length differences. Item means for both measures are plotted in the top graphs of Fig. 1.

The classification data conform well with a phonological processing account, providing clear evidence that listeners are attuned to the phonological context in which t-deletion occurs. Only DB targets were classified as words regardless of pronunciation. Listeners rarely classified the deleted pronunciations of the DB words as pseudowords, although they did so slightly more often than the citation pronunciations. In contrast, the deleted pronunciations of the CB words were classified as pseudowords 77% of the time. The citation forms of these stimuli were rarely classified as pseudowords.

A 2×2 ANOVA on the subject and item classification data showed the interaction of bias and pronunciation was reliable, $F1(1,47) = 572.65$; $F2(1,32) = 69.28$; min $F'(40) = 61.8$ (all p-values were less than .05 unless noted otherwise). Planned comparisons showed that within each bias condition the increase in pseudoword responses from the citation to the deleted pronunciation was reliable (DB: $F1(1,47) = 29.97$; $F2(1,17) = 4.72$; min $F'(23) = 4.08$; CB: $F1(1,47) = 712.86$; $F2(1,17) = 101.60$; min $F'(23) = 88.93$).[2]

Despite the low pseudoword classification rates to deleted variants like *couner*, the RT data show that there is a substantial cost in processing time. RTs were 93 ms slower to the deleted than citation pronunciations of the DB words. With so few reaction times contributing to the mean in the CB deleted condition (only 23%), statistical comparisons involving this condition should be viewed cautiously. Nevertheless, for the few instances in which these pronunciations were classified as words, responses were 359 ms slower than to their citation form. The bias by pronunciation interaction was reliable by items, $F2(1,32) = 6.90$, p < .01, but not by participants, $F < 1$ (due to large individual differences). Within each bias condition, the slowdown found to deleted pronunciations was reliable for the DB words, $F1(1,47) = 49.60$, $p < .001$; $F2(1,17) = 11.90$; $p < .003$; min $F'(26) = 9.6$, and CB words,

$F1(1,47) = 5.17$, $p < .03$; $F2(1,17) = 16.73$, $p < .001$; min $F'(64) = 3.95$, p < .06. Note also that RTs to the deleted pronunciations of the DB words were much faster than RTs to the deleted pronunciations of the CB words (by 246 ms), showing that even when both were heard as words, the former were processed far more efficiently than the latter.

Much the same pattern of results is found when the data are broken down by word length, but in addition, there is an effect of length, with fewer nonword classifications and faster RTs to the three-syllable than two-syllable items. Although word length effects were not large enough to yield statistically reliable differences in the classification data, substantial differences were found in the RT data. For the CB words, there was a main effect of word length, $F2(1,16) = 5.15$, $p < .04$, with responses to three-syllable words being fastest, and a main effect of pronunciation, $F2(1,16) = 15.89$, $p < .001$, with the citation pronunciations being responded to faster than the deleted pronunciations. For the DB words, there was a reliable interaction between pronunciation and word length, $F2(1,16) = 4.14$, $p < .05$, which was due to the RT difference between the deleted and citation pronunciations being larger for the two-syllable than three-syllable words.

Discussion

The data are consistent with the proposal that variant recognition is aided by a phonological inference mechanism. /t/-deleted variants were classified as words only in the phonological environment where deletion frequently occurs in production. Otherwise listeners generated an overwhelming number of nonword responses. The RT data in the DB condition replicate results showing a slowdown in processing the variant form relative to its citation form. This outcome was obtained with both two-syllable and three-syllable words.

Such selective sensitivity to the phonological environment has also been demonstrated in a cross-modal priming experiment (Ranbom & Connine, 2007; Experiment 3). Although the purpose of their experiment was different, the conditions were almost identical. Just as in the current

---

[2] The same outcomes are obtained when analyses were performed on logistic transformations of the data.

experiment, words with medial /nt/ clusters were used in the DB condition, where listeners heard the citation and deleted variants as primes and had to make a lexical decision to the word's written form. In the CB condition, words with other medial clusters (e.g., sp, ld, mp) were used, but unlike in the current experiment, most had a reduced vowel in the second syllable. Despite these small differences, the results were qualitatively the same. For the DB words, the deleted variant produced less priming compared to the citation form. Importantly, for the CB words, the deleted form showed no evidence of priming.

If the current results reflect the operation of a phonological process, then they should replicate with newly learned words. This prediction was tested in Experiments 3 through 5 using the Ganong (1980) paradigm. To provide continuity across paradigms, it was first necessary to demonstrate that the results of Experiment 1 replicate with the Ganong paradigm as well.

Experiment 2

The Ganong paradigm was used because it provides a means of comparing the magnitude of lexical activation generated by citation and variant pronunciations. In the task, listeners categorize word-initial phonemes on a phonetic continuum as being one of two possibilities (e.g., /k/ or /p/). Each step is prepended to a context (e.g., ounter) that forms a word at one endpoint and a pseudoword at the other (e.g., counter–pounter). When listeners are presented a step from the middle of the continuum, which is perceptually ambiguous, they are biased to label the segment in a lexically consistent manner (e.g., responding /k/ given ounter).

The magnitude of this labeling bias can serve as a measure of the strength of lexical activation (Pitt, in press; Pitt & Samuel, 2006). A larger bias implies stronger activation because lexical influences are the cause of the bias. Quantitatively, the lexical effect is measured relative to another context (e.g., owder) that biases labeling toward the other continuum endpoint (e.g., /p/). If this additional context is held constant across pronunciations (e.g., citation vs. deleted), it functions as a reference from which to measure changes in the strength of lexical activation. The size of the lexical effect is referred to as a *lexical shift* because of how the categorization functions shift apart in the middle of the phonetic continuum (see Fig. 2).

In the experiment, the sizes of lexical shifts produced by deleted and citation pronunciations of a DB and a CB word were compared. If the results replicate what was found in Experiment 1, lexical shifts should be found for both pronunciations of the DB word, but only the citation form should yield a shift for the CB word. Pitt (in press) studied the DB conditions alone, and found both pronunciations produced lexical shifts, but that for the deleted pronunciation was smaller. A similar finding was expected here. Categorization response times were also expected to replicate the RT data of Experiment 1, with slower responses when hearing a deleted than citation form.

Method

Stimuli

The same two variables in Experiment 1 were crossed, whether the target word was a member of the class of words in which medial /t/ frequently or rarely deletes (deletion-biased vs. citation-biased) and how the medial /
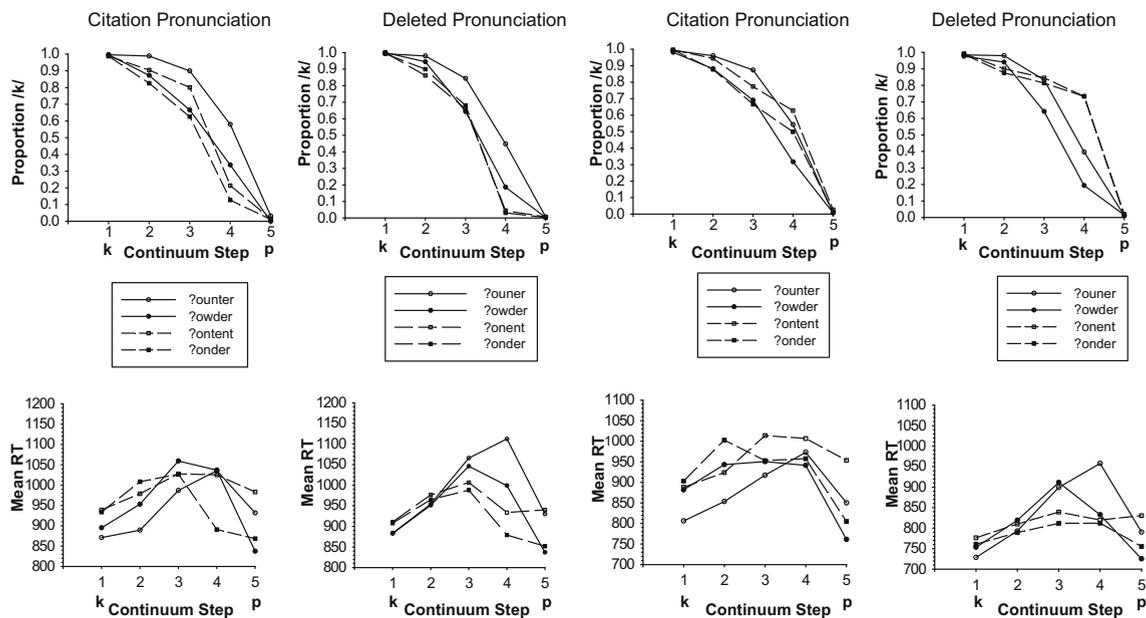

Fig. 2. Mean labeling and RT functions for the citation pronunciations (first column of graphs) and deleted pronunciations (second column of graphs) of the deletion-biased (dashed lines) and citation-biased (solid lines) words. The graphs on the left half of the page are from Experiment 2, and those on the right half are from a replication.

t/ was pronounced ([t] vs. nasal flap). Of the words used in Experiment 1, the best pair to use was *counter* and *content*. They are the most similar phonologically and lend themselves to creating word-nonword continua by morphing the initial /k/ to /p/.

