Exploring the role of exposure frequency in recognizing pronunciation variants

Mark A. Pitt[1], **Laura Dilley**[2,3], and **Michael Tat**[1]

[1] Department of Psychology, Ohio State University

[2] Department of Communicative Sciences and Disorders, Michigan State University

[3] Department of Psychology, Michigan State University

Abstract

Words can be pronounced in multiple ways in casual speech. Corpus analyses of the frequency with which these pronunciation variants occur (e.g., Patterson & Connine, 2001) show that typically, one pronunciation variant tends to predominate; this raises the question of whether variant recognition is aligned with exposure frequency. We explored this issue in words containing one of four phonological contexts, each of which favors one of four surface realizations of word-medial /t/: [t], [ʔ], [ɾ], or a deleted variant. The frequencies of the four realizations in all four contexts were estimated for a set of words in a production experiment. Recognition of all pronunciation variants was then measured in a lexical decision experiment. Overall, the data suggest that listeners are sensitive to variant frequency: Word classification rates closely paralleled production frequency. The exceptions to this were [t] realizations (i.e., canonical pronunciations of the words), a finding which confirms other results in the literature and indicates that factors other than exposure frequency affect word recognition.

In casual speech, talkers pronounce words in ways that deviate from their canonical pronunciations. For example, talkers of American English often flap intervocalic /t/s. For some words (e.g., *pretty*), flapping occurs with such frequency that the flapped variant (e.g., [prɪɾi]) is much more common than its citation form (e.g., [prɪti]). For communication to succeed, listeners must learn to recognize these alternative pronunciations of words. How does this occur?

A partial answer to this question is that learning occurs through exposure. Listeners encode the variation they experience to the degree (frequency) they experience it, thereby tuning their perceptual system to the pronunciation variability found in the environment (e.g., Connine, 2004; Connine, Ranbom, & Patterson, 2008; Ernestus & Baayen, 2007; Mitterer & Ernestus, 2006). Although there may be other means of recognizing pronunciation variants (generalization of rules), the ever-growing literature on statistical language learning (Gomez, 2007; Saffran, 2003) and exemplar theoretic models of language perception and production (Bybee, 2001; Johnson, 2006; Pierrehumbert, 2003) suggest that an experience-based account is both plausible and likely.

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, pitt.2@osu.edu, http://lpl.psy.ohio-state.edu.

Research that speaks to the influence of variant exposure on variant processing has examined variation word-finally and word-medially. Although there are inconsistencies across studies in need of resolution, overall, the results suggest a strong link between exposure and recognition. In an analysis of a corpus of Dutch speech, Mitterer and Ernestus (2006) found that word-final /t/ reduction was more frequent after /s/ than /n/. This bias in production was also found in perception: listeners were more likely to report /t/ at the end of a nonword when the preceding segment was /s/ than /n/. Mitterer and McQueen (2009) extended these findings to show that influences of exposure frequency on perception span a word boundary. Word-final /t/ is reduced more when the following word begins with /b/ than /n/, and participants' responses show the same bias.

A number of studies have also examined word-internal pronunciation variation, with the goal of answering processing and representational questions about how recognition of a variant form of a phonological unit (e.g., segment or morpheme) differs from that of the canonical (i.e., phonemic or full) form of the unit. Using counts from an analysis of the Switchboard corpus of American English (Godfrey, Holliman, & McDaniel, 1992), Connine (2004) selected words whose dominant pronunciation of medial /t/ was [ɾ] (e.g., *pretty*). Listeners had to classify the initial phoneme on a word-nonword continuum (e.g., *pretty-bretty;* Ganong, 1980) when the medial /t/ was pronounced as [t] or as [ɾ]. For steps in the middle of the continuum, larger biases in stop labeling from the following context were found for the [ɾ] realization than for the [t] (canonical) realization, suggesting that the flapped variant generated greater lexical activation.

