



Published in final edited form as:

*J Exp Psychol Hum Percept Perform.* 2009 June ; 35(3): 896–910. doi:10.1037/a0013160.

## The strength and time course of lexical activation of pronunciation variants

**Mark A. Pitt**

*Ohio State University*

### Abstract

Spoken words undergo frequent and often predictable variation in pronunciation. One form of variation is medial /t/ deletion, in which words like *center* and *cantaloupe* are pronounced without acoustic cues indicative of syllable-initial /t/. Three experiments examined the consequences of this missing phonetic information on lexical activation. In Experiments 1, the Ganong (1980) paradigm was used to measure the strength of activation of /t/-deleted variants, comparing labeling and response time results with their citation counterparts. Phonemic restoration was used in Experiment 2 to generalize the results. In the final experiment, Experiment 1 was replicated with a large number of trials so that the time course of activation could be mapped. Results show lexical influences on labeling begin sooner and reach a higher level for the citation than the /t/-deleted variant, although the overall shapes of their activation profiles are similar.

---

Phonological analyses of corpora of conversational speech show that words are pronounced in different ways (e.g., *just* -> *jus*; *probably* -> *probly*). This variation can be severe (Ernestus, Baayan, & Schreuder, 2002; Johnson, 2004) and in its milder forms is reasonably predictable on the basis of the surrounding phonological environment, the type of variation, word frequency, speaker, morphological complexity, and other variables (Ernestus, 2000; Ernestus, Lahey, Verhees, & Baayan, 2006; Jurafsky, Bell, Gregory, & Raymond, 2001; Patterson, LoCasto, & Connine, 2003; Raymond, Dautricourt, & Hume, 2006). The ubiquity and variety of pronunciation variation is a challenge for understanding spoken word recognition. How do listeners recognize a pronunciation variant as the word intended by the speaker?

Research into this question has explored the consequences of pronunciation variation on word processing. There is a clear distinction between processing words that vary in ways commonly encountered in conversation versus variation that is unexpected. Words that undergo common forms of variation (e.g., vowel deletion, flapping, assimilation) prime their citation counterparts more strongly than a phonologically unrelated control word. In contrast, primes that deviate in uncommon ways (e.g., *horse* pronounced as *gorse*) show minimal or no evidence of lexical activation of the intended word (Connine et al, 1993; Marslen-Wilson, Nix, & Gaskell, 1995; McLennan, Luce, & Charles-Luce, 2003, 2005; Ranbom & Connine, 2007).

---

Correspondence information: Mark A. Pitt, Department of Psychology, 1835 Neil Avenue, Ohio State University, Columbus, OH 43210-1222, office (614) 292- 4193, fax (614) 688-3984, E-mail: pitt.2@osu.edu, <http://lpl.psy.ohio-state.edu>. Address correspondence to Mark Pitt, Department of Psychology, Ohio State University, 1835 Neil Avenue, Columbus, OH, 43210. Pitt.2@osu.edu

Mark A. Pitt, Department of Psychology, Ohio State University.

**Publisher's Disclaimer:** The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at <http://www.apa.org/journals/xhp/>.

How strongly pronunciation variants activate lexical memory depends on when and how activation is measured. Tasks that probe longer-term or more abstract representations tend to yield few differences in performance as a function of pronunciation variant. Using semantic priming, Deelman and Connine (2001) and Sumner and Samuel (2005) showed that primes differing in how a word-final /t/ is pronounced (e.g., released, unreleased, glottalized) generated comparable amounts of semantic priming. McLennan, Luce, and Charles-Luce (2003, 2005) obtained similar results in studies examining lexical activation of words whose medial /t/ can be flapped instead of released (e.g., [prɪtɪ] vs [prɪti]). Using a repetition priming paradigm in which primes and targets are presented in consecutive blocks of trials instead of sequentially on a single trial, response times (shadowing and lexical decision) were comparable regardless of the pronunciation of the prime or target, again suggesting abstraction from the surface form.

Tasks that tap more immediate form-based representations (standard repetition priming, phoneme monitoring, lexical decision) often show a cost in encoding pronunciation variation (Deelman & Connine, 2001; Janse, Nootboom, & Quene, 2006; Ranbom & Connine, 2007). For example, LoCasto and Connine (2002) found that mismatches in pronunciation between prime and target (*cabinet*->*cabnet*; *cabnet*-> *cabinet*) almost always induced a response slowdown to the target in auditory-auditory repetition priming. Janse et al (2006) obtained a similar pattern in a study examining lexical activation of words that undergo word-final /t/ variation. In a cross-modal repetition priming experiment, /t/-deleted variants (e.g., *jus* for *just*) were less effective primes than variants in which /t/ was instead released or unreleased.

That word processing can be slowed when a pronunciation variant is encountered makes sense on purely informational grounds if one assumes the citation form is the only one represented lexically: Deviations from the form stored in memory will slow processing. However, this outcome is counterintuitive in light of the fact that a pronunciation variant can be heard much more often than its citation form. Analyses of speech corpora consistently show that this is often the norm for listeners. Words with final /t/ are unreleased much of the time in conversational Dutch (Ernestus, 2000). The same is true in English for certain words (Neu, 1980; Deelman & Connine, 2001). More generally, specific phonological environments lead to consistent and frequent pronunciation variation in some words (Bell et al, 2003; Patterson, LoCasto, & Connine, 2003; Raymond et al, 2006).

Nevertheless, even when a variant's frequency is much higher than the citation form, the latter generates greater lexical activation than the former. Ranbom and Connine (2007; see also LoCasto & Connine, 2002) most recently demonstrated this using words such as center ([sentə] pronounced as [senə], without the [t]). In their analysis of the Switchboard speech corpus (Godfrey & Holliman, 1997), medial /t/ deletion,<sup>1</sup> or nasal flapping, was found to be very common in some words, occurring over 75% of the time. Although exposure to the citation forms of these words is comparatively infrequent, listeners performance in lexical decision and cross-modal priming paradigms showed that they are processed more efficiently. Reaction times were faster and accuracy was higher to the citation form.

The purpose of the present investigation was to define further the nature of the processing advantage for the citation form over the more common pronunciation variant. There were two goals. The first was to replicate past findings with complementary paradigms that can provide a measure of activation strength. Response time has been the primary dependent measure in prior work, and in the context of lexical decision experiments, has been interpreted as reflecting

---

<sup>1</sup>*deleted*, as used here, refers to [the] fact that acoustic cues indicative of syllable-initial /t/ are not visible in the spectrogram and /t/ is not heard upon close listening to the signal. A nasal flap might contain acoustic evidence signaling /t/, but what that evidence is and whether it is sufficient to induce perception of /t/ are unanswered questions.

differences in lexical processing between the realizations. If the locus of such effects is indeed lexical, and reflects differences in activation, then they should be found with other paradigms that require lexical memory for differences to emerge between conditions. Two paradigms that fit this bill in multiple ways are the Ganong (1980) paradigm, used in Experiment 1, and phonemic restoration (Samuel, 1981; Warren, 1970), used in Experiment 2. With both, the measures of lexical influences are computed from two-choice classification responses from which a measure of activation strength can be derived. Both also have a long history of being sensitive to lexical manipulations, so they are appropriate for asking questions about the degree of lexical activation of pronunciation variants. Finally, both measure lexical processing indirectly, in that lexical effects are computed from participants' decisions about properties of the stimulus other than its lexical status.

The second goal of the study was to map the time course of activation for a citation form and its pronunciation variant. A highly common pronunciation variant could very well generate an amount of activation that is equivalent to its citation counterpart, only the rate of activation might be slower due to the deviation in pronunciation. If this is the case, the peak of the activation function for the variant would be later in time, but could be just as high as that of the citation form. The time pressure placed on participants to respond quickly in lexical decision tasks may well tap activation at an early point in time, and thus miss the later evolution of the variant's activation profile. Alternatively, pronunciation variation could permanently handicap lexical activation, in which case no matter when activation is tapped, it would be lower for the variant. These possibilities were explored in Experiment 3.

## Experiment 1a

In the first experiment, the Ganong (1980) paradigm was used, in which listeners classify a word-initial phoneme as one of two alternatives (e.g., /s/ or /ʃ/). These phonemes are prepended to a context (e.g., *enter*) that forms a word at one endpoint and a nonword at the other (e.g., *center-shenter*). When listeners are presented a step from the middle of the continuum, which is acoustically ambiguous, they are biased to label the initial segment in a lexically consistent manner (e.g., responding /s/ given *enter*).

The magnitude of the labeling bias can serve as a measure of the strength of lexical activation, with a larger bias implying stronger lexical activation (Pitt & Samuel, 2006). Quantitatively, the lexical effect is measured relative to another context (e.g., *elter*) that biases labeling toward the opposite continuum endpoint (e.g., /ʃ/). If this additional continuum is held constant across pronunciation conditions (e.g., citation vs. /t/-deleted), it serves as a reference against which to measure changes in the magnitude of lexical activation. The size of the lexical effect is referred to as a *lexical shift* because of the way the labeling functions spread apart in the middle of the phonetic continuum.

I measured the sizes of the lexical shifts for two words that undergo frequent /t/ deletion, once when pronounced in their citation form, and once when pronounced with a nasal flap. The high frequency with which these words are heard in their reduced form in conversation suggests that they will generate robust lexical shifts, so at issue is the relative sizes of the shifts. If /t/-deleted variants generate less lexical activation, their lexical shifts should be smaller.

The time required to classify the phoneme has proven to be a valuable additional source of information about word processing, supporting and clarifying interpretation of the labeling data (Connine & Clifton, 1988). For example, if the deleted variant generates a lexical shift that is similar in size to the one found with the citation form, is it at a cost of slower responding (i.e., tapping activation later in time)? RT data were collected and analyzed to answer this question in order to further determine the consequences of processing pronunciation variants.