A corresponding nonword-word continuum is necessary to serve as a reference against which to measure lexical influences. The two words chosen for this purpose were *powder* for *counter*, and *ponder* for *content*.

Stimuli were recorded as in Experiment 1, including new tokens of *counter* and *content*. Analyses of their citation and deleted pronunciations were performed to ensure the acoustic properties of the realizations were as intended. The citation pronunciations contained a stop closure (averaging 36 ms) after the nasal followed by a stop burst plus aspiration (49 ms) prior to vowel onset. The deleted pronunciations contained only nasal murmur in this same region.

Word-initial /p/–/k/ continua were created by blending clear tokens of each phoneme in various proportions (Pitt & McQueen, 1998). Tokens of /k/ and /p/, excised from instances of *counter* and *pounter*, were identified whose VOTs were similar in duration (93 ms). They were then excised from the words immediately before the first pitch period of the vowel, blended in 21 ratios (steps of .05), and then prepended to tokens of all three contexts (*ounter, ouner, owder*) to create the *counter* test set: *counter-pounter, couner-pouner, cowder-powder*. Because /t/ lies between /p/ and /k/ on a place-of-articulation continuum, one might wonder whether the ambiguous steps sounded like /t/. None did, nor did listeners ever report hearing utterances that began with /t/ in post-experiment questionnaires.

A pilot experiment was conducted to identify steps in the middle of the /p/–/k/ continuum at which lexical influences were substantial. Ten listeners heard twelve presentations of the endpoints (steps 1 and 21) of each continuum plus seven steps taken from the middle of the continuum. Labeling functions were created from the averaged participant data. Three middle steps that yielded large differences in labeling between the two contexts (i.e., a lexical shift) and yielded this shift over a different range of the dependent measure (e.g., 0.6–0.9; 0.4–.07; 0.2–0.5) were selected. This last requirement meant that the middle steps were never adjacent steps on the 21-step continuum, but were separated by at least one step. The three middle steps plus the two endpoints made up the five steps on each of the three continua of the *counter* test set.

The steps described in the preceding two paragraphs were repeated to create the *content* test set (*content-pontent, conent-ponent, conder-ponder*). This was necessary because the initial vowels in *content* and *ponder* are different from that in *counter*.

Design

How the target word was pronounced (citation vs. deleted) was manipulated between participants and the pronunciation bias (DB vs. CB) was manipulated within participants. This design was chosen to guard against the citation form biasing how listeners would respond to the corresponding deleted variant if they were presented together.

Procedure

The equipment was the same as in Experiment 1. Two buttons on a response box situated in front of participants were labeled with the letters that correspond to the two initial phonemes (e.g., *k* and *p*). After listening to instructions to respond quickly yet accurately, participants sat through a 14-trial practice session before proceeding to the test session. Each step on the four continua was presented 12 times (randomized within blocks of 20 trials), for a total of 240 trials. One group heard the *counter-pounter, cowder-powder, content-pontent,* and *conder-ponder* continua (citation-pronunciation condition). The other heard the *couner-pouner, cowder-powder, conent-ponent,* and *conder-ponder* continua (deleted-pronunciation condition). On each trial listeners were given 2500 ms to respond after stimulus offset. Once all listeners had responded, there was a 1500 ms pause before the next trial. After the experiment, listeners were given a surprise recall test in which they were asked to write down all of the utterances (words and pseudowords) heard during the experiment.

Participants

Forty-five participants served in the experiment, 21 in the citation-pronunciation condition and 24 in the deleted-pronunciation condition.

Results

The proportion of /k/ responses at each step on each continuum was calculated for each participant, and then averaged across listeners to yield mean labeling functions, which are plotted in the graphs on the left half of Fig. 2. Data from the group who heard the citation pronunciations are in the first column. Those from the group who heard the deleted pronunciations are in the second column. The CB functions are plotted with dashed lines and the DB functions are plotted with solid lines.

Lexical influences on phoneme categorization are present when the labeling functions diverge in opposite directions to make lexically consistent responses (e.g., /k/ given *ounter* and /p/ given *owder*) most frequent. This lexical shift was quantified by calculating the mean difference in labeling between the three middle steps on the citation (or deleted) continuum and the reference continuum (Pitt & Samuel, 1993).

Comparison of the labeling functions within and across graphs shows that the citation and deleted pronunciations of the DB word produced lexical shifts, but for the CB words, only the citation pronunciation produced a shift. For the DB words, the size of the shift was .20 for *counter*, $t(20) = 4.45$ and .17 for *couner*, $t(23) = 3.78$. For the CB words, *content* yielded a .11 shift, $t(20) = 4.2$ and *conent* a slight reversal ($-.02$). The drop in shift size from the citation to the deleted pronunciation was reliable for the CB words (.13; $t(40) = 6.67$) but not for the DB words (.03).

The RT data were aggregated following the same method as the labeling data. For each participant, response

times at each continuum step were averaged separately for each of the two continua of a test set. Individual participant data were then combined and the mean RT functions for each context are graphed in the bottom row of Fig. 2.

Evidence of lexical influences are most often largest at the endpoints of the continuum because the lexically consistent context reinforces interpretation of the endpoint step for one continuum but conflicts with it for the other (e.g., /k/ given *ounter* vs. *owder*; Connine & Clifton, 1987). Lexical effects are quantified by taking the difference in RT across endpoint steps. For the k-biased context, this was steps 1–3. For the /p/-biased, it was steps 4 and 5.

Except for the *ounter* context, lexical effects emerged only at the /p/ endpoint. In the left graph (citation pronunciations), /k/ responses were 53 ms faster when the context was *ounter* than *owder*, $t(20) = 2.04$. On the /p/ side of the continuum the two functions reverse, as they should, but the RT speed up to *?owder* was not reliable because it is present at only step 5, not step 4. For *?ontent* and *?onder*, there is a large (125 ms) and reliable advantage for *?onder* at the /p/ endpoint (125 ms, $t(20) = 5.30$), but the functions are quite similar at the /k/ endpoint. In the right graph (deleted pronunciations), RT differences are present only at the /p/ endpoint. In the DB condition, responses averaged 103 ms faster given *?owder* than *?ouner*, $t(23) = 4.47$. In the CB condition, the lexical RT advantage to *?onder* over *?onent* was somewhat smaller than this (72 ms), but still reliable, $t(23) = 4.68$.

Because inferences from these data hinge in part on a null lexical shift with *?onent*, it is worthwhile to demonstrate the reliability of the results. Experiment 2 was therefore rerun. The only change was that the three middle steps (2–4) on the *?ouner* and *?onent* continua were replaced with three adjacent steps from the middle of the 21-step /p/–/k/ continuum. The purpose of this change was to increase the opportunity for lexical influences to emerge.