Connine, Ranbom, and Patterson (2008) generalized this finding to words that undergo deletion of an unstressed vowel (*camera -> camra*). Corpus analyses (Patterson, LoCasto, & Connine, 2003) guided the selection of words that underwent vowel deletion frequently or infrequently. When the two groups of words were pronounced without the vowel, lexical decision responses were faster and more accurate to the words that underwent frequent vowel deletion. Just the reverse was found, faster responding to words that underwent infrequent vowel deletion, when the stimuli were spoken with the unstressed vowel. This reversal in responding as a function of the vowel's presence is compelling evidence of the tight coupling between exposure frequency and variant processing. This picture, however, is complicated by the results from studies that have found either no effect of exposure frequency or a violation of exposure frequency, with the canonical pronunciation being processed more efficiently than a more frequent variant. For example, using a paradigm similar to that of Connine (2004), Pitt (2009) found the canonical pronunciation of a word that undergoes frequent /t/-deletion (e.g., *center* spoken as [sɛntɚ]) generates greater lexical activation than the more frequent /t/-deleted variant (e.g., *center* spoken as [sɛnɚ]). Although the different type of variation in the two studies (flapped /t/ versus deleted /t/) may partially explain the discrepancy, the advantage for the canonical form has been reported in other studies, including ones using flaps (Ernestus & Baayen, 2007; McLennan, Luce, & Charles-Luce, 2003, 2005; Tucker & Warner, 2007). Relatedly, Janse, Nooteboom, and Quené (2007) investigated word-final /t/ reduction in a fixed /st#b/ context. In a corpus of spoken Dutch, they found that unreleased word-final /t/ occurred more frequently than released word-final /t/; however, in corresponding perception experiments they found a null effect of variant frequency on processing of the two types of variants.

Taken together, these findings suggest that the relationship between variant frequency and variant recognition is not straight forward. In particular, the canonical form of a word sometimes violates what would be expected on the basis of a purely frequency-of-exposure account. The current study sought to build on this body of work in two ways. One was to expand the scope of inquiry to variants of words that are rarely heard (from common to uncommon) for the purpose of creating a more detailed profile of variant processing than

has been obtained to date. These data can then be compared against their frequency of usage in the language to permit a thorough evaluation of the exposure-frequency hypothesis. The more closely recognition correlates with usage, the greater the support for the hypothesis.

The second aim of the study was to consolidate findings in the literature concerning processing of different variant types in a single study. Differences in variant processing have been found across studies, using different types of variation, and using various methodologies. By examining the processing of multiple forms of variation in the same experiment, we eliminated many of these sources of potential variability and hoped to develop a clearer picture of how variant frequency relates to variant processing.

To achieve these goals, we studied the processing of words with medial /t/ variation. The many allophones of /t/ make this phoneme ideal for testing the exposure hypothesis and for examining the issue of consistency across types of variation. Word-medial /t/ can be realized as [t], a glottal variant [ʔ], a flap [ɾ], or /t/ can be deleted, denoted here as [.]; processing of the glottal variant, [ʔ], is particularly under-studied. Furthermore, different phonological environments favor one allophone (i.e., type of variation) over another (McMahon, 2002; Raymond, Dautricourt, & Hume, 2006; Shockey, 2003), making a given phonological context variably conducive to each of the allophones of /t/. Taking this reasoning one step further, for each phonological context, which also corresponds to a unique set of words, variants can be rank-ordered in terms of phonological conduciveness. According to the exposure hypothesis, those ranked high in terms of phonological conduciveness should be recognized easily (since they readily occur in the language), whereas those ranked low should not.

We tested this proposal not just in one phonological context (i.e., one rank-ordering of conduciveness to the four allophones) but in four distinct phonological contexts. Each of these contexts is most conducive to one of the allophones, creating an experimental design with 16 conditions (four contexts by four allophones). Experiment 1 established the frequency with which each allophone is produced in American English in words containing the four phonological environments. These data were then used to inform stimulus selection and make comparisons in Experiment 2, in which recognition of the variants was measured.

Experiment 1

Experiment 1 aimed to establish the frequency with which the four allophones of word-medial /t/ ([t], [ʔ], [ɾ], and [.]) occur in each of four phonological environments. The environments were selected because they have been described as favoring (i.e., being especially conducive to) production of one of the four variants. A memory-demanding production task was used to elicit speech that was sufficiently casual to generate allophones of /t/.

Participants—Participants were 40 undergraduates enrolled in Introductory Psychology at Ohio State University. American English was their native (first) language and no one reported hearing difficulties. All were naïve to the purposes of the experiment.