## Method

**Stimuli**—Two words were identified that not only had the appropriate phonological structure but also satisfied the criteria necessary to use the Ganong paradigm (e.g., identification of an appropriate word to use in the creation of the reference continuum). They were *center* and *counter*. *Center* occurs 53 times in the Buckeye Speech Corpus (Pitt et al, 2007; Pitt, Johnson, Hume, Kiesling, & Raymond, 2005; including derivatives like *centered*), with the medial /t/ never once being realized as [t]. *Counter* was much less frequent, but was also never pronounced with as [t] in its twelve occurrences. These pronunciation frequencies are very similar to those reported by Ranbom and Connine (2007), who used the Switchboard corpus (Godfrey & Holliman, 1997).

Word-initial phonetic continua were created using reference words that matched the targets in number of syllables, lexical stress, and the initial vowel, but whose initial consonant differed by one phonetic feature. The reference word for *center* was *shelter*, necessitating the construction of a /s-ʃ/ continuum. The reference word for *counter* was *powder*, necessitating the creation of a /k-p/ continuum.

Tokens of the target words (citation and deleted forms) and reference words were recorded onto DAT tape in a sound-dampened room, digitally transferred to a PC where they were downsampled to 16kHz (7.8kHz low-pass filtered), and then stored as individual sound files. Analyses of the citation and deleted pronunciation were performed to ensure the acoustic properties of the two pronunciations were as intended. The citation forms all 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 none of these cues. Instead, medial /n/ was coarticulated with the following vowel.

The phonetic continua were created by blending clear tokens of each endpoint in various proportions (Pitt & McQueen, 1998). To make the /s-ʃ/ continuum, tokens of each fricative were identified that were similar in duration (216 ms), excised from the word at the zero crossing prior to the first pitch period of the vowel, and then digitally combined, sample by sample, in steps of .05, ranging from a clear /s/ (100% /s/ plus 0% /ʃ/) to a clear /ʃ/ (0% /s/ plus 100% /ʃ/). These 21 steps were then prepended to three contexts (different tokens of *enter*, *elter* and a token of *ener*) after removing their initial fricative, again at vowel onset. This yielded the three continua for the *center* test set: *center-shenter*, *cener-shener*, and *celter-shelter*.

For the /k-p/ continuum, tokens of /k/ and /p/ 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 the 21 ratios, and then prepended to tokens of all 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/.

Two pilot experiments were conducted to identify steps in the middle of the two phonetic continua at which lexical influences were substantial, one using the *center-shenter* and *celter-shelter* continua, and the other using the *counter-pounter* and *cowder-powder* continua. Twenty listeners (10 for each pair of continua) 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. The three middle steps that yielded large differences in labeling between the two contexts (i.e., a lexical shift) were selected. These three middle steps plus the two endpoints made up the five steps on each of the three continua of a test set.

Labeling in the Ganong paradigm is sensitive to the size of a lexical neighborhood, which is defined as the number of words that deviate from the target word by a single substitution, deletion, or insertion of a phoneme (Newman, Sawusch, & Luce, 1997). In the current experiment, neighborhood effects could be mistaken for lexical effects if the difference in neighborhood structure of the endpoints on the citation continuum (e.g., *counter-pounter*) is not comparable to that between the endpoints on the deleted continuum (*couner-pouner*). That is, if there were a density advantage (i.e., more neighbors) for *counter* relative to *couner*, then a larger lexical shift for *counter* could be attributed to neighborhood differences, not pronunciation differences.

For the *counter* test set, neighborhoods were small (2–6 words) and neighborhood density (sum of log frequency counts) increased from the word to the nonword endpoints on the two continua (*counter*=1.08 vs. *pounter*=3.15; *couner*=0.6 vs. *pouner*=4.60). To the extent that differences between such small neighborhoods affect word perception, there is a density advantage for *counter* (1.08–3.15=–2.93) over *couner* (0.6–4.60=–4.54), but in both cases it is biased towards the nonword endpoint. Analysis of the *center* test set showed that neighborhood size spanned a larger range (3–12 words) and density was larger at the word endpoints (*center*=5.06 vs *shenter*=2.54; *cener*=6.84 vs *shener*=1.45). The density advantage is for the /t/-deleted pronunciation (*center*: 5.06–2.54 = 2.62; *cener*: 6.84–1.45=5.39). Similar results are obtained if the number of neighbors is used as the measure of neighborhood density. These analyses show that lexical neighborhood does not consistently covary with pronunciation of the target in a manner that could render interpretation of the data ambiguous.

**Procedure**—A different group of listeners was tested on each of the two test sets. Stimulus presentation was blocked such that participants responded to each step on the three continua before a step was presented again, except for steps on the reference continuum (e.g., *selter-shelter*), which were presented twice for every presentation of the steps on the other two continua (e.g., *center-shenter* and *cener-shener*) so that listeners heard /s/-biased and /ʃ/-biased contexts equally often. Presentation within each block was randomized. With five steps per continuum, each block contained 20 trials. There were ten blocks in all, for a total of 200 trials.

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. Two buttons on a response box situated in front of participants were labeled with the letters that correspond to the two phonemes (e.g., *k* and *p*, or *s* and *sh*). After listening to instructions to respond quickly yet accurately on each trial, participants sat through a 14-trial practice session before proceeding to the test session. On each trial listeners were given 2500ms to respond after stimulus offset. Once all listeners had responded, there was a 2500ms pause before the next trial. Stimulus presentation and response collection were controlled by computer. After the experiment, listeners were given a surprise recall test in which they were asked to write down all of the utterances (words and nonwords) heard during the experiment.

**Participants**—Participants were 36 undergraduates enrolled in Introductory Psychology. All participated to fulfil a course requirement and none reported hearing difficulties. Half heard the *counter* test set and half the *center* test set. Although data were not collected on their place of birth or where they were raised, given the demographics of the OSU student body (> 90% native to Ohio), it is a good bet that these participants are exposed to medial /t/ variation at the rates found in the Buckeye corpus.

## Results

**Lexical Shift Analyses**—The proportion of /k/ (or /s/) responses at each step on each continuum was calculated for all participants. These data were then averaged across listeners



within each test set and are graphed in the top row of Figure 1, with the *counter* data on the left and the *center* data on the right. Lexical influences on phoneme categorization are present when the labeling functions diverge so that responses to the lexically consistent category (e.g., /s/ given *enter* and /ʃ/ given *elter*) are most frequent. This lexical shift was quantified by calculating the mean difference in labeling between the three middle steps on the citation (and deleted) continuum and the reference continuum (Pitt & Samuel, 1993). All p-values were less than .05 unless reported otherwise.

Lexical shifts were obtained with the citation forms of both test sets, with the sizes of the shifts differing noticeably across continua (/k-p/ vs /s-ʃ/). In both cases the size of the shift was statistically reliable (?*ounter* shift: .10;  $t(17)=3.42$ ; ?*enter* shift: .64;  $t(17)=14.45$ ; ? denotes the perceptual ambiguity along the continuum). Also across both graphs, one can see that a smaller lexical shift was obtained with the /t/-deleted variant, suggesting reduced lexical activation. Although these shifts were reliable in their own right (?*ouner* shift: .08;  $t(17)=2.45$ ; ?*ener* shift: .53;  $t(17)=14.39$ ), the drop in shift size from the citation to the deleted pronunciation was not reliable for ?*ounter*,  $t < 1$ , but was for ?*enter*,  $t(17)=4.57$ .

**Reaction Time Analyses**—The RT data were aggregated in the same way as the categorization data. For each participant, response times at each continuum step were averaged separately for each of the three continua of a test set. Individual participant data were then combined and the mean RT functions for each test set are graphed in the bottom row of Figure 1. As with the labeling data, we are interested in how the two test functions (e.g., ?*ounter* and ?*ouner*) are positioned relative to the reference function (e.g., ?*elter*) and to each other. However, unlike the labeling data, the direction of the shift is not fixed, but reverses across the phonetic continuum as the lexical status of the strings change from words to nonwords. On the /k/ side of the continuum, responses should be faster in the /k/-biased context (e.g., *ounter*) than in the /p/-biased context (*owder*), causing the ?*ounter* function to lie below the ?*owder* function. These two functions should swap positions on the /p/ side of the continuum, where the *owder* context should speed /p/ responses and the *ounter* context should slow them down. As with the labeling data, of interest is the relative position of the /t/-deleted function (*ouner*).

Lexical influences on response speed were analyzed separately on each side of the continuum. The cross-over point between the reference (?*owder*) and citation (?*ounter*) function was used to split the continuum in two. For both the *counter* and *center* test sets, this point fell between steps 3 and 4, resulting in steps 1–3 being grouped together and steps 4–5 being grouped together. Results on the /k/ and /s/ sides of the continua (steps 1–3) are of most interest because this is where facilitory effects from lexical activation of the /t/-deleted variant should be found.

For the *counter* test set, a large effect of lexical status is found with ?*ounter* and ?*owder*, as would be expected. From step 1 through 3, responses were an average of 53 ms faster in the *ounter* context than in the *owder* context, clear evidence of lexical biases affecting categorization speed,  $t(17)=2.32$ . At steps 4 and 5, the /p/ side of the continuum, there was a 63 ms word RT advantage for the ?*owder* function,  $t(17)=2.85$ .