The data from this replication are graphed in the right half of Fig. 2. The methodological change had no effect on the outcome. Lexical shifts were obtained again in the labeling data for the citation pronunciations of the CB and DB words (column 3), but only the deleted pronunciation of the DB word produced a shift (column 4). Lexical shifts were reliable in all comparisons except for *?onent* (*?ounter*: .16; $t(20) = 4.95$; *?ontent*: .10; $t(20) = 3.27$; *?ouner*: .14; $t(20) = 4.79$; *?onent*: .02; $t(20) = .87$, $p < .40$). The two-way ANOVA on biased pronunciation and /t/ realization yielded a main effect of bias, $F(1,80) = 7.18$, but not one of realization, $F(1,80) = 2.14$, $p < .15$. The interaction of bias by realization was reliable, $F(1,80) = 4.2$. Comparisons within each pronunciation condition showed that the .02 difference between *?ounter* and *?ouner* was not reliable($t < 1$), but the .08 difference between *?ontent* and *?onent* was, $t(20) = 2.18$.

In the RT data, lexical effects are evident with the two citation pronunciations. There was a 66 ms RT advantage for the *?ounter* context on the /k/ side of the continuum, $t(20) = 4.13$. On the /p/ side, responses were 61 ms faster for *?owder*, $t(20) = 2.51$. For *?ontent*, there was a 47 ms lexical effect on the /k/ side, $t(20) = 2.04$, $p < .06$, and a 86 ms advantage for *?onder* on the /p/ side, $t(20) = 5.09$.

For the two deleted pronunciations, differences between functions are even more clear-cut in the replication.

For *?ouner*, there is a slight (21 ms), though nonsignificant, lexical effect at the /k/ endpoint. In contrast, the reversal of the two functions is quite robust on the /p/ side (95 ms; $t(20) = 4.60$). For *?onent*, the data are qualitatively different. The *?onent* function sits above the *?onder* function by a small but relatively fixed amount from steps 1–4. Only at step 5 (/p/ endpoint) do the functions begin to separate and show an expected lexical advantage for *?onder*. Statistical comparisons on both endpoints were nonsignificant. Note that this same trend is visible in the data from the first testing (column 2), and is what would be expected if listeners heard both stimuli at the /k/ endpoint (*conent* and *conder*) as pseudowords (i.e., no lexical effect) and heard only p*onder* as a word at the /p/ endpoint.

Discussion

As in Experiment 1, the results of Experiment 2 clearly show that lexical activation is highly selective, even to the point of treating similar words differently when they undergo the same form of reduction. The pattern of results across conditions is what should be found if the phonological environment triggered application of a process that inferred the presence of a medial /t/. *Counter*, *couner*, and *content* generated lexical shifts, but *conent* did not.

After the experiment, participants completed a questionnaire in which they were asked to recall the stimuli. Analyses of the responses showed further how the deleted variants of the DB and CB words were perceived very differently. *Couner* was recalled 100% of the time, and was spelled as *counter* 80% of the time, with the medial "t" restored in its written form, even though these listeners only ever heard the deleted variant. In contrast, *conent* was not only recalled much less often (only 29% of the time), but it was always classified as a pseudoword, as in Experiment 1.

The results of Experiments 1 and 2 establish the conditions necessary for phonological inference to be a plausible mechanism by which pronunciation variants are recognized. Across stimuli and paradigms, phonologically viable environments led to perception of the variant as a word and robust lexical activation, whereas a phonologically unviable context yielded a nonword percept and no evidence of lexical activation.

Experiment 3a

The power and appeal of an inference mechanism is its ability to generalize. Presumably it develops by encoding across many words the commonalities defined by the environment in which the specific type of phonological variation occurs. This knowledge can then be applied readily to future encounters with the same environment to aid word recognition. If an inference process is sufficient for recognizing pronunciation variants, then the results obtained in the DB condition of Experiment 2 should generalize to a newly learned word whose phonological structure makes medial /t/-deletion predictable, and therefore restorable. That is, not only should the citation form of the word generate a lexical shift, but its /t/-deleted variant should as well.

This prediction was tested in Experiment 3a using a learning paradigm. The experiment began with a short exposure phase that was modeled after one developed by Gaskell and Dumay (2003) and Dumay & Gaskell, 2007). They had listeners monitor for phonemes in novel words (e.g., cathedruke) repeatedly in an exposure session. In a test session one week later, using pause detection, they found that those novel words now showed evidence of lexicalization. Pauses embedded in phonologically related words (e.g., cathedral) were detected much more slowly than in control items, indicating that the novel word had become an effective lexical competitor, slowing response times.

An attractive property of this methodology is that only the novel word's phonological form is learned. Semantic and syntactic knowledge are not required to learn the word, although Leach and Samuel (2007) show that it can dramatically improve lexicalization. Brief exposure paradigms like this have proven similarly successful in studying listeners' sensitivities to the statistical properties of language (Magnuson, Tanenhaus, Aslin, & Dahan, 2003).

In Experiment 3, the exposure phase and test phase were separated by seven days. During the exposure phase, listeners heard the citation pronunciation of a new word. During the test phase (Ganong paradigm), the pronunciation of the newly learned word varied across conditions. In one condition, it was pronounced in its citation form. In the other, the deleted pronunciation was presented. If a lexical shift is found with the citation form, the results demonstrate that a lexical representation of the novel word was formed. If a lexical shift is also found with the deleted form, the data would demonstrate generalization, and be convincing evidence of phonological inference at work.

Method

Design

There were two variables, both with two levels and both manipulated between subjects. One was whether listeners participated in an exposure phase in which they heard the canonical pronunciation of a novel word (senty or surnty), and the other was whether, in the test phase using the Ganong paradigm, listeners heard the canonical form of the novel word (senty or surnty) or its /t/-deleted variant (seny or surny). Two sets of stimuli were used rather than one to ensure results replicated. The choice of context (enty and urnty) was dictated by the requirements of the Ganong paradigm.

Stimulus continua

Stimulus creation followed the procedure described in Experiment 2. For the enty test set, three contexts (/enti/, /eni/, /eti/) were appended to steps from a word-initial /s/– ʃ/ continuum to create three continua (senty-shenty, seny-sheny, setty-shetty). enty was the canonical context, eny was the deleted context, and etty served as the reference context. The phonological environment /nti/ was chosen because /t/ deletion occurs often in this environment and it occurs in a number of words (e.g., twenty, plenty, county, bounty), some of which were used in Experiment 1.

The /s/–/ʃ/ continuum was created by blending tokens of each fricative, spliced from tokens of /eti/, in 21 proportions.[3] Three ambiguous steps were identified from the results of a brief identification test in which listeners labeled all steps prepended to the vowel /e/. Together with the two endpoints they formed the five-step continuum. The fricative was 170 ms in duration and the durations of the contexts were as follows: enty, 524 ms; eny, 471 ms; ety, 447 ms.

The preceding continua creation procedure was repeated for the urnty test set (/enti/, /eni/, /eki/). urnty was the canonical context, urny the deleted context, and urky served as the reference context. A second /s/–/ʃ/ continuum had to be created because the vowel following the fricative was different. Three ambiguous steps were again identified in a pilot experiment in which listeners labeled all steps on the fricative continuum. The fricative was 206 ms in duration and the durations of the contexts were as follows: urnty, 570 ms; urny, 508 ms; urky, 530 ms. The test sets were manipulated between participants.

Acoustic analysis of the tokens showed the "citation" pronunciations had a stop closure followed by a burst release, whereas the /t/-deleted forms contained only nasalization during this interval. In the token of ety, /t/ had a stop closure and burst release.

Procedure

The testing procedure differed depending on whether listeners were in the exposure or no-exposure condition. In the exposure condition, testing took place over two days, separated by one week. On day 1 (exposure phase) participants were familiarized with the "citation" pronunciation. The session began with a 48-item lexical decision experiment to confirm that listeners classified the citation form as a nonword. Half of the items were words and they were equally likely to be one, two, or three syllables in length. Responses buttons were labeled "word" and "nonword". Participants were instructed to respond accurately and quickly. After stimulus presentation, there was a 1500 ms timeout. A 2000 ms pause preceded the next trial. There were 16 practice trials.