Stimuli and Design—Four sets of 22 lexical items were identified which contained a word-medial /t/ in one of four distinct types of phonological environment that were expected to facilitate elicitation of predominantly one of the four types of medial /t/ allophone: [ʔ], [ɾ], [t], or [.]. These phonological environments were derived both from published descriptions of phonological contexts associated with distinct word-medial /t/ variants for American English (e.g., McMahon, 2002; Raymond et al., 2006; Zue & Laferriere, 1979), as well as counts collected for the present study on frequency of usage of different word-

medial /t/ variants in lexical items drawn from the Buckeye Corpus of conversational speech (Pitt et al, 2007). For the phonological environment predicted to favor [t] ("Favors [t]"), the /t/ occurred in poststress position at the onset of an unstressed syllable and was preceded by a voiceless stop consonant or voiceless fricative or /l/. For the phonological environment predicted to favor [ɾ] ("Favors [ɾ]"), /t/ occurred in poststress, intervocalic position; note that for some items, the preceding vowel was a rhotic diphthong. Also, the following syllable lacked /n/ (though for two words the following syllable had another nasal, /ŋ/ - *getting, meeting*). Next, for the phonological environment predicted to favor [ʔ] ("Favors [ʔ]"), /t/ occurred in poststress position before an unstressed syllable containing /n/; for all these items, the preceding phoneme was also voiced. Finally, for the phonological environment predicted to favor [.] ("Favors [.]"), /t/ occurred in poststress position after /n/, so that this /t/ was expected to often be deleted in lieu of producing a nasal flap for the word-medial consonant sequence (Raymond et al., 2006). The following syllable in these items lacked a nasal, with one exception (*mounting*). A sentence frame was constructed for each of the lexical items; see Appendix for materials. The 88 target sentences were combined with 106 filler sentences, the main purpose of which was to mask the repetitiveness of the orthographic and phonological structure of target items. One stimulus list was created by randomly permuting the sentences, with the one constraint that three target sentences were not adjacent. To counterbalance order of presentation, a second list was created by reversing the order of items in the first one. Equal numbers of participants were randomly assigned to both lists.

Procedure—Participants were tested individually in a sound-dampened room. They sat in front of a computer monitor and microphone. On each trial, participants were given 2.5 seconds to read a short sentence on the computer screen. It was then erased and three seconds later a single, semantically related word appeared (corresponding to items in parentheses in the Appendix). Participants were instructed to remember the sentence, and then upon seeing the additional word, to integrate it with the sentence to form a new sentence, and speak it aloud into the microphone. The recording window was six seconds in duration; the next trial began two seconds after the window ended. The experiment began with a 40-trial warm-up session, the purpose of which was to acclimate participants to the experimental session so as to induce a casual speaking style by the time test sentences were presented. Participants were given one break half-way through the experiment.

Analysis—Two trained phonetic analysts used the symbol set from the Buckeye Corpus to phonetically transcribe target words in the produced speech. Both analysts had experience labeling phonetic properties of spontaneous speech using these conventions, and both were naïve to the purposes of the study. Based on phonetic labels, each token was assigned to one of the four allophonic variant categories ([t], [.], [ɾ], [ʔ]). Good correspondence was observed between the analysts in rates with which tokens of items were assigned to the four variant categories (mean $r^2 = 0.89$).[1] To find a single value characterizing rates of producing /t/ allophones for each experimental item, the average rate of token assignment to these categories for the two analysts was determined.

Results and Discussion

Table 1 reports means and standard deviations in rates of allophone usage across items in the four phonological contexts. Of interest are the rates of realization for a given context

---

[1]Tokens of word-medial /t/s which were assigned a phonetic label of "d" were coded as [t] due to the phonetic similarity of unaspirated /t/ and /d/, while tokens which were assigned a label of [ch] were also coded as [t] due to the decomposability of the voiceless affricate into /t/ and /ʃ/. Agreement was determined by calculating correlation coefficients across target item types in rate of assigning tokens as [t], [.], [ʔ], or [ɾ] and taking the mean correlation coefficient across these four categories.

(column). Looking down the columns, for three of four phonological contexts – those favoring [t], [ɾ], and [ʔ] – there was a single surface allophonic realization that predominated, corresponding to the variant favored by the respective phonological context. For example, in the Favors [t] context, [t] was the dominant surface realization, with 86% of tokens in this category showing [t]s. For Favors [.] context, realizations of allophones were closely split between two types: full [t] (54%) and deletions (46%).2

Looking across rows, we can infer the degree to which each surface realization is specific to a phonological context. [t] and [.] occur in each of the four phonological contexts to some (nonzero) degree, although [t] is relatively uncommon in contexts favoring [ɾ] or [ʔ]. In contrast, [ʔ] and [ɾ] are highly restricted with respect to the phonological environments in which they occur. For example, [ɾ] occurs very seldom or never in phonological environments other than the environment in which it is favored. The glottal allophone, [ʔ], shows a similar pattern in that it appears to occur hardly ever in phonological environments other than that in which it is the favored variant.3