/t/ deletion affected response time in a way that suggests *couner* activated lexical memory more weakly than *counter*. The ?*ouner* function lies between the ?*ounter* and ?*owder* functions across the continuum. The effect of reduced activation is visible across steps 1–3, with RTs averaging 38ms slower when the context was *ouner* than *ounter*,  $t(17)=2.17$ . Note that RTs given *ouner* are only slightly (15ms) faster than those given *owder*. The weaker lexical activation by the /t/-deleted variant aided responding on the /p/ side of the continuum, by reducing the magnitude of the RT slowdown. At steps 4 and 5, the difference between the ?*ouner* and ?*owder* functions was 33ms ( $t(17)=1.21$ ,  $p < .24$ ), half the size of the slowdown found

with the *?ouner* function (63 ms). The 30ms RT difference between the *?ouner* and *?owner* functions at steps 4 and 5 was not reliable,  $t(17)=1.64$ ,  $p<.12$ .

The RT functions of the *center* test set are not dome-shaped like those of *counter*, with the endpoint RTs being fastest and responses to the middle steps being slowest. Rather, RTs were much faster at the word than nonword endpoints, indicating that the lexical contexts had strong effects on response speed across the entire phonetic continuum. Statistical comparisons showed that at the /s/ endpoint, /s/ responses were reliably faster in the *enter* than *elter* contexts,  $t(17)=5.70$ . The reverse was true at the /ʃ/ endpoint,  $t(17)=5.20$ . This same pattern emerged when the deleted variant, *?ener*, was compared with *?elter* (/s/ endpoint:  $t(17)=2.19$ ; /ʃ/ endpoint:  $t(17)=5.69$ ). As with the counter test set, comparison of the *?enter* and *?ener* functions shows that the citation form generated greater lexical activation than the deleted form. Response times on average were 46 ms faster to *?enter* than to *?ener* at the /s/ endpoint,  $t(17)=2.29$ . At the /ʃ/ endpoint, RTs were also faster to *?enter* than to *?ener* (35 ms), but not reliably,  $t(17)=1.94$ ,  $p=.25$ ; this last outcome is the reverse of what was found for the *counter* test set.

Before discussing the implications of the results, the experiment was rerun to address a methodological concern: Are the effects found with the deleted forms influenced by the presence of the citation forms, or vice versa? The within-subject design of Experiment 1a could have biased how listeners perceived or responded to the deleted variants. For example, the presence of the citation pronunciation might have exaggerated differences between it and the deleted variant, increasing the chances that listeners heard *center* and *counter* as nonwords, thus reducing lexical effects. In Experiment 1b, the deleted and citation continua were presented to different groups of listeners.

## Experiment 1b: Replication between listeners

### Method

The experimental setup and procedure were identical to Experiment 1a except for the between-groups design. With two words (*center* and *counter*) and two pronunciations of each, four groups of participants were tested. All also heard the appropriate reference continuum. Each continuum step was presented 12 times, for a total of 120 trials. For the *center* test set, 40 new participants were tested, 20 in each group. For the *counter* test set, there were 23 participants in each group.

### Results

The data were aggregated and analyzed following the procedure in Experiment 1a. Mean labeling and RT functions are shown in Figure 2. The labeling data closely resemble what was found when the experiment was run within-subjects. The only significant departure are the labeling shifts for the *counter* test set. As can be seen, large and statistically reliable lexical shifts were found with both pronunciations (*ouner*: .23,  $t(22)=11.01$ ; *owner*: .16,  $t(22)=6.59$ ). The difference between them was reliable,  $t(44)=2.08$ . Inspection of the graph shows that the two reference functions are fairly close to each other. What differs most are the *?ouner* and *?owner* labeling functions. Proportion /k/ responses given *ouner* were much greater than given *owner* in the middle of the continuum. Much the same outcome was obtained with the *center* test set. The lexical shift for *enter* was .70,  $t(19)=18.79$ , and that for *ener* was .60,  $t(19)=10.96$ . The .10 difference in shift size between pronunciations approached statistical significance,  $t(38)=1.76$ ,  $p<.09$ .

The RT data also resemble those of Experiment 1a. *?ouner* generated reliably faster RTs (30ms) compared to *?owder* at the /k/ endpoint (steps 1–3),  $t(22)=2.53$ . Just the reverse was obtained at the /p/ endpoint (45ms;  $t(22)=3.58$ ). As in Experiment 1a, *?owner* RTs were not

reliably different from *?owder* RTs at the /k/ endpoint, 18 ms;  $t(22)=1.86$ , but much slower than them at the /p/ endpoint (49ms;  $t(22)=3.56$ ).

Despite replicating the basic RT pattern, there is one salient difference in RTs across experiments, which may explain why labeling shifts were so much larger in Experiment 1b: RTs to the *counter* test set were on average 212ms faster in Experiment 1b than in 1a. Comparison of the two RT graphs across Figures 1 and 2 shows the distributions overlap minimally, with the slowest mean RT in Experiment 1b (758ms) being 85ms faster than the fastest mean RT in 1a (833ms). The faster responding in Experiment 1b (*counter* test set only) appears to have tapped lexical activation earlier in time, when it was not only greater, but differences in activation between the citation and deleted forms were greater as well.

For the *center* test set, response times fell within a similar range as in Experiment 1a, and the data patterned similarly as well. At the /s/ endpoint, *?enter* generated faster RTs than *?elter*,  $t(19)=4.29$ . The reversal at the /ʃ/ endpoint was also reliable,  $t(19)=5.94$ . Comparisons of *?ener* with *?elter* yielded a reliable differences at the /ʃ/ endpoint,  $t(19)=6.42$ , but not the /s/ endpoint,  $t(19)=1.26$ ,  $p<.22$ . Just as in Experiment 1a, responses to *?ener* were slower than to *?enter* across the continuum. However there was too much variability between groups of participants for the difference to reach statistical significance at either endpoints (/s/ and /ʃ/:  $t<1$ ).

## Discussion

A summary of the results of Experiment 1 is presented in Table 1. Rows 1 and 2 contain the sizes of the lexical shifts for the citation and deleted pronunciations. The differences between the two are shown in the third row. Scanning across the data, one cannot help but notice that the shift is more than twice as larger for the *center* than *counter* test sets (.105 vs .045). Although the drop in shift magnitude for the *counter* test sets can appear puny in Experiment 1a, it is most informative to evaluate these numbers in the context of the absolute size of the lexical shifts themselves (rows 1 and 2). The percent change in shift magnitude is a more meaningful measure because it normalizes for such differences. These values are shown in parentheses in the third row. Three are quite comparable, with the right-most value being somewhat higher. The labeling data show clearly that words whose phonological structure leads to frequent medial /t/ deletion activate their lexical representations, but consistently less so than the citation forms.

The RT data reinforce the above claims. The bottom row of Table 1 contains the mean difference in milliseconds between the citation and deleted RT functions (steps 1–3), but only for the within-subject comparisons because between-group differences were too variable and inconsistent. Positive values represent faster responses in the citation context relative to the deleted context. Responses in the citation context were always faster by a healthy margin, indicating faster lexical activation given the canonical pronunciation.

One unexpected outcome was the failure of the *?enter* and *?ener* functions to cross over at steps 4 and 5 and show an RT advantage for *?ener*. This happened in both Experiments 1a and 1b, but only with the *center* test set. As mentioned above, a likely reason for this outcome is the fricative continuum itself. Even clear tokens of /s/ and /ʃ/ can be confusable, especially when embedded in a lexical context. Steps at the /ʃ/ endpoint may have been sufficiently ambiguous to be susceptible to lexical biases, thereby extending the lexical advantage at the /s/ endpoint to the /ʃ/ endpoint. The absence of dome-shaped RT functions, which were found with the /k-/p/ continuum, supports this interpretation.

Inspection of the RT and labeling data together raises interesting questions about the temporal dynamics of processing pronunciation variants. In both the *counter* and *center* test sets, the



deleted pronunciation produced a large lexical shift, but at a cost: RTs were substantially slower than to the citation form. From what is known about the time course of lexical effects in the Ganong paradigm (Miller & Dexter, 1987; Pitt & Samuel, 1993, 2006), there is good reason to think that the labeling shift for the deleted variant would shrink in size if listeners responded as fast as they did to the citation pronunciation. This coupling of lexical influences across dependent measures can be used profitably to study the time course of processing the two pronunciations, and is done in Experiment 3.

One might wonder whether listeners heard the deleted variants as the intended word or were aware of the slight differences in pronunciation of the deleted and citation forms. In a surprise recall test after the experiment, listeners were asked to write down all of the words and nonwords they heard. Most (78%) recalled the test words (e.g., *counter*). All who did classified them as a words. Of the few of who distinguished between the two pronunciations (13%), all classified the deleted form as a word. Most listeners did not notice or did not bother to distinguish between the deleted and citation pronunciations. Even the listeners who heard only the deleted variant (half the participants in Experiment 1b) restored the /t/ when recalling the stimuli. This response pattern does not reflect a reluctance on the part of listeners to report nonwords. They did so freely; nonword endpoints (e.g., *selter*, *pounter*) were reported just as often as word endpoints.

## Experiment 2: Converging evidence with phonemic restoration

The results of Experiment 1 replicate prior work in showing that words that undergo frequent medial /t/ deletion in casual speech yield lexical activation, but not as much as that found with the citation form. The results build on this literature by providing one estimate of the difference in activation strength, 21% when averaged over the four percentages in Table 1.