The second part of the exposure phase consisted of a phoneme monitoring experiment. Its purpose was to familiarize participants with the citation form without explicitly asking them to learn the word. Eight two-syllable nonwords, one of which was the target word (senty or surnty, depending on test set), were presented in a randomized order in each of 28 blocks of trials. The to-be-monitored phoneme, whose corresponding letter was specified visually on a computer monitor in front of participants, was constant across trials within a block, and occurred in no more than half of the stimuli. There were seven phonemes (/d, k, l, n, r, s, t/), with each being specified four times. They were chosen because they constituted the majority of consonants across words and occurred in

---

[3] Although it would have been desirable to use a /p/-/k/ continuum to maintain continuity with Experiment 2, I decided to switch to a /s/-/ʃ/ continuum because the perceptual ambiguity of the fricative induces large lexical shifts (Pitt, in press). If generalization occurs, the setup must be sufficiently sensitive to detect it.

multiple positions across words. In addition, the three consonants in the target word were included to ensure that listeners actively listened to its pronunciation.

The letter corresponding to the target phoneme was displayed continuously during a block of trials. Listeners were instructed to press a single button rapidly upon detecting the target phoneme. A 2500 ms timeout followed auditory presentation of the stimulus. There was then a 1500 ms pause before the next trial. The experiment began with eight practice trials.

In the test phase (day 8), participants sat through what amounted to a shortened version of Experiment 2. Participants heard either the citation continuum (e.g., *senty-shenty*) or the deleted continuum (e.g., *seny-sheny*). The reference continuum (e.g., *sety-shetty*) was the same across both conditions. Each step was presented 12 times, for a total of 240 trials across the two continua. Other procedural details were the same as in Experiment 2.

In the no-exposure condition, day 1 was omitted and participants completed only the test phase.

Participants

Participants were recruited from the same pool as Experiment 1. A total of 86 served across the four conditions of the *enty* test set, and 86 across the conditions of the *urnty* test set. For each test set, approximately half of the participants heard the citation continuum during the test phase, half with exposure and half without. The remainder heard the deleted continuum during the test phase, again divided equally between the two exposure conditions.

Results and discussion

Analysis began by assessing accuracy in the exposure phase (day 1). Performance was high across all groups of listeners. In the lexical decision task, mean accuracy was never lower than 96% in any condition, with never more than two participants (4 total) categorizing the target stimulus (*senty* or *surnty*) as a word in any condition. Participants considered these items nonwords upon first encounter.[4] Mean accuracy in phoneme monitoring was never lower than 94%, indicating that participants actively listened to the stimuli.

The labeling data were aggregated and analyzed as described in Experiment 2. Shown in the top row of Fig. 3 are the mean labeling and RT data for the participants who were exposed to *enty* and tested on the *enty-etty* continuum. The results suggest that the exposure phase succeeded in inducing listeners to process *senty* as a word. The frequency of /s/ responses was higher in the *enty* context than that in the *ety* context, particularly in the middle of the continuum. This .12 lexical shift is reliable, $t(19) = 3.62$. The size of the shift is particularly impressive because the reference function is a nonword-nonword continuum (*setty-shetty*), so the shift is due solely to the *enty* context. The RT data also show evidence of *senty* being lex-

icalized. Response times at the /ʃ/ endpoint were on average 77 ms slower in the *enty* than *ety* contexts, $t(19) = 2.99$. Responses were slowed because /ʃ/ was inconsistent with the bias induced by *enty* toward /s/.

The smoothness of these functions and the locations of where they diverge, when this occurred in the data, are representative of what was found across the other conditions in the experiment. Because of the large number of graphs that would be needed to display all of the data (16 total), it seemed prudent to switch to a presentation format that preserved the information of most interest but condensed the space required to do so. This was done by using bar graphs to display the sizes of the mean labeling and RT shifts. Those for the *enty* test set are shown in the middle row of Fig. 3. Those for the *urnty* test set are in the bottom row. The data are grouped by the type of exposure participants received (bar shading) and whether the citation or deleted test continua were presented on day 8.[5]

The left-most pair of bars in the middle and lower graphs demonstrate that exposure to the citation pronunciation is required for that word to generate a lexical shift. Without exposure (unfilled bars), a minuscule shift was produced, which in neither case was statistically reliable (?*enty*: −.01; ?*urnty*: .02). With exposure (hashed bars), large and reliable shifts were obtained (?*enty*: .12, $t(19) = 3.62$, as described above; ?*urnty*: .11; $t(20) = 3.16$). The increase in shift size from the no-exposure to the exposure conditions was reliable for both contexts, ?*enty*: $t(38) = 2.59$; ?*urnty*: $t(40) = 3.04$. These data show that the exposure session was sufficient to induce lexicalization, a precondition for assessing generalization.

The right-most pair of bars in each graph are the data from the conditions in which generalization to the /t/-deleted variant was assessed. They both show that after exposure to the citation pronunciation, there was no generalization to the /t/-deleted variant. The labeling shifts in the exposure condition are small and comparable in size to those in the no-exposure conditions (?*eny*: no-exposure; −.02; exposure = .02; ?*urny*: no-exposure; −.04; exposure = .002). Differences between the exposure and no-exposure conditions were not statistically reliable for either context.

The RT data are graphed on the right-hand side of Fig. 3. The bars represent the mean difference in response time found over steps 4 and 5, with positive values indicating a slowdown in the test context (e.g., *enty*) relative to the control context (e.g., *ety*). Analyses focused on only these steps because, just as is shown in the RT graph at the top of the Figure, RTs across steps 1–3 overlapped each other across the two continua, and were thus minimally informative. The two cases (of eight) in which this was not true are mentioned below.

The RT data parallel the labeling data in showing no generalization to the pronunciation variant. The left-most pair of bars again shows that exposure during the phoneme monitoring session led to lexicalization. For the *enty*

---

[4] Exclusion of the data from these listeners did not alter the results, so their data were retained in the analyses.

[5] The labeling and RT graphs for all of the conditions in Fig. 3 are available on the publications page of the author's web site.
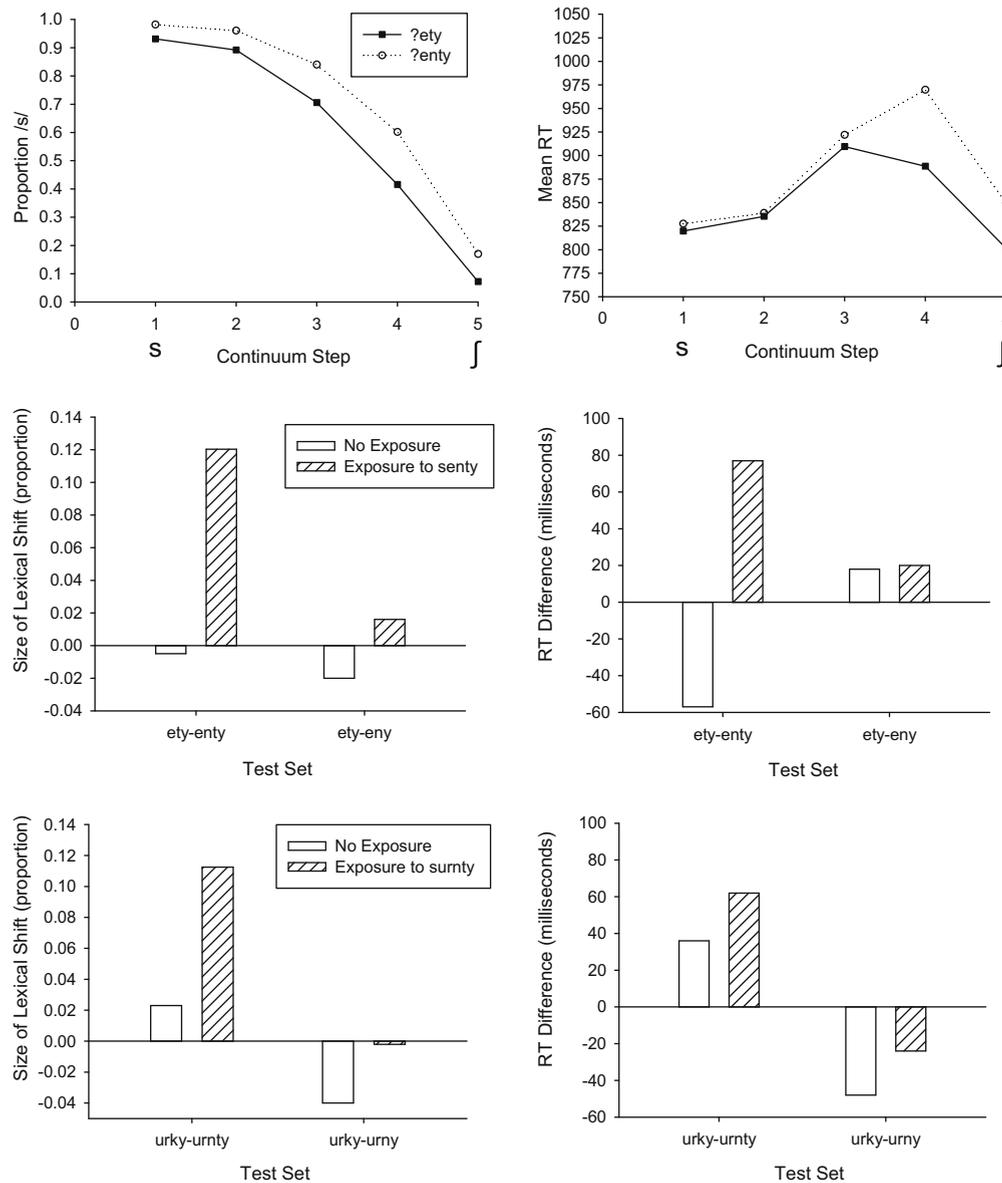
Fig. 3. Data from Experiment 3. The top two graphs represent prototypical labeling and RT functions when exposure on day 1 transfers to the test session on day 8. The middle and bottom rows contain graphs depicting the sizes of the lexical shifts across conditions in Experiment 3. The middle row contains data using the *enty* test set and its variants and the bottom row data using the *urnty* test set and its variants.