To address how successful our production task was at eliciting casual speech, data in Table 1 were compared with the realization of words which occurred in an analysis of 19 talkers from the Buckeye Corpus. We focused on a subset of 24 of the experimental items that occurred at least 14 times in that corpus (mean *n* = 50; these are marked with asterisks in the Appendix). The data are shown in Table 2, and are remarkably similar to those in Table 1, indicating that the production study was quite successful in eliciting casual speech. Looking down columns, for all four phonological contexts there was a single type of surface allophonic realization that heavily predominated, and that variant type corresponded to the one favored by the respective phonological context. (Note that there was only one item with the minimum specified frequency for the phonological context favoring [ʔ] in the corpus, so these numbers cannot be generalized and no standard deviation could be calculated.)

One difference across tables is that the phonological context favoring [.] shows a predominance of [.] realizations and many fewer instances of [t] than in the production study. This suggests that spontaneous speech was somewhat more casual than the production-study speech. Moreover, looking across columns, [t] and [.] again occur in each of the four phonological contexts to some (nonzero) degree. In contrast, [ʔ] and [ɾ] are once more highly restricted with respect to the phonological environments in which they can occur.4

Another means of demonstrating the close correspondence between the production and corpus data is to correlate the rate of allophone use for each of the four realizations of /t/.

2Across individual items in phonological contexts favoring [t], [ɾ], and [ʔ], a single phonetic variant predominated (i.e., occurred in ≥ 65% of productions of the item), which was the one favored by the respective phonological context. The one exception was the word *sentence* in the Favors [ʔ] category, which showed 56% [ʔ] productions (and 38% [t] productions). In contrast, for items in the Favors [.] category, 11 items showed a pattern of [t] predominating (i.e., ≥ 65% of productions of the item had that variant). Likewise, eight items (*center, counter, county, fantasy, interview, plenty, twenty, wanted*) showed a pattern of [.] predominating while the remaining three items (*encounter, ninety, pointed*) showed more equal distributions between [.] and [t] variants.
3Note that for surface realizations of [ʔ] in the corpus in Table 2, the mean and standard deviation of the rate of [ʔ] realizations in phonetic contexts predicted to favor [ɾ] is somewhat higher than for the production experiment; however, this is due entirely to one item, *getting*, which is the only item in this category that also had a nasal in its second syllable, a phonetic attribute conducive to [ʔ]. When this item is removed, the mean and standard deviation in rate of [ʔ] usage in the 'Favors [ɾ]' environment both drop to ~0%, suggesting even more similarity with the production study results.
4A full characterization of rates with which variants are associated with particular phonological environments would require determining the probability with which each of these environments occurs in the ambient language (cf. Bayes' Rule), as well as the rates in which they occur in all other contexts. The former is known to be a very challenging issue (Goldwater, 2007), and the latter would require a full acoustic-phonetic analysis of a speech corpus to identify how often similar phonetic variation occurred in other phonological environments. These analyses were outside the scope of the present study, but it seems reasonable to suppose that phonological contexts are relatively comparable in their frequency in the language.

Correlation coefficients were calculated for the 24 words shared in the data sets. This analysis showed good agreement, with the mean correlation being 0.87.

In summary, the data from the production task and the corpus analysis are in good agreement in establishing a baseline of /t/ allophone use in this dialect of American English in the four phonological contexts. By far the most frequent surface realization is that favored by the phonological context (although the data are somewhat equivocal for [.]). The results are equally clear about the types of realizations that are rare or virtually nonexistent. Under the hypothesis that listeners encode the frequency of variant usage (e.g., Pierrehumbert, 2003), recognition of the variants should closely mirror their production frequency. This hypothesis was tested in Experiment 2 by having listeners make lexical decision judgments to words from all 16 cells.

Experiment 2

Method

Participants—Participants were 64 undergraduates enrolled in Introductory Psychology at Ohio State University. American English was their dominant language and no one reported hearing difficulties.