In Experiment 2, the comparison was repeated using phonemic restoration (Samuel 1981; Warren, 1970), which was chosen because it complements the Ganong paradigm in important ways. One is that, like Ganong, restoration measures lexical effects indirectly. Listeners' attention is focused on deciding whether a segment is present in a word, and their ability to do so is influenced by lexical variables. However, the two tasks lead to opposite predictions about how lexical effects will influence performance. In the Ganong paradigm, the size of the labeling bias increases positively with the strength of lexical influences. In phonemic restoration, the listener's task is to discriminate segments that have been replaced with noise from those to which noise has been added. As lexical influences increase, discrimination difficulty increases, with listeners finding it harder to tell whether a segment is noise-added or noise-replaced.

In terms of stimulus choice, phonemic restoration is more flexible than the Ganong paradigm, which can be used only when stimuli satisfy a number of criteria. Not only must there be a pair of words available to create the word-nonword and nonword-word continua, but a suitable phonetic continuum must be created, which can be tricky. Phonemic restoration does not have these limitations, which enabled the use of much larger and more variable set of stimuli.

If the results of Experiment 1 replicate using phonemic restoration, then discrimination should be worse (lower  $d'$ ) for the citation than deleted pronunciations. Longer words generate more lexical activation than shorter ones (Pitt & Samuel, 2006), so an additional drop in discriminability should be found with three-syllable words. In their study on processing vowel-deleted variants, LoCasto and Connine (2002) found vowel deletion disrupted lexical activation less in three- than two-syllable words, so a similar effect of length was expected here.

## Method

**Stimuli**—Eighteen words served as the target stimuli (listed in the Appendix). All possess the same phonological structure in which medial t-deletion is common (/nt/ preceded by a full vowel that receives primary stress and a following vowel that is unstressed), and all exhibit t-deletion in the Buckeye speech corpus (Pitt et al, 2007). The ubiquity of this form of pronunciation variation can be appreciated when it is considered that for two-syllable words alone, t-deletion occurred 75% of the time (678 tokens, 51 types). Equally important, only 5% of ts were released, with the remaining pronunciations being labeled as [ʃ] or [d]. Ranbom and Connine (2007) report similar statistics in the Switchboard corpus (Godfrey & Holliman, 1997).

Half of the target words were two syllables and half were three syllables in length. In four of the three-syllable words, /nt/ straddled the first and second syllables; in five it straddled the second and third syllables. More words than this were not used because I wanted to limit the test to those words for which we had the most information about production variation.

Fillers were words (132) and nonwords (150) from one to three syllables in length, each length occurring with similar frequency. Nonwords were created by changing consonants and vowels of words not already used in the experiment. Nonwords were included out of concern that if a deleted variant was heard as a nonword, it would stand out from the words, be responded to differently, and potentially alter participants' response strategy. With an equal number of nonwords, this problem was eliminated. Such a large number of fillers was included to ensure that the deleted variants were not conspicuous.

Targets (including new tokens of *center* and *counter*) and fillers were recorded using the equipment and procedure of Experiment 1. Deleted forms of the targets were spoken in a relaxed and casual style. Acoustic analyses of the stimuli showed no evidence (visual or auditory) of stop closure or burst release cues in the deleted forms, but both were present in the citation forms (closure plus release averaged 84ms across tokens). This difference in pronunciation resulted in the deleted variants being 55ms shorter than the citation forms (655ms vs. 710ms).

The first vowel in the target words was chosen as the phoneme to be altered (replaced by noise or mixed with noise). This choice was made to avoid interfering with processing the medial /t/ in the citation form, the critical manipulation between conditions, and also its immediately adjacent phonemes (preceding /n/ and following vowel). Because word-final phonemes in some of the two-syllable words immediately followed /t/ (e.g., *county*, *plenty*), this left the first and second phonemes as viable choices, of which the latter seemed preferable because it is not at a word boundary.

The phonemes chosen to be altered in the fillers varied in position within the word, in phone class (vowel or consonant), and stress of the syllable containing the altered phoneme (stressed or unstressed) to camouflage the targets.

To construct the added and replaced versions of each stimulus, the stretch of the to-be-altered speech signal was identified and measured using auditory and visual cues. The sound files and their corresponding measurements were fed into a program that generated the added and replaced tokens of each stimulus, making four versions of each target word (deleted-added, deleted-replaced, citation-added, citation-replaced) and two versions of each filler (added, replaced). Signal correlated noise was used (see Samuel, 1996).

The altered vowels (replaced and added) were then excised from the target words and pretested to ensure that discrimination out of context was comparable across the citation and deleted

pronunciations. If it were not, then differences found when the listeners responded to the vowels embedded in the words could be misleading. Replaced and added versions of each vowel were presented once each to 39 listeners in a random order, which paralleled the design of the main experiment, described below.  $d'$  was calculated separately for each vowel. Mean  $d'$  across the 18 words in each pronunciation condition was 2.622 (citation) and 2.800 (deleted), demonstrating extremely good and comparable discriminability. The .178 difference is not statistically reliable,  $t < 1$ .

## Design and Procedure

To make this experiment a stringent test of the hypothesis that deleted variants produce less lexical activation than their citation forms, each target word (citation or deleted form) was presented only once to a participant, which is the opposite of what is done in the Ganong paradigm, where listeners hear a small set of stimuli many times. The combination of a small set of targets, and the fact that signal detection analyses are observation-hungry, made it impossible to design an experiment whose data could be analyzed by participants. With 18 targets, each presented only once, there would be too few observations per condition to obtain trustworthy  $d'$  values.

The experiment was therefore designed to compare item performance across conditions by having a large number of participants respond to the various pronunciations of each target word, each of which appeared in a different version across four stimulus lists. All of the fillers, half added and half replaced, occurred in each list. For two of the lists, the added version was presented; for the other two, the replaced version was presented.

Listeners were tested up to four at a time using the same equipment as in Experiment 1. Two buttons on the response boxes were labeled “added” and “replaced.” Listeners were instructed to press the button labeled “added” if they thought “noise was added to the letter sound” or “replaced” if they thought “the letter sound was replaced with noise.” The 300 test trials followed an 18-trial practice session. Listeners had 2500ms to respond after stimulus offset, and there was a 2000ms pause before the next trial began.

**Participants**—Participants were 120 new undergraduates from the same pool as Experiment 1. Thirty were tested in each of the four lists.

## Results and Discussion

$d'$  (a measure of discriminability) and beta (a measure of response bias) were calculated for the citation and deleted forms of each of the 18 target words. Mean values, along with the mean proportion of false alarms (responding “added” when the phoneme was replaced) and misses (responding “replaced” when noise was added to the phoneme) are listed in the first row of Table 2.

The results replicate and generalize what was found in Experiment 1. The  $d'$  data show a clear discrimination advantage for the deleted variants.  $d'$  is higher when the word was the deleted variant than the citation form. The .446 difference in  $d'$  was statistically reliable,  $t(17) = 3.24$ , with 14 of 17 words (one tie) showing the effect in the observed direction. Inspection of the misses and false alarms shows that the source of the poorer discrimination for the citation form is a higher frequency of responding “added” when the segment was in fact replaced with noise, just what would be expected if citation forms produced stronger lexical activation. The difference in false alarms between the citation and deleted conditions was reliable,  $t(17) = 4.08$ , with 15 of 18 items showing this effect in the same direction. The difference between the miss rates was not reliable. The virtually equivalent betas across pronunciations indicates that

listeners applied the same criteria when responding to both types of words, showing a slight bias to respond “replaced.”

When the results are analyzed by word length, not only is the same pattern found, but expected differences emerged between the two- and three-syllable words (lower rows of Table 2). Discrimination was better for deleted than citation forms for both two- and three-syllable words (two-syllable:  $t(8)=2.83$ ; 8 of 9 words; three-syllable:  $t(8)=1.79$ ,  $p<.11$ ; 6 of 8 words, with one tie). Comparison across rows shows that discrimination in the deleted and citation conditions dropped sharply from the two-syllable to the three-syllable words,  $t(8)=2.93$ , replicating past findings showing that lexical influences increase with word length (LoCasto & Connine, 2003; Pitt & Samuel, 2006).

The miss and false-alarm data parallel those found overall. Miss rates were low and impressively constant across word length. Where systematic differences show up is again in the false alarm data, which show that discrimination difficulties were due to an increase in restoration when listeners heard the citation form. False alarms were reliably higher to the citation than deleted pronunciations (two-syllable:  $t(8)=2.68$ ; three-syllable:  $t(8)=4.00$ ). The main effect of word length was also reliable,  $t(8)=3.44$ , but the interaction of word length with pronunciation was not. The small differences in betas between the citation and deleted forms was not reliable for either word length. However, there was an effect of word length on bias, with beta being reliably higher (by .20) for three-syllable than two-syllables words,  $t(16)=2.71$ .

Unlike in Experiment 1, there was no reference (i.e., nonword) condition from which to determine whether the /t/-deleted forms generated any lexical activation. Although filler nonwords were included in the experiment, the phonemes that were altered and their locations in the nonwords were not matched to the target words. Despite these differences, a comparison of the two was undertaken to provide a preliminary answer to this question. Eighteen filler nonwords (nine of each length) were chosen whose altered phoneme was sonorous (e.g., vowel, liquid), thereby matching them to the targets as closely as possible. The position of the altered phoneme within the nonwords could not be matched well; only 25% occurred in the first syllable, which is where they occurred in the targets. Mean  $d'$  for the nonwords was 1.39, .23 units greater than that obtained with the /t/-deleted forms (1.16). Although this difference was not statistically reliable ( $p=.21$ ), it is in the direction expected if the deleted variant activated lexical memory.

The data of Experiment 2 replicate those of Experiment 1 in showing that /t/-deleted variants activate lexical memory to a lesser degree than their citation forms. The results generalize this observation to more words, words that vary in length, and using a different paradigm. For two-syllable words, which were used in Experiment 1, the drop in activation strength is in the same range as but somewhat higher than that found with the Ganong paradigm (30% vs 21%). Even though words like *county* and *internet* are heard most often without medial /t/ in conversation, /t/ deletion has a detrimental effect on lexical activation.