test set, response times to the initial fricative in the *enty* context were 77 ms slower than in the *ety* context, $t(19) = 2.99$. In the corresponding no-exposure condition, the −57 ms difference was also statistically reliable, $t(22) = 3.26$, but as can be seen in the graph, the effect was in the opposite direction, with response times being faster to the citation than control items. This RT advantage was obtained across all five steps of the /s/–/ʃ/ continuum and was virtually constant, with the mean difference being −51 ms. For the *urnty* test set, a large slowdown (60 ms) was also obtained in the exposure condition, $t(20) = 3.30$. Without exposure, a slowdown of about half that amount (36 ms) was obtained, $t(20) = 2.07$, p < .052, but as with

the no-exposure condition using the *enty* test set, this difference was not restricted to steps 4 and 5 of the continuum, but was constant across all steps (mean of 38 ms).

It is unclear why listeners in these no-exposure conditions responded faster to the items on one continuum than another. This same pattern was found in Experiment 2 (second and fourth columns in Fig. 2). What is common across these four conditions is that no lexical shift was obtained in the labeling data. In the absence of lexical influences, listeners sometimes display a RT bias for one continuum.

The right-hand pair of bars again represent data from the test of generalization, where listeners were exposed

to the citation form (e.g., *enty*) and then tested on the /t/-deleted pronunciation (e.g., *eny*). As with the corresponding labeling data, evidence of generalization is nonexistent. Overall, RT differences were much smaller than those found when listeners labeled the citation pronunciations. For *?eny*, there was a 20 ms slowdown at the /ʃ/ endpoint after exposure and an equivalent slowdown (18 ms) with no-exposure. Neither difference was statistically reliable. For *?urny*, reversals were obtained. With exposure, there was a −24 ms reversal, which approached significance, $t(21) = 1.75$, $p < .09$. With no-exposure there was an even larger reversal of −48 ms, $t(21) = 2.59$, $p < .02$.

Discussion of these data will be postponed until after presenting data from one additional condition, which was run to discount an uninteresting cause of the null lexical shifts found with the /t/-deleted continua (*seny-sheny*, *surny-shurny*).

Experiment 3b

The failure to find reliable shifts with the /t/-deleted contexts might make one wonder whether there is something unusual about these stimuli that would prevent them from producing lexical shifts under any circumstances. If this were the case, then any conclusions about variant processing would be misleading. To address this concern, I reran the exposure condition for the *seny* test set, but instead of presenting the citation pronunciation during the exposure phase, the /t/-deleted pronunciation was presented instead. If the stimuli are problematic, then a null effect should be found again. Otherwise a lexical shift should be obtained because, just as with the citation pronunciations in Experiment 3a, listeners will have learned the /t/-deleted variant as a new word.

Method

The experiment was identical to Experiment 3a except that only the *eny* test set was used. Participants were exposed to the /t/-deleted variant on day 1 and tested on the corresponding /t/-deleted continuum on day 8. Twenty-five listeners were tested.

Results

The labeling and RT data were analyzed using the same method as in Experiment 3a. There was a healthy .076 shift in the labeling data, $t(24) = 2.73$, and a 58 ms slowdown in the RT data, $t(24) = 2.96$. These results demonstrate that there is nothing unusual about the /t/-deleted variants that prevented listeners from forming a lexical representation of them. With exposure, they are lexicalized just like their citation counterparts.

Discussion

The results of Experiment 3 identify a situation in which phonological inference fails. If inference were sufficient for variant recognition under the current circumstances, generalization from the citation to the deleted pronunciation should have been found. This did not happen for either *seny* or *surny*.

These results are at odds with studies on regressive place assimilation that have shown generalization (Gaskell & Marslen-Wilson, 1998; Mitterer & Blomert, 2003). One reason for the different outcome could be differences in the types of pronunciation variation. When a word-final segment assimilates, there is often information in the acoustic signal corresponding to its surface realization, although its clarity can vary tremendously (Dilley & Pitt, 2007; Nolan, 1992). This information in part defines an assimilable environment. In contrast, with medial /t/ deletion, the severity of the reduction itself may leave too little, if any, acoustic information to ascertain whether /t/ should be restored. Related to these differences in the type of variation is their location in the word, medial vs final. It may be that deletion is more complete word-medially, and phonological inference is insufficient to rectify some types of variation in this location. In this regard, it should be remembered that the theory was originally developed in the context of word-final variation, so the present study can be viewed as testing its bounds.

Lexical processes could compensate for the limitation of phonological inference in this situation. With medial /t/ deletion, inference may not be possible unless listeners have encoded the variant. If this is the case, then the question arises as to what is the nature of lexical memory for encoding variants? Representational accounts of variant recognition offer a few possibilities.

The key to generalization lies in ensuring lexical memory can accommodate variation. Underspecification theory (Lahiri & Reetz, 2003) achieves generalization by having underspecified lexical representations. Only marked phonological features of a word are stored in a word's lexical entry. Unmarked features are not, making listeners insensitive to their variation. The place of articulation of /t/ is unmarked, permitting generalization to stop consonants with other places of articulation (e.g., /k/, /p/). Thus, pronunciation variants such as [senki] should be correctly recognized as *senty*, assuming the former is not a word. However, the complete deletion of /t/, as in the current experiment, causes the marked feature [-continuant] (i.e., stop consonant) to be absent as well, which would prevent generalization of *seny* to *senty*, as was found in Experiment 3. Underspecification theory correctly predicts the current results.

The more popular solution to the problem of generalization is to encode pronunciation variation directly in memory (Johnson, 2006; Pierrehumbert, 2001; Ranbom & Connine, 2007). Although the current data cannot discriminate among these alternative proposals, they do speak to one issue that is common to all three – the role of variant exposure in representation formation and updating. The data of Experiment 3 can be probed to determine whether brief exposure to the variant was sufficient to produce early signs of its lexicalization.

If immediate exposure alone were sufficient for generalization, then at a minimum, small lexical shifts should have been found when listeners were exposed to the citation pronunciation (e.g., *senty*) on day 1, and tested on the deleted variant (e.g., *seny*) on day 8. Listeners first heard

the deleted variant during the practice trials on day 8, yet that exposure did not affect labeling in the test session to yield a lexical shift. Even if one assumes that the effects of exposure are not instantaneous or that repeated exposure is required, the evidence from Experiment 3 is not encouraging. Analyses of data from the last half of the labeling session on day 8, where a shift is most likely to be found, showed not even a trend in the predicted direction (?eny = .01; ?urny = −.03). After exposure, a memory consolidation period on the order of one day or one week might be required to encode the variant in memory.