Design—A 4 (phonological context) × 4 (surface realization) repeated measures design was used. As in Experiment 1, phonological context favored one of four allophones of word-medial /t/: [t], [ɾ], [ʔ], or [.]. The second factor was the surface realization of the word-medial /t/, with four levels: [t], [ɾ], [ʔ], and [.]. Four stimulus lists were constructed, each with four conditions; within each list, the conditions that were created by pairing level of phonological context with level of surface realization were counterbalanced using a Latin Square design, so that the items in each condition appeared in only one list. There were thus 16 pronunciation conditions (4 levels of phonological context × 4 levels of surface realization). Sixteen participants were randomly assigned to each of the 4 lists.

Stimuli—The selected experimental stimuli consisted of 72 isolated words (68 bisyllabic, 4 trisyllabic) from Experiment 1, which corresponded to the first 18 items in each of the four phonological context conditions in the Appendix. Filler items (228) were included to mask the manipulation of pronunciation variation, which occurred primarily in bisyllabic words. There were 92 monosyllabic and 92 trisyllabic utterances, approximately equally split between words and nonwords (created by changing between one and three phonemes in English words, depending on length). Because it was not known beforehand how listeners would hear all of the bisyllabic variants, 44 bisyllabic nonwords (created as described above) were also included to ensure listeners would classify at least some bisyllables as nonwords.

Several tokens of each stimulus were recorded by a male talker onto DAT using a Tascam DA-30MKII DAT recorder connected to an N/D308A cardioid microphone via a Yamaha MV802 mixer at 48 kHz. Uncommon pairings of surface realization and phonological context were rehearsed multiple times to ensure fluent pronunciation. The recordings were digitally transferred to a PC and downsampled to 16 kHz with lowpass filtering applied at 7.8 kHz to prevent aliasing. Tokens of target words were checked for accuracy in pronunciation by the second author. Moreover, a number of steps were taken to ensure that tokens representing distinct levels of the two independent variables (phonological context and surface realization) differed from one another only on the critical dimensions, and not in extraneous ways. For example, across items, care was taken to ensure similar pronunciation of the same word with different phonetic variants: unstressed syllables were pronounced

with reduced vowels, words had similar global pitch and rhythmic characteristics, and any creaky voicing occurred only at the end of the word.

In addition, care was taken to ensure that the acoustic realizations of a given variant type were similar across phonological contexts, to prevent potentially confounding systematic covariation between precise phonetic realization of a variant and phonological context type. For example, given that [ɾ] has several acoustic manifestations (de Jong, 1998), consistency across contexts was maximized by selecting productions of [ɾ] evidencing a closure plus short burst. To further ensure consistency in precise phonetic realizations of variants across phonological contexts, an acoustic analysis of relevant phonetic characteristics (closure duration, burst duration, duration of irregular voicing) of (non-deleted) variants was undertaken. Table 3 reports closure duration and VOT across all four phonological contexts for surface realizations of [t] and [ɾ]. For both allophones, there were no significant differences across phonological contexts in closure duration ([t]: $F(3,68) = 2.038$, $p < .117$, [ɾ]: $F(3,68) = 2.148$, $p < .102$). Moreover, there was no significant difference across phonological contexts in VOT for [ɾ], $F(3,48) = .592$, $p < .623$, but there was a significant difference in this variable for [t], $F(3,68) = 6.538$, $p < .001$; post-hoc Tukey's HSD tests revealed a shorter VOT for the environment of Favors [t] compared to the other three groups ($p < .05$), but no additional differences. Moreover, Table 4 reports the duration of intermittent irregular voicing in the waveform, defined as the total duration of periods of silence and/or nonmodal voicing, in the region of word-medial /t/ for surface realizations of [ʔ]; no significant differences across phonological contexts in this dependent measure were observed, $F(3,68) = 1.622$, $p < .192$.

All tokens of target words were saved as separate sound files. With four realizations of each of the 72 target, one for each of four surface realizations of word-medial /t/, there were 288 target stimuli.

Procedure—Participants were tested in groups of four in sound-dampened rooms. They sat in front of a computer keyboard and LCD monitor, and were instructed to press one of two keys on a button board to indicate whether the utterance heard over headphones was a "word of English" or a "nonsense word." The instructions stressed fast and accurate responding. A computer controlled stimulus presentation and response collection. There was a 2.5 second timeout after stimulus offset. A two-second pause preceded presentation of the next word. Twenty-four practice trials preceded the 300 test trials. The experiment lasted 50 minutes.

Results and Discussion

Data analysis began by focusing on the frequency with which the items across conditions were classified as words. For each item, the percentage of *word* classifications was calculated. Condition means were then computed, and these are shown in Table 5. Of the 288 realizations, 12 could be heard as another word of English (e.g., winter -> winner; sweeten -> Sweden). Responses to these items were removed from the data because it is unclear which of the two words listeners heard (i.e., either interpretation would elicit a *word* response).