### Experiment 3: Mapping the time course of lexical activation

In the final experiment, the time course of lexical activation of the two pronunciations was compared. Although the labeling data of Experiment 1 show that the deleted variant generates less lexical activation overall, the RT data suggest that their time courses are different as well: Activation of the deleted form is slower than the citation form. Together these results suggest that activation of the citation form begins earlier, or possibly rises more steeply once initiated. Which of these alternatives is the case, and does the citation form maintain this advantage across its activation history?

Too few observations were collected per continuum step in Experiment 1 to partition the data by response speed (Fox, 1984) to address these issues. Experiment 1a was therefore rerun with three times as many responses per step to make it possible to analyze the evolution of the lexical shift over time.

## Method

Except for a few procedural changes, the experimental setup was identical to Experiment 1. Only the *counter* test set was used. For each step, 48 observations were collected (960 total trials), and listeners were instructed to keep their responses within a specified range by providing feedback about response speed on the computer screen in front of them. The purpose of the feedback was to ensure participants produced enough fast responses to measure lexical activation early in its evolution. Because labeling shifts were largest for the *counter* test set when RTs were fast, this procedural change was necessary to induce fast responding. Mean RT was calculated for each participant after every block of 20 trials, and written feedback was displayed on the computer screen for 6.2s before the next trial began to encourage participants to respond within the desired range. The words “good pace” were printed on the screen when the mean RT fell between 500 and 900 ms. A value less than 500 ms generated feedback of “too fast,” and one slower than 900 ms, “respond faster.” A new group of 29 participants from the same population as Experiment 1 was tested.

## Results and Discussion

The methodology was successful in generating fast responding, with a median group RT of 655ms. The effect of this shows up in the overall analyses in two ways. One is that the lexical shift for the citation form is much larger than that found in Experiment 1a (.16 vs .10), but only slightly larger for the deleted form (.095 vs .08). These shifts are shown in the top half of Figure 3. Both are statistically reliable (*?ounter*:  $t(28)=10.24$ ; *?owner*:  $t(28)=6.49$ ), and now with the increased separation between them, so is their difference,  $t(28)=7.13$ .

Fast responding by participants also muted RT differences across continua (compare Figures 1 and 3). On the /k/ side of the continuum, there were much smaller but still reliable RT advantages for *?ounter* over *?owder* and *?owner* (16ms,  $t(28)=3.01$ ; 21ms,  $t(28)=3.56$ , respectively). On the /p/ side of the continuum, the functions for the two pronunciations behaved similarly, slowing down relative to *?owder* (*?ounter*: 28ms,  $t(28)=5.1$ ; *?owner*: 37ms,  $t(28)=5.8$ ).

With RTs to *?ounter* and *?owner* more comparable, larger differences emerged in their lexical shifts; this outcome forecasts what the RT partition analysis will reveal. The overall results also provide yet another demonstration that the deleted variant produces significantly less and slower lexical activation than the citation form.

The time course of activation of the two variants was investigated by measuring lexical shifts in consecutive 100ms windows starting at 400ms (Pitt & Samuel, 2006). The actual method of calculating their size was identical to the overall analysis, except that only responses within each 100ms RT bin were used. To ensure reliable measurement of shift size, there had to be at least three observations per step in the calculation of the lexical shift for each participant. When this criterion was not met, the group mean was used for that participant, which occurred at most five times (in the slowest bin, 900–1300 ms). This last RT bin was an aggregate of data from a larger window because there were so few observations at this end of the RT distribution.

Mean lexical shifts in each RT bin are shown the bottom graphed of Figure 3 as a function of the pronunciation of the word. The lexical shift is consistently larger for *?ounter* than *?owner* across all RT bins. Inspection of the shapes of these activation functions reveals a few



differences in their evolution. Most noticeable is the rise in activation from 400 to 500 ms. For ?*ounter*, the rise is quite steep (.12 units), peaking at 500ms before dropping off slightly (.035) over the remainder of the bins. For ?*owner*, the slope is much shallower (.07). In addition, ?*owner* shows a slight peak later in time, in the 700ms bin, before trailing off minimally. A two-way ANOVA with pronunciation and RT bin as factors produced main effects of both variables (Pronunciation:  $F(1,28)=33.01$ ,  $p<.01$ ; RT bin:  $F(5,28)=3.22$ ,  $p<.01$ ), but no interaction, indicating that the trajectory of lexical activation, once initiated, is similar for the two variants, just lower for ?*owner*.

That ?*ounter* activation rose at a rate almost double that of ?*owner* from the 400 to 500 ms bin (.12 vs .07) warranted closer inspection. Statistical comparison of this difference only approached significance,  $F(1,28)=2.94$ ,  $p<.10$ , however, because of a high degree of variability in shift size in the 500 ms bin; some listeners showed large shifts and others much smaller ones. Although the data are suggestive of the idea that citation forms generate a more rapid rise in activation than deleted forms, firm conclusions are premature at present.

What is not equivocal in Figure 3 is that lexical influences on labeling begin earlier for the citation than deleted form. In the 400ms bin, ?*ounter* generated a small but reliable lexical shift,  $t(28)=2.37$ ,  $p<.03$ , which is what would be expected if lexical activation is just getting underway. ?*owner*, in contrast, generated almost none. The difference between the two pronunciations in this RT bin is reliable,  $t(28)=2.13$ ,  $p<.04$ . Another comparison that brings out the sizeable difference in onset of activation is to compare shift size across bins. The mean shift in the 400ms bin for ?*ounter* is only slightly smaller than that in the 500ms bin for ?*owner* (.055 vs .079;  $t(28)=1.08$ ,  $p<.29$ ), yet it occurs 100ms earlier in time. Finally, if one extrapolates backwards in time, phoneme identification responses earlier than 400 ms should show no sign of lexical influences, identifying the onset of activation of ?*ounter*. There were only nine participants who generated enough responses between 200 and 400 ms after stimulus onset to compute stable lexical shifts. For both ?*ounter* and ?*owner*, mean shifts were .00 and .01, respectively. At this early stage in word processing, listeners have heard so little of the utterance that activation is too weak to bias labeling even to ?*ounter* (Miller & Dexter, 1988).

To understand what acoustic properties in the words differentiate the initial rise in activation, waveforms of the final portions of both pronunciations were time-aligned to the x axis in the lower graph, *counter* above *couner* (offsets spliced at 600ms). The stop burst, and perhaps closure, appear critical for generating maximal lexical activation. Evidence of *counter's* influence on labeling begins during it, in the 400ms bin. Its absence in *couner*, where in place of the burst there is nasalization, leads to weaker and slower lexical activation, which does not appear until well into /æ/. This analysis highlights the sensitivity of the perceptual system to acoustic properties of the citation pronunciation of words (Warren & Marslen-Wilson, 1987). At the same time, it demonstrates the tolerance of the perceptual system to the absence of some cues, which enables recognition of pronunciation variants to succeed.

## General Discussion

Across three experiments using two paradigms and three dependent measures, clear and consistent evidence was obtained that words which undergo frequent medial /t/ deletion generate less lexical activation than their citation counterparts. In Experiments 1 and 3, results with the Ganong (1980) paradigm showed that the citation form generates labeling shifts that are larger than the deleted variant. This was obtained twice with the *counter* test set (a third time was not reliable) and twice with the *center* test set. The RT data also showed there was a cost in responding to the deleted variant, producing slower response times on the lexically consistent (/k/ or /s/) side of the continuum.

The results using phonemic restoration reinforce these data. Because lexical influences are detrimental to performance in this paradigm, discrimination should have been worse if the citation form generated the most lexical activation. This is exactly what was found, with better discrimination (larger  $d'$ ) obtained with the deleted than citation form. This result also replicated across two-syllable and three-syllable words.

The time course analysis in Experiment 3 provided details of the delay in processing the pronunciation variant and clarified the consequences of weaker activation of the pronunciation variant. Lexical influences on labeling begin later in time and are uniformly weaker over the course of activation of the variant. As the lower graph in Figure 3 shows, this permanent drop in activation is temporally linked to the realization of [nt] as [ɾ].

The present findings add to a growing literature that suggests immediate lexical activation of pronunciation variants is impaired (Janse et al, 2006; LoCasto & Connine, 2002; Ranbom & Connine, 2007). They build on this work by quantifying the impairment, a reduction in activation on the order of 20–30%, and by mapping the activation time course of the two pronunciations. The one exception in this literature is an experiment by Connine (2004), who used a variant of the Ganong paradigm to study the processing of variants in which intervocalic /t/ flaps (e.g., [bɛɾɔ] vs [bɛtɔ]). Although the experiment was not designed to measure lexical shifts, listeners were biased to a greater degree to make lexically consistent responses when they heard the flapped variant than the citation form, the opposite of what was found in the present study.

Because Connine's results were obtained with multiple word pairs and replicated, they are not likely an anomaly. One cause of the different outcomes may be the frequency with which flapping occurs in some phonological environments (96% of the time; Patterson et al, 2003). For these words, listeners hear the flapped variant almost exclusively, which may provide these forms with an initial processing advantage. The different results across studies may also be attributable to differences in the response collection method. In the current study, fast responding on the button board was paramount. In contrast, Connine had listeners circle on an answer sheet one of the two response alternatives (/b/ or /p/), and instructions did not emphasize response speed. With additional time to process the stimulus, post-perceptual processes may have had undue influence on participants' judgments.