In addition to exposure, another requirement for generalization might be to associate the variant with its citation pronunciation. Otherwise the variant might have been processed as a different word. Listeners hear pronunciation variants most often in conversations, where the surrounding sentence and discourse aid in interpreting reductions in speech. Listeners will make inferences, implicitly or explicitly, about the intended word if its identity is not clear. This inference should contribute to lexicalization of a variant because of the association made between the two pronunciations; they become lexically isomorphic. In Experiment 3, instructions to participants on day 8 made no mention of the fact that one stimulus was in any way related to one from day 1. Given that one week transpired between the two testing sessions and the /t/-deleted variant was one of four endpoint stimuli presented on day 8, this association may not have been obvious to participants. If this hypothesis is correct, then generalization should be found when the association is established prior to the test phase, which is what was tested in the next experiment.

Experiment 4

Except for a small methodological change in the procedure, Experiment 4 was identical to Experiment 3. Prior to the start of the labeling session on day 8, listeners heard a few tokens of the /t/-deleted variant (seny or surny) in a dialog between experimenters. They mention that a word from day 1 will be presented again on day 8, but the experimenters pronounce the word only in its /t/-deleted form. It was up to participants to infer the association between the two pronunciations across days. The experimenters never spoke the citation form or discussed how the pronunciations differed. If this slight modification causes listeners to process the deleted variant similarly to its citation form, then a lexical shift should be found.

Method

Two conditions from Experiment 3a were used. In one there was exposure to senty with testing on seny, and in the other there was exposure to surnty with testing on surny. The only modification to the procedure was that at the start of the experimental session on day 8, listeners heard a short verbal exchange between two experimenters about the experiment. The dialog, which was prerecorded (the text is listed in the Appendix A), contained five occurrences of the /t/-deleted variant, which referred to the name of the experiment. Listeners heard the dialog over headphones just prior to the start of the practice session. It was not unusual for them to hear some of the instructions over the headphones because they also experienced it on day 1. A microphone on a short stand was in plain view as the participants walked into the testing cubicles. The dialog was recorded in the control room housing the audio and computer equipment, complete with doors opening and other sounds, to give the impression of a real conversation taking place at that moment in time, upon which participants were eavesdropping. The five tokens of the variant, which were not copies of a single token, were spoken by a talker different from the one who recorded the materials for the experiment. Twenty-one participants heard the enty test set and 24 the urnty test set.

Results and discussion

The data from both test sets are shown in the top half of Fig. 4, with the ?eny data above the ?urny data. For ?eny, a robust .11 lexical shift was obtained in the labeling data, t(20) = 2.11. The RT data show a similar pattern, with responses across steps 4 and 5 being 42 ms slower in the ?eny than ety context, t(20) = 3.25. Further evidence of the lexical influences can be found at the /s/ endpoint (steps 1 and 2), where response times were faster in the ?eny than ?ety context by 30 ms, t(20) = 1.92, p < .07. For ?urny, a similarly-sized lexical shift (.09) was obtained in the labeling data, t(23) = 2.75. In the RT data, the pattern is the same as that found in the ?eny graph above it, only the trends are too weak at both endpoints to yield statistically reliable differences. Except for this one null effect, the results across continua are consistent in showing that inclusion of the dialog at the start of day 8 induced generalization to the /t/-deleted variant.

To appreciate the effect that hearing the dialog had on responding, the graphs in the bottom half of Fig. 4 are from the identical conditions in Experiment 3a, except that there was no dialog before the test session on day 8. The ?eny data are again above the ?urny data. Comparison across graphs, primarily the labeling data, makes it evident that incidental exposure to the variant in the dialog led to generalization. This claim is reinforced by statistical comparisons across experiments. The increase in shift size from Experiment 3a to 4 was marginally reliable for ?eny (.09), t(42) = 1.79, p < .08, and slightly more robust (.13) for ?urny, t(44) = 2.43. Neither analysis reached statistical significance in the RT data.

It is reasonable to wonder whether the dialog itself, without exposure to the citation form on day 1, is sufficient to produce the data in Fig. 4. That is, perhaps from just listening to the dialog, lexical representations of seny and surny formed. To test this hypothesis, a control experiment was run in which a group of 20 new listeners participated in day 8 only, without ever hearing the citation pronunciation, senty. Only the seny test set was used. A few words were changed in the dialog so that it did not refer to a past experimental session. The lexical shift was minuscule (−.01) and in the wrong direction. The RT data showed no biases either. As in the control conditions of Experiment 3a (white bars), without exposure on day 1, seny was processed as a nonword.
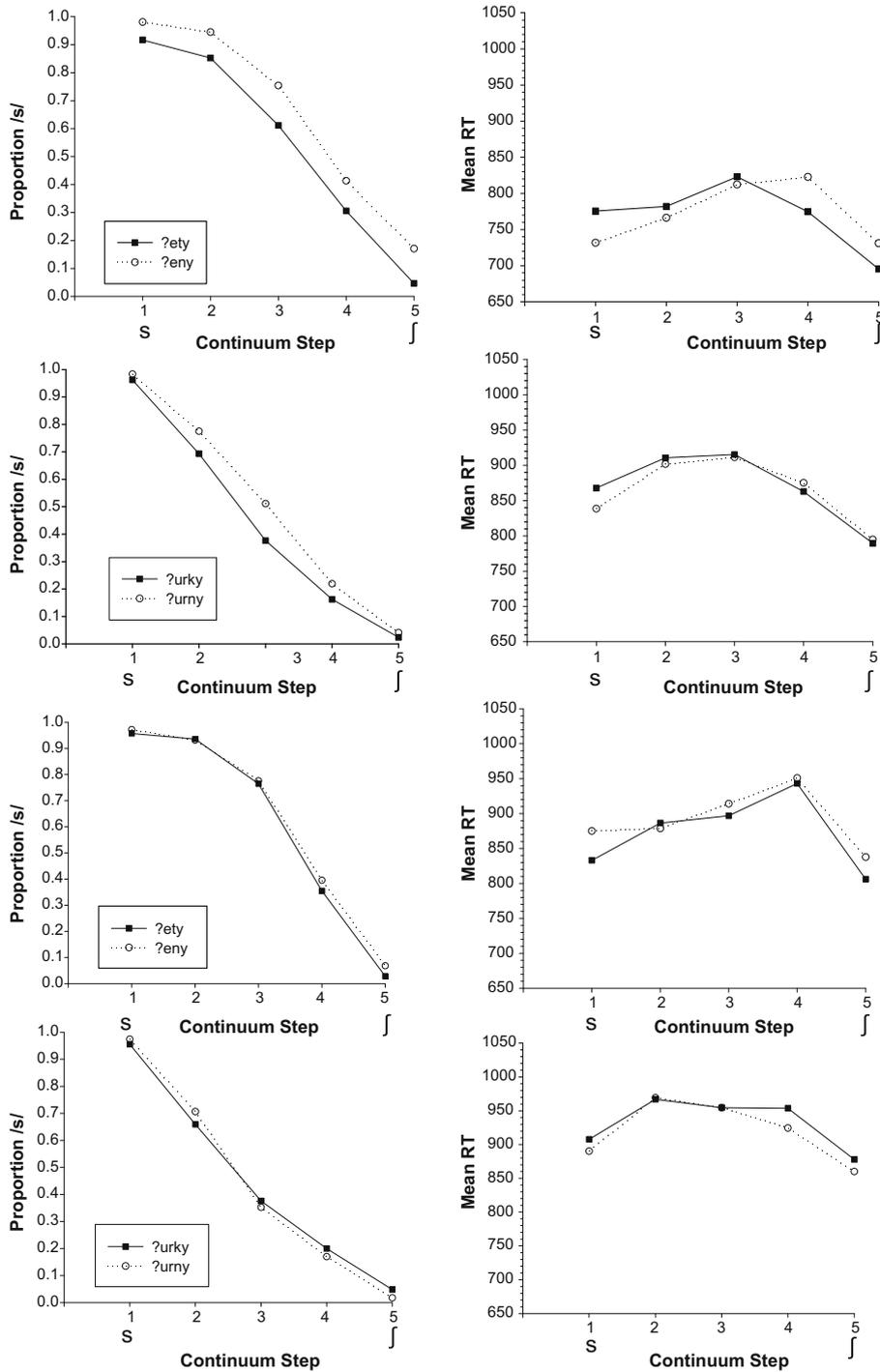
Fig. 4. Mean labeling and RT functions in Experiment 4 (top half of graph) and in the corresponding conditions in Experiment 3a (bottom half of graph). Data in the first and third row were obtained with the *?eny* test set and those in the second and fourth row with the *?urny* test set.