As in Experiment 1, comparisons of interest are the cells within each column. Separate logit mixed-effects models (Jaeger, 2008), as implemented in the lme4 package (Bates & Maechler, 2009; R core development team, 2009), were used to analyze the data in the four cells in each favored phonological context. Preliminary analyses indicated that items and subjects should be treated as random factors in all models. Realization served as the fixed factor. The results of likelihood ratio tests showed that realization improved the fit of all models over the model with only random factors, indicating a statistically significant effect

of realization on percentage of *word* classifications ([.]: $\chi^2(3) = 41.56$, *p*<.001; [t]: $\chi^2(3) = 112.25$, *p*<.001; [ɾ]: $\chi^2(3) = 83.73$, *p*<.001; [ʔ]: $\chi^2(3) = 97.57$, *p*<.001). Comparisons between the conditions in each column were conducted by repeating the mixed-model analysis multiple times to obtain all pairwise comparisons. Letter subscripts on the cell means indicate which conditions were statistically different from one another.

The classification results provide some clear evidence that variant recognition mirrors frequency of variant usage in the language. Across three of the phonological contexts ([t], [ɾ], [ʔ]), the data pattern is similar to those in Tables 1 and 2, in several respects. Items containing favored realizations (diagonal) were classified as words greater than 92% of the time. Just as importantly, classification rates were much lower with uncommon realizations of /t/ in these environments. In the Favors [ʔ] context, words with [.] and [ɾ] realizations were labeled words less than 22% of the time. In the Favors [ɾ] context, the uncommon pronunciation of [.] was heard as a word a similar amount of time (19%); when the pronunciation was [ʔ], reports of words increased to 47%, but still half that found with the favored pronunciation (94%).

The Favors [t] context also shows impressive listener selectivity in what realization of /t/ counts as a word, with classification of the [.] and [ʔ] realizations dropping below 20%. The high classification rate to words with the [ɾ] realization (86%) in this context is likely due to the fact that [ɾ] in most items was realized as an interval of closure duration, followed by a burst release, consistent with observed data on acoustic realization of American English [ɾ] (de Jong, 1998; Zue & Laferriere, 1979). These characteristics thus show similar acoustics to what one would expect for the realization of /t/ preceding an unstressed syllable lacking a nasal (Zue & Laferriere, 1979), and we attribute the high word classification rate to this acoustic similarity.

The results in the Favors [.] context show the trend found in the other three phonological contexts, only it is weaker. Classification of words containing the favored realization, [.], was reliably lower than that with [t], and although it was 13% higher than words with [ʔ] as the surface form, the difference from [.] was not significant. Inspection of responses to words with /t/ realized as [.] identified three (*mounting, ninety, rental*) that contributed most to this lower word classification rate. Their removal increases the mean for the favored realization in the Favors [.] context to 88%.

The one outcome that consistently violates the predictions of an exposure account are the high classification rates to the [t] realizations in the [.], [ɾ], and [ʔ] phonological contexts. On the one hand, good recognition of these canonically pronounced words is to be expected because clear, robust, and veridical segmental cues to the /t/ are present. On the other hand, based on exposure frequency alone (Table 2), they should be classified as words no greater than other, similarly infrequent surface realizations of /t/.

The low word classification rates in many of the cells prevented performing an equivalent analysis on the reaction time (RT) data. The exceptions to this are the cells in which classification rates are reasonably high, which are the favored and [t] realizations for each phonological context. Mean RTs to words in these conditions, measured from word onset, are shown in Table 6. Statistical analyses were again carried out using a mixed-effects model on the data in each favored condition, with log RT as the predicted variable, subjects and items as random factors, and realization as a fixed factor. Word duration and the frequency with which the dominant variant occurred in Experiment 1 were also added as predictors. The latter was used instead of lexical frequency given the focus of this study; variant frequency should predict reaction time to the extent that listeners encode the

frequency of variant realization. RTs less than 200 ms or greater than 2000 ms were removed from the data (<1%).

On the basis of the data in Tables 1 and 2, one might expect responses to be fastest to the favored realization because it is the most frequent. A consideration of the classification data might suggest this prediction should be modified, with RTs to the canonical and favored realizations being comparable because their classification rates are comparable. In the phonological contexts