What are the implications of the results for theoretical accounts of how pronunciation variants are recognized? Two approaches they pose challenges for are underspecification (Lahiri, 1999; Lahiri & Reetz, 2003) and phonological inference (Gaskell & Marslen-Wilson, 1998). In underspecification, words are represented lexically as features, but only phonetic features that are marked are represented. Coronal place of articulation is assumed to be unmarked (i.e., default value), which means it is not lexically represented. The virtue of this representational format is that recognition is insensitive to pronunciation variation involving unmarked features, thus ensuring variants are correctly recognized. One implication of this is that processing should not suffer when the underlying (i.e., intended) segment is a coronal, even if the surface realization is different (e.g., dorsal or labial place of articulation). In the current study, coronality varied across the citation and deleted pronunciations (i.e., presence versus absence of /t/). According to underspecification theory, listeners should have been insensitive to this variation because the feature is unmarked. *Couner* should have been processed just as efficiently as *counter*, which is not what was found. The data would seem to argue against the underspecification approach.

In phonological inference (Gaskell & Marslen-Wilson, 1998), listeners apply phonological rules to recover the underlying phonemes in a pronunciation variant of a word. Rule application is triggered when the specific phonological environment is encountered. Medial /t/ deletion is

most often found in a post-stressed syllable containing a reduced vowel (e.g., *counter*), making it a viable context from which to derive a phonological rule. In the current experiment, when *counter* is heard, /t/ should be restored by application of the phonological inference process.

Ideally, phonological inference should operate without a processing cost. Inference must occur in real time for a fast-moving and fleeting signal like speech. If this were true, one might expect lexical activation of *cener* to be identical to *center*. It is also reasonable to assume that inference exacts a cost on processing efficiency. The phonological environment must be identified, and this could include subsequent context. Activation of *cener* would therefore be delayed relative to *center*, as was found in Experiment 3. However, because the inference process should restore the deleted /t/, one would expect asymptotic activation of *cener* to reach that of *center* at a later point in time. That this outcome was not found makes inference seem less plausible, but it may indicate nothing more than inference is inferior to [t] actually being present in the signal. Without additional assumptions like these, which are at a level of detail yet to be articulated in the approach, phonological inference is somewhat inadequate.

How else can the current results be explained? The finding that the citation form generates more robust (earlier and stronger) activation than the /t/-deleted counterpart, despite the latter being experienced much more frequently, also poses a challenge for a strict exemplar account of recognition, in which experience is the primary determiner of lexical formation (Johnson, 2006; Pierrehumbert, 2001). By relaxing this assumption slightly, an account of variant recognition emerges that is in line with the current data.

The ubiquity of the processing advantage for the citation pronunciation over the variant across paradigms, tasks, and stimuli (Janse et al, 2006; Ranbom & Connine, 2007; Sumner & Samuel, 2005) suggests that lexical memory gives priority to perceptual distinctiveness over experience with a particular variant. The more acoustic features that match a lexical entry, the greater its lexical activation. With informationally rich lexical representations, confusions should be minimized. *Center* will not be confused with *sender* or *sinner*. That lexical memory places a premium on distinctiveness is supported by an ever-growing literature that shows word processing is sensitive to subphonemic variation, suggesting that fine-grained acoustic detail, some of it context-conditioned, is encoded in memory (Davis, Marslen-Wilson, & Gaskell, 2002; McMurray, Tanenhaus, & Aslin, 2002; McQueen, Dahan, & Cutler, 2003; Salverda, Dahan, & McQueen, 2003; Whalen, 1981).

Sensitivity to the statistical regularities in the environment is also crucial for memory formation and updating. By itself, a distinctiveness (or tolerance) account would fail to differentiate acceptable pronunciation variants that are encountered often in conversation (*counter*) from mispronunciations (*touner*). A straightforward means of differentiating the two would be to encode in the lexical representation the frequency with which a variant is encountered. This could be implemented in a number of ways. For example, if words are represented lexically as a matrix of features, a path through the matrix that matches the variant pronunciation would represent the variant. Paths would solidify or strengthen as a function of exposure, and gradually differentiate themselves from unexpected pronunciations. In those rare cases where the pronunciation variant is heard almost exclusively (e.g., flapping; Patterson et al, 2001), the representation of the variant might be so well developed that it could reach an activation level comparable to that of the citation form (Connine, 2004; McLennan et al, 2005).

Central to this explanation of variant recognition is learning through exposure. Lexical representations are not static, but develop as a function of experience with the input. The frequency and variety of pronunciation variation in spoken language requires a highly adaptable system. Because current computational models of spoken word recognition lack a learning mechanism, they would overgeneralize, processing pronunciation variants and

mispronounced words in the same way, as comparable decreases in activation. Despite this shortcoming, it is worthwhile to show that these models behave as expected, and can in fact simulate the results found in the current experiment.

The results from a simulation test with the ARTphone model (Grossberg, Boardman, & Cohen, 1997), using default parameter values, are shown in Figure 4. In this simulation, the only word represented in the lexicon was the citation pronunciation of *center*. Inputs to the model were *center* and *center*. Activation functions are identical until the medial /t/, which is to be expected given that the inputs are the same up to this point. As was found with the empirical data, activation of the variant is weaker after this point and remains so by a similar amount until activation dissipates. Also shown are the activation functions for the medial /t/ given the two inputs. The connectivity that gives rise to resonance in ARTphone ensures /t/ will be activated given both inputs, only more strongly when /t/ is actually present in the signal. TRACE (McClelland & Elman, 1986) yields a qualitatively similar pattern of data.<sup>2</sup>

ARTphone was derived from Adaptive Resonance Theory (Carpenter & Grossberg, 2003), a network architecture for learning, but does not have the ability to learn. There are few obstacles to augmenting ARTphone so it could learn, and there is every reason to believe it would perform correctly once enhanced: Neural networks can learn to generalize appropriately in a similar situation, as Gaskell (2003) showed in a network model of regressive place assimilation. Further steps in this direction, in which models learn from the environment, will be needed to mimic the subtleties and complexities of recognizing pronunciation variants.

### Acknowledgments

This work was supported by research grant DC004330 from the National Institute on Deafness and Other Communication Disorders, National Institute of Health. Erik Tracy, Jessica Snider, Sally Blechschmidt, and Samuel Dewitt were a great help in testing participants and scoring data. I also thank James Pooley for running the ARTphone simulation, the Max Planck Institute for Psycholinguistics for support while writing this manuscript, and James Magnuson and an anonymous reviewer for constructive feedback.

### Appendix

Target words used in Experiment 2.

Two syllable	Three syllable
center	accounting
counter	advantage
county	apprentice
lantern	entertain
painted	internet
plenty	interview
printer	pentagon
twenty	percentage
wanted	quantity

<sup>2</sup>These simulation results are available at the author's web site.