Along these same lines, one might wonder whether participants who made a connection between *seny* and *senty* on day 8 used this knowledge strategically to bias their responding toward *seny*. Although participants were not asked about this possibility, strategic influences would be expected to increase across the experiment as use of the

strategy improved. No such trend in performance (e.g., larger lexical shift) was found from the first to the second half of the experiment.

What is impressive (and surprising) about the results of Experiment 4 is the immediacy with which hearing the dialog affected phoneme labeling. The labeling session be-

gan less than twenty seconds after the dialog ended, yet large lexical influences were measured within the next ten minutes. These effects did not grow steadily over the test session, but are relatively stable, decreasing only slightly from the first to the second half of the experiment (e.g., ?eny: first half = .12, second half = .10). This outcome suggests that once an association is made between a pronunciation variant and a previously learned form of the word, lexical influences on processing the variant begin soon thereafter.

Experiment 5

The final experiment in this study examined whether the generalization found in Experiment 4 can also be found in other phonological contexts. The phonological contexts of *enty* and *urnty* permit /t/ deletion, so the results of Experiment 4 could be specific to word forms whose phonological structure undergoes predictable variation. If this is the case, it would suggest that knowledge of phonological variation contributes to generalization.

This hypothesis was tested by returning to the citation-biased stimuli of Experiment 2. Recall that although *content* produced a lexical shift, its /t/-deleted variant, *conent*, did not. If the results of Experiment 4 are due solely to lexical processes, then exposure to a nonword like *shontent*, with the same word-medial phonological structure as *content*, should yield generalization to the /t/-deleted form, *shonent*. A failure to do so would suggest that viability of the variation (e.g., phonological inference) aids generalization.

Method

Except for a change in test stimuli, the methodology was identical to Experiment 4. The new word participants were exposed to on day 1 to promote learning was *shontent*. Generalization to *shonent* was measured on day 8. The *ontent* context was chosen because its /t/-deleted form, *onent*, did not produce a lexical shift in Experiment 2. A /ʃ/-/s/ continuum was chosen to maintain continuity with the preceding experiments. A new fricative continuum had to be created. Results from a pilot experiment were used to select three ambiguous steps. Along with the two endpoint steps, each was prepended to a token of *onent* and *onage*, the context that served as the reference function. A separately recorded token of *shontent* was used during testing on day 1. The dialog that was presented at the start of testing on day 8 was rerecorded using *shonent*. There were 23 participants in the two-day experiment (exposure condition). Another 23 served in the corresponding no-exposure condition, participating only in the labeling session on day 8.

Results and discussion

The labeling and RT graphs are shown in Fig. 5, with the data from the exposure condition on the left and the data from the no-exposure condition on the right. Comparison of the lexical shifts across conditions indicates that generalization did not occur. Although the lexical shift in the exposure condition was statistically reliable, $t(22) = 2.73$,
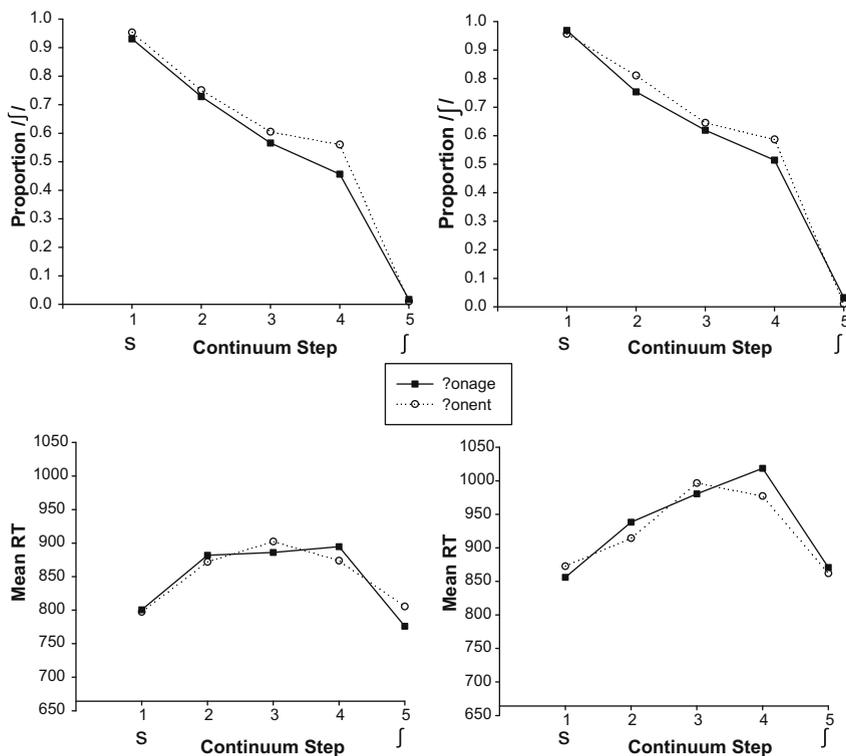

Fig. 5. Mean labeling and RT functions in Experiment 5. The graphs on the left are from the exposure condition and those on the right from the no-exposure condition.

it is about half the size of that found in Experiment 4 (.055 versus .10), and only slightly larger than that in the no-exposure condition (.052), which was not statistically reliable. The two lexical shifts are not statistically different from each other. In both conditions, listeners were slightly biased to respond /ʃ/ in the *onent* context, with the bias being largest at step 4.

The RT functions in the exposure condition resemble those in the no-exposure condition. There is neither a reliable RT advantage for *?onent* at the /ʃ/ endpoint nor a reliable RT slowdown at the /s/ endpoint.

The results of Experiment 5 suggest that for words whose phonological structure infrequently results in deletion of medial /t/, training with the citation form on day 1 is not sufficient to generalize to the /t/-deleted form on day 8. When compared with the results of Experiment 4, the findings show that generalization occurs more readily in a phonological context in which variation is common. One explanation for this is that phonological inference facilitates generalization.

The current results should not be interpreted as suggesting that generalization from *shontent* to *shonent* is impossible. A more elaborate training session would probably be sufficient to obtain a lexical shift. Evidence that listeners can in fact generalize to /t/-deleted forms of such words can be found in Table 1, where citation-biased words were pronounced without the medial /t/ 11% of the time in the Buckeye corpus. Although generalization is possible with new words like *shontent*, the results of Experiments 4 and 5 show that it is easier with new words whose phonological structure resembles words in which /t/ deletion is frequent. Listeners' sensitivity to this variation facilitates generalization.

General discussion

Pronunciation variation in speech makes the process of matching a spoken word with its representation in memory difficult. Phonological inference, a possible solution to the problem, was evaluated by exploring whether an inference rule is applied when a variant of a newly learned word is encountered. Experiments 1 and 2 established the preconditions necessary to test generalization of a phonological inference mechanism by showing that nasal-flapped variants are only processed as words when they occur in a very specific phonological environment. In Experiments 3 and 4, a perceptual learning paradigm was used to test generalization. The results showed that generalization to a pronunciation variant is not automatic, but requires a specific type of exposure in which the variant can be associated with the already-learned (canonical) pronunciation. The results of Experiment 5 suggest that generalization occurs more readily when the particular form of variation is common in words with the same phonological structure.