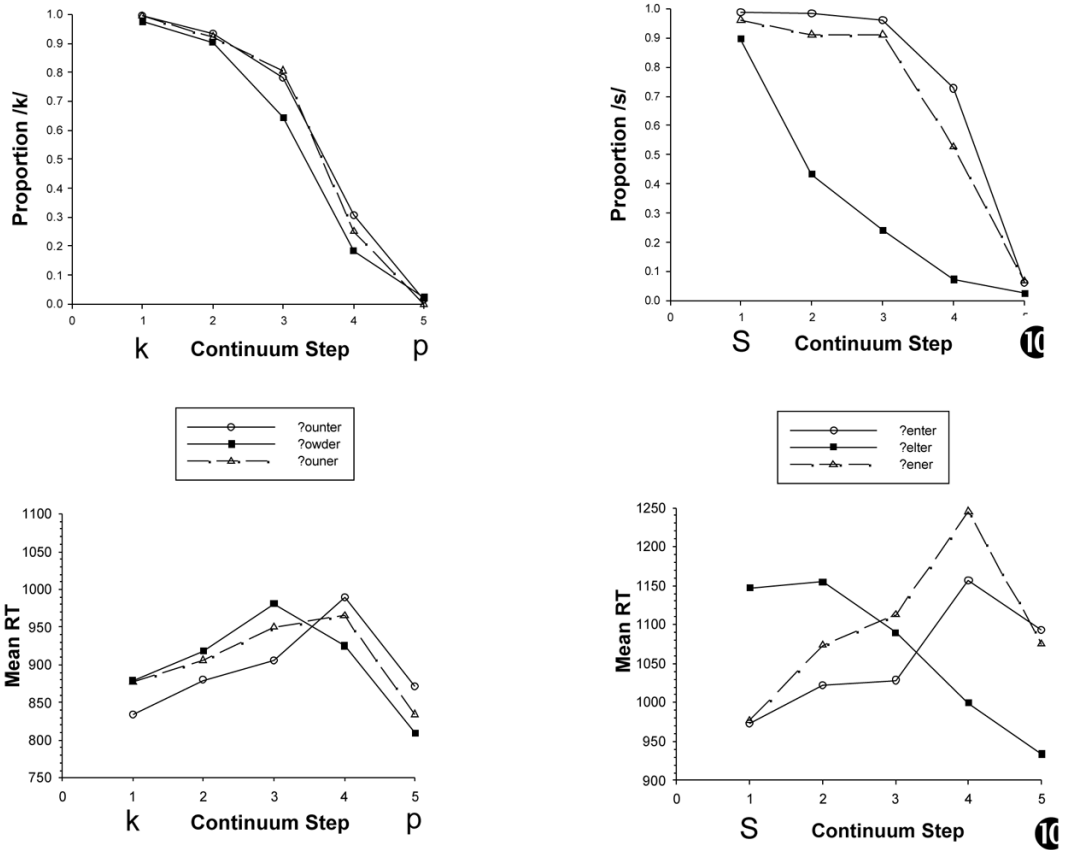
## References

- Bell A, Jurafsky D, Fosler-Lussier E, Girard C, Gregory M, Gildea D. Form variation of English function words in conversation. *Journal of the Acoustical Society of America* 2003;113:1001–1024. [PubMed: 12597194]
- Brent MR, Cartwright TA. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 1996;61:93–125. [PubMed: 8990969]
- Carpenter, G.; Grossberg, S. Adaptive Resonance Theory. In: Arbib, MA., editor. *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press; 2003. p. 87-90.
- Connine CM. It's not what you hear, but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review* 2004;11:1084–1089. [PubMed: 15875980]
- Connine CM, Blasko DG, Titone D. Do the beginnings of spoken words have special status in auditory word recognition? *Journal of Memory and Language* 1993;32:193–210.
- Connine CM, Clifton C. Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 1987;13:291–299. [PubMed: 2953858]
- Davis MH, Marslen-Wilson WD, Gaskell GM. Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 2002;28:218–244.
- Deelman T, Connine CM. Missing information in spoken word recognition: Nonreleased stop consonants. *Journal of Experimental Psychology: Human Perception & Performance* 2001;27:656–663. [PubMed: 11424652]
- Ernestus, M. A corpus-based study of the phonology-phonetics interface. Utrecht: LOT; 2000. *Voice Assimilation and Segment Reduction in Casual Dutch*.
- Ernestus M, Baayen H, Schreuder R. The recognition of reduced word forms. *Brain and Language* 2002;81:162–173. [PubMed: 12081389]
- Ernestus M, Lahey M, Verhees F, Baayen RH. Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America* 2006;120:1040–1051. [PubMed: 16938990]
- Fox RA. Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance* 1984;10:526–540. [PubMed: 6235317]
- Ganong WF. Phonetic categorization in auditory perception. *Journal of Experimental Psychology: Human Perception and Performance* 1980;6:110–125. [PubMed: 6444985]
- Gaskell G. Modeling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics* 2003;31:447–463.
- Gaskell G, Marslen-Wilson WD. Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 1998;24:380–396. [PubMed: 9554092]
- Godfrey, JJ.; Holliman, E. *Switchboard-1 Release 2*. Linguistic Data Consortium; Philadelphia: 1997.
- Grossberg S, Boardman I, Cohen M. Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* 1997;23:483–503.
- Janse E, Nootboom SG, Quene H. Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes* 2007;22:161–200.
- Johnson, K. *Spontaneous Speech: Data and Analysis*. Tokyo: The National Institute for Japanese Language; 2004. *Massive reduction in conversational speech*.
- Johnson K. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 2006;34:485–499.
- Jurafsky, D.; Bell, A.; Fosler-Lussier, E.; Girard, C.; Raymond, W. Reduction of English function words in Switchboard. *Proceedings of the International Conference on Spoken Language Processing (ICSLP-98)*; 1998.
- Lahiri A. Speech recognition with phonological features. *Proceedings of the 14th International Congress of Phonetic Sciences* 1999:715–718.

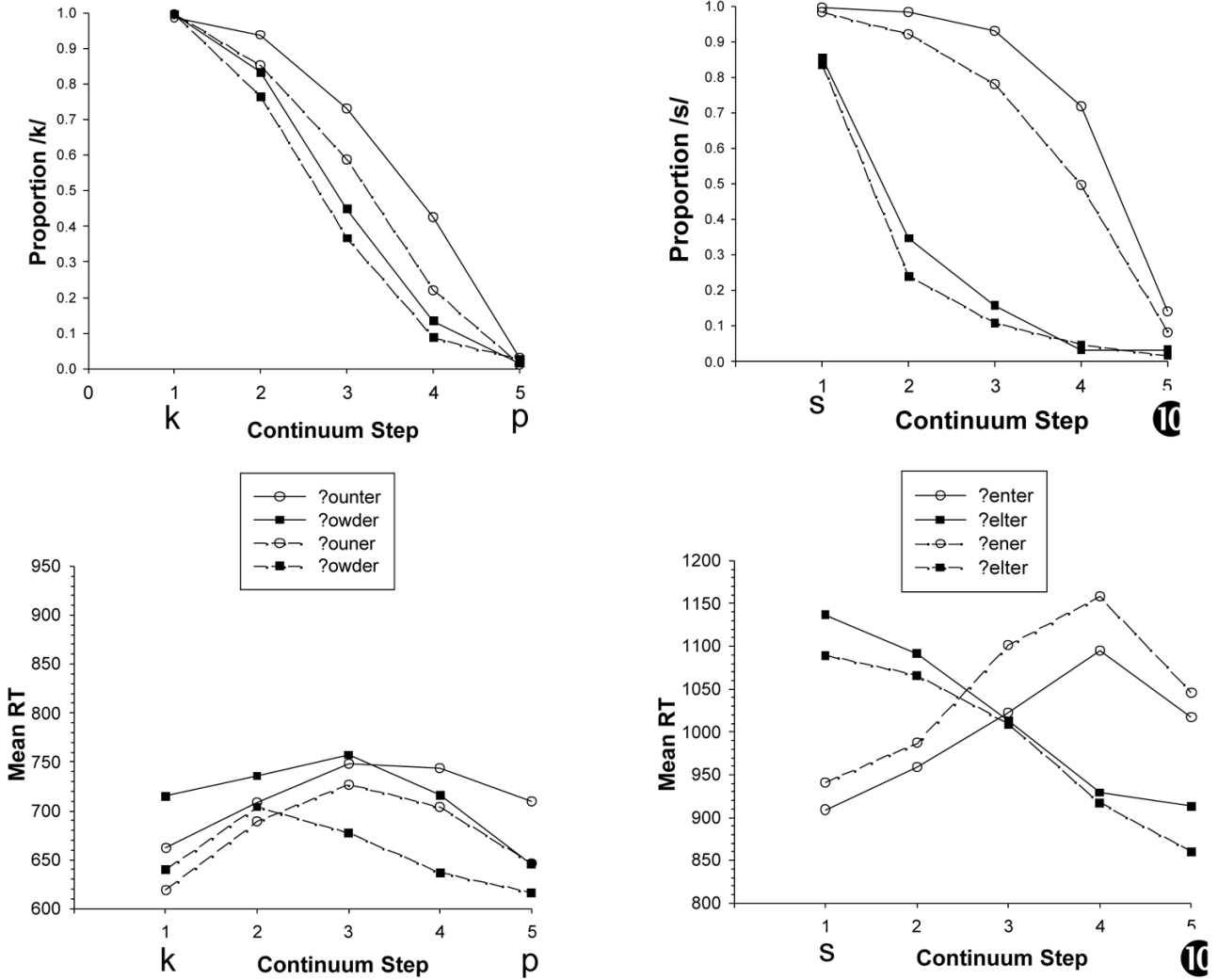


- Lahiri A, Reetz H. Retroflexes and Dentals in the FUL-Model. Proceedings of the 15th International Congress of Phonetic Sciences 2003:301–304.
- LoCasto PC, Connine CM. Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysics* 2002;64:208–219. [PubMed: 12013376]
- McClelland JL, Elman JL. The TRACE model of speech perception. *Cognitive Psychology* 1986a;18:1–86. [PubMed: 3753912]
- McLennan CT, Luce PA, Charles-Luce J. Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory and Cognition* 2003;29:539–553.
- McLennan CT, Luce PA, Charles-Luce J. Representation of lexical form: Evidence from studies of sublexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance* 2005;31:1308–1314. [PubMed: 16366791]
- McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. *Cognition* 2002;86:B33–B42. [PubMed: 12435537]
- McQueen, J.; Dahan, D.; Cutler, A. Continuity and gradedness in speech processing. In: Schiller, NO.; Meyer, AS., editors. *Phonetics and Phonology in language comprehension and production: Differences and similarities*. New York, NY: Mouton de Gruyter; 2003. p. 39-78.
- Marslen-Wilson WD, Nix A, Gaskell G. Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes* 1995;10:285–308.
- Miller JL, Dexter ER. Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 1988;14:369–378. [PubMed: 2971767]
- Neu, H. Ranking of constraints on /t,d/ deletion in American English: A statistical analysis. In: Labov, W., editor. *Locating Language in Time and Space*. New York: Academic Press; 1980.
- Newman RS, Sawusch JR, Luce PA. Lexical Neighborhood Effects in Phonetic Processing. *Journal of Experimental Psychology: Human Perception and Performance* 1997;23:873–889. [PubMed: 9180048]
- Patterson D, Connine CM. Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica* 2001;58:254–275. [PubMed: 11641632]
- Patterson D, LoCasto PC, Connine CM. Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica* 2003;60:45–69. [PubMed: 12802088]
- Pierrehumbert, JB. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J.; Hopper, P., editors. *Frequency effects and emergent grammar*. Amsterdam: John Benjamins; 2001. p. 137-157.
- Pitt, MA.; Dille, L.; Johnson, K.; Kiesling, S.; Raymond, W.; Hume, E.; Fosler-Lussier, E. *Buckeye Corpus of Conversational Speech*. Columbus, OH: Department of Psychology, Ohio State University (Distributor); 2007. (2007; Final release) [www.buckeyecorpus.osu.edu]
- Pitt MA, Johnson K, Hume E, Kiesling S, Raymond W. The Buckeye Corpus of Conversational Speech: Labeling Conventions and a Test of Transcriber Reliability. *Speech Communication* 2005;45:89–95.
- Pitt MA, McQueen JM. Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language* 1998;39:347–370.
- Pitt MA, Samuel AG. An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance* 1993;19:699–725. [PubMed: 8409855]
- Pitt MA, Samuel AG. Word Length and Lexical Activation: Longer is Better. *Journal of Experimental Psychology: Human Perception and Performance* 2006;32:1120–1135. [PubMed: 17002526]
- Ranbom LJ, Connine CM. Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language* 2007;57:273–298.
- Raymond WD, Dautricourt R, Hume E. Word-internal /t, d/ in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 2006;28:55–97.
- Salverda AP, Dahan D, McQueen JM. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 2003;90:51–89. [PubMed: 14597270]

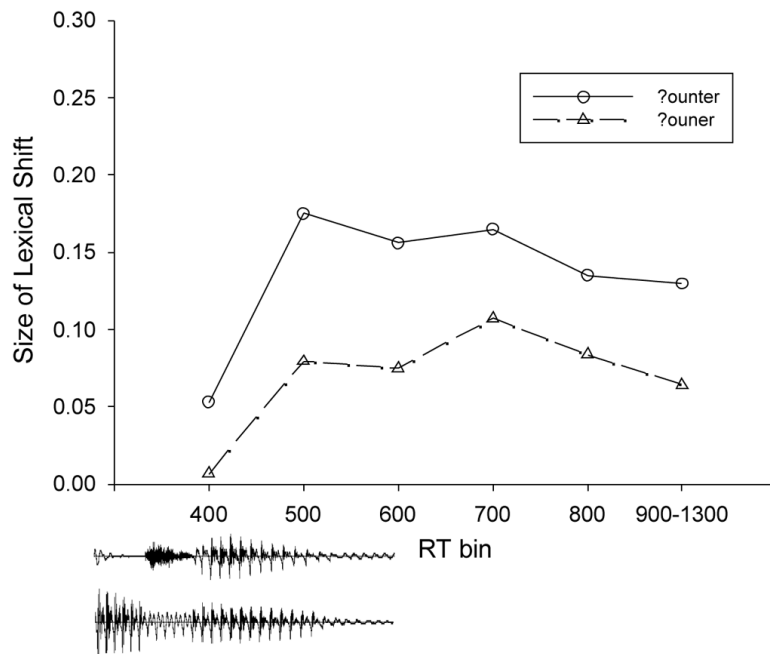
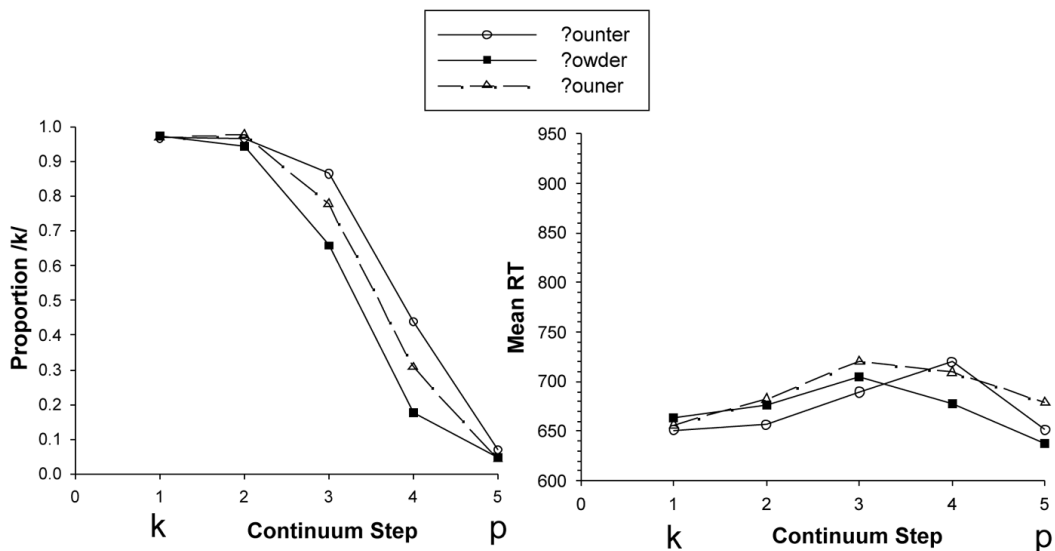
- Samuel AG. Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General* 1981;110:474–494. [PubMed: 6459403]
- Samuel AG. Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General* 1996;125:28–51.
- Sumner M, Samuel AG. Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language* 2005;52:322–338.
- Warren P, Marslen-Wilson W. Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics* 1987;41:262–275. [PubMed: 3575084]
- Warren RM. Perceptual restoration of missing speech sounds. *Science* 1970;167:392–393. [PubMed: 5409744]
- Whalen DH. Effects of vocalic formant transitions and vowel quality on the English [s]-[ʔ] boundary. *Journal of the Acoustical Society of America* 1981;69:275–282. [PubMed: 7217525]



**Figure 1.** Labeling shifts and reaction time functions for the *counter* and *center* test sets.

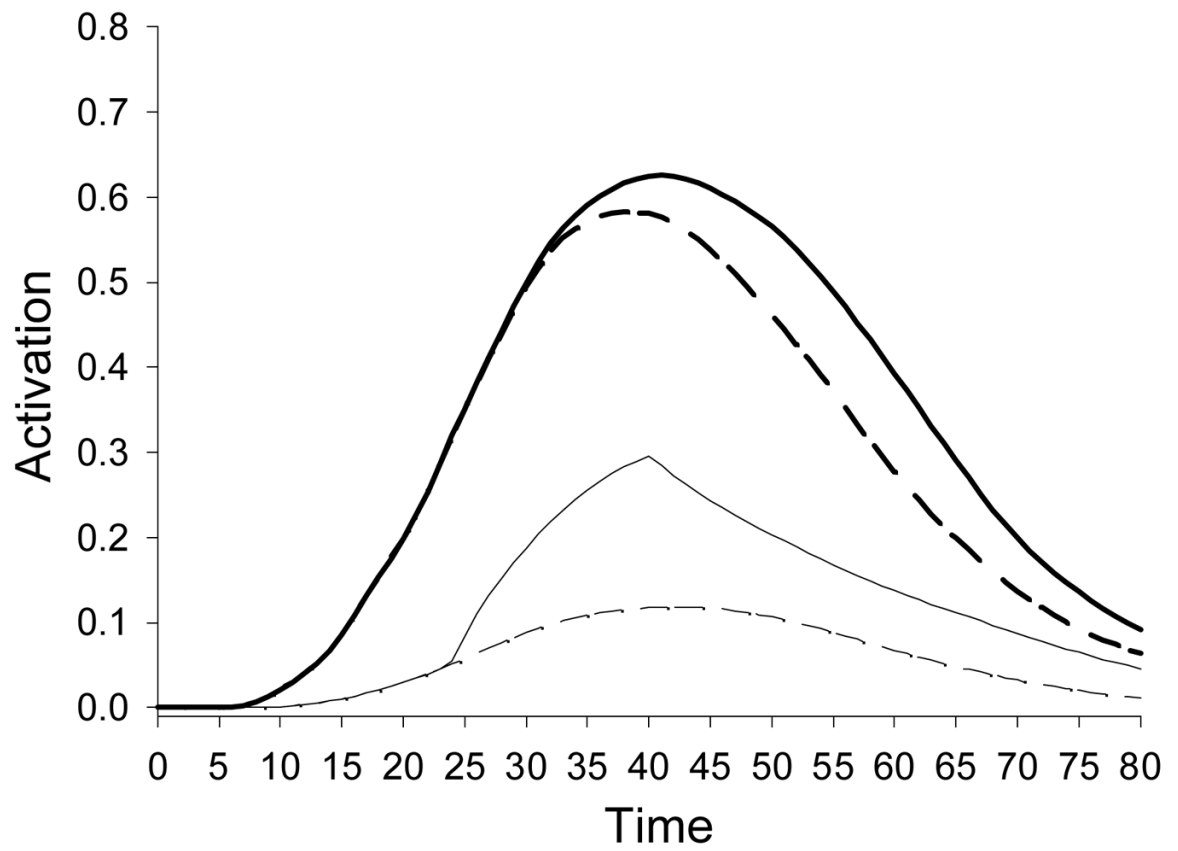


**Figure 2.** Labeling shifts and reaction time functions for a between-subjects replication of Experiment 1a. The data for the *counter* test set are on the left, and the *center* test set are on the right. Because pronunciation of the target word was manipulated between subjects, there are two functions for the reference word (e.g., *owder*), one for the group of listeners who heard the full form of the target word (e.g., *ounter*, solid line) and one for the group that heard the deleted variant (e.g., *ouner*, dashed line).



**Figure 3.** The top two graphs contain labeling shifts and reaction time functions for the *counter* test set. The lower graph displays the size of lexical shifts for counter and ouner across reaction time bin.





**Figure 4.**  
Word and phoneme activation functions in the ARTphone network.

**Table 1**

Mean lexical shift size for citation and deleted forms in Experiments 1a and 1b.

Pronunciation	Center Test Set (center-cener- shelter)		Counter Test Set (counter-couner-powder)	
	1a (within)	1b (between)	1a (within)	1b (between)
Citation	.64	.70	.10	.23
Deleted	.53	.60	.08	.16
Difference	.11 (17%)	.10 (14%)	.02 (20%)	.07 (32%)
RT difference	74ms		38ms	

Table 2

Mean signal detection measures in Experiment 2.

	d'		Beta		False Alarms		Misses	
	Citation	Deleted	Citation	Deleted	Citation	Deleted	Citation	Deleted
Mean	.714	1.160	.702	.696	.650	.487	.142	.146
2-syllable	1.113	1.606	.554	.650	.547	.350	.125	.134
3-syllable	.315	.715	.850	.743	.752	.625	.159	.157