The results of this study provide solid evidence that phonological inference alone is insufficient to explain the recognition of variants of new words. Because generalization occurred only after lexicalization of the citation form *and* its association with the pronunciation variant,

the results also show that lexical involvement is necessary. Word-specific variation must be legitimized in order for a variant to generate lexical activity, and the results identify one legitimizing condition: meaningful exposure. Representational accounts such Ranbom and Connine (2007) and Pierrehumbert (2001) are supported by these results because they accurately predict that exposure to the variant is a necessary condition for generalization.

The necessity of lexical involvement in variant recognition is not only supported by the current data, but also by a consideration of the challenges posed by other words in the language and more extreme forms of speech reduction. There are words in English (e.g., *many, bunny*) whose pronunciation is phonologically similar to [kauni], making it difficult for a phonological process alone to determine whether a medial /t/ should be restored. In other words, as one reviewer put it, there is no motivation based solely on phonological grounds for interpreting *eny* as *enty*. Furthermore, in cases of extreme reduction (*probably->*[prai]; Ernestus, Baayen, & Schreuder, 2002) recognition by phonological rule alone would be quite challenging. Which rules should be applied and in what order?

These observations might make one wonder whether a representational account is sufficient for variant recognition. After all, a word's lexical representation is probably the most reliable source of information. As mentioned in the introduction, convincing evidence why a representational account may not be enough comes from experiments showing regressive place assimilation with stimuli that listeners perceive as nonwords (Darcy et al., in press; Gaskell & Marslen-Wilson, 1998; Mitterer & Blomert, 2003). This is clear evidence of phonological knowledge not only being recruited in the service of variant recognition, but also operating without the aid of lexical knowledge.

When variation occurs at the end of a word, it is more complex than word-internal variation because the conditioning environment spans a word boundary. The integration of information across words, which is required to recover the underlying form, is likely beyond the scope of lexical processes, which are traditionally considered to be only word-internal. Lexical knowledge aids in recognizing the word currently being encoded, not one that just occurred or will occur. Phonological inference processes, in contrast, have no such constraints. Although potentially sensitive to word boundaries, they are not limited by them. Phonological inference may be a necessary mechanism that aids variant recognition, operating most visibly in those environments in which lexical knowledge is inadequate.

When the broader literature on speech reduction is taken into account, solely representation-based proposals are challenged to provide a comprehensive account of variant recognition. It may well be that lexical processes dominate word-medially and phonological processes do so word-finally. Precisely because of this possible tradeoff, the hybrid proposal of Gaskell and Marslen-Wilson (1998), whereby phonological and lexical processes operate together to ensure accurate variant recognition, seems

most suitable at the present time. LoCasto and Connine (2002) offered a similar account when studying the processing of words that undergo vowel deletion (e.g., *camera* -> *camra*). They suggested lexical processes play a more dominant role than phonological inference in variant recognition. Given that they investigated word-medial variation too, it may not be a coincidence that the current assessment jibes with their conclusions.

Finally, the current data also provide details about the conditions necessary for generalization. They show that exposure alone is not sufficient, at least not when testing follows immediate exposure. In neither Experiment 3a nor the control condition of Experiment 4 did brief exposure to the variant result in a subsequent lexical shift. What is sufficient for generalization is for listeners to process the variant in a way that connects it with the already-stored and perceptually similar word form. This was achieved with the short dialog, and may be how listeners successfully process newly encountered pronunciation variants on a regular basis. The discourse provides the interpretive context in which to infer the intended meaning of the utterance. The encoding of a variant's acoustic realization (i.e., generalization to the new surface form) may be a by-product of making the correct inference, which can occur after just a few encounters, as in Experiment 4. By this proposal, variant encoding comes about through elaborative processing, not just exposure, and may well engage the same mechanisms responsible for the formation of new lexical entries (Leach & Samuel, 2007), although elaboration may need to be greater when variation is less predictable (Experiment 5). In the dialog between experimenters in Experiment 4, *seny* was used as a noun, referring to the name of the experiment. Although providing this information may have contributed to lexicalization, it alone was not enough, as the results of the control condition in Experiment 4 showed. Prior exposure to the citation form was also necessary. Under the right conditions, generalization is seamless. If it were not, pronunciation variants would disrupt conversation more than they do.

In conclusion, the prevalence and severity of pronunciation variation in conversational speech poses a nontrivial challenge for understanding how spoken words are recognized. The growing fields of word learning and variant recognition were merged by using a learning paradigm to study variant processing, which was viewed as a type of word learning in which the listener has a lexical representation of the word intended by the speaker, but the surface form produced by the speaker is unfamiliar. Generalization from the former to the latter requires linking the two realizations in a meaningful way.

Acknowledgments

Appendix A

A.1. Test words used in Experiment 1

Deletion-biased pronunciation		Citation-biased pronunciation	
Two syllable	Three-syllable	Two-syllable	Three-syllable
Center	Accounting	Contact	Contagious
Counter	Advantage	Content	Continue
County	Apprentice	Contest	Discontent
Lantern	Entertain	Context	Fantastic
Painted	Internet	Contour	Guarantee
Plenty	Interview	Mentor	Intention
Printer	Pentagon	Pontiff	Intestine
Twenty	Percentage	Quantize	Plantation
Wanted	Quantity	Syntax	Volunteer

Note: Except for *twenty*, only the first /t/ was deleted in those words containing multiple /t/s.

A.2. Dialog between two experimenters using the target word senny (Experiment 4)

Experimenter 1: Hey uh, how do you turn off the mic again?

Experimenter 2: Oh yeah, the mic. You're gonna need to actually unplug the mic to turn it off. Oh yeah, what's the name of this experiment again?

Experimenter 1: Senny. And a good way to remember that is to recall that it was one of the words that the participants heard on the first day of testing. One of them was Senny.

Experimenter 2: Oh yeah, I do remember one of the words being Senny. Are they going to be hearing Senny again today?

Experimenter 1: Yep, Senny is one of them that they'll hear.

References

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

Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (in press). Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry, & R. van de Vijver (Eds.), *Variation and gradience in phonetics and phonology*. Mouton De Gruyter.

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.

Dilley, L., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America, 122*, 2340–2353.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*, 35–39.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*, 162–173.

Ernestus, M., Lahey, M., Verhees, F., & Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America, 120*, 1040–1051.

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

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*, 105–132.

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.

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

Gow, D. W., Jr., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*, 279–296.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*, 485–499.

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., & Raymond, W. (1998) Reduction of English function words in Switchboard. In *Proceedings of the international conference on spoken language processing (ICSLP-98).*

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55*(4), 306–353.

LoCasto, P. C., & Connine, C. M. (2002). Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysics, 64*, 208–219.

Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics, 68*, 178–183.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition. *Journal of Experimental Psychology: General, 132*, 202–227.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics, 65*, 956–969.

Mitterer, H., Csepe, V., & Blomert, L. (2006a). The role of perceptual integration in the recognition of assimilated word forms. *The Quarterly Journal of Experimental Psychology, 59*, 1395–1424.

Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006b). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science, 30*, 451–479.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce. Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73–103.

Mitterer, H., & McQueen, J. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 244–263.

Neu, H. (1980). Ranking of constraints on /t, d/ deletion in American English: A statistical analysis. In W. Labov (Ed.), *Locating language in time and space.* New York: Academic Press.

Nolan, F. (1992). The descriptive role of segments: Evidence for assimilation. In D. R. Ladd & G. J. Docherty (Eds.), *Papers in laboratory phonology II: Gesture, segment prosody* (pp. 261–280). Cambridge, MA: Cambridge University Press.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 924–940.

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

Patterson, D., LoCasto, P. C., & Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica, 60*, 45–69.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and emergent grammar* (pp. 137–157). Amsterdam: John Benjamins.

Pitt, M. A. (in press). The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology: Human Perception and Performance.*

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). *Buckeye corpus of conversational speech* (2007; final release). Columbus, OH: Department of Psychology, Ohio State University (Distributor). <http://www.buckeyecorpus.osu.edu>.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347–370.

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

Pitt, M. A., & Samuel, A. G. (2006). Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1120–1135.

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language, 57*, 273–298.

Raymond, W. D., Dautricourt, R., & Hume, E. (2006). Word-internal /t, d/ in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change, 28*, 55–97.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*, 322–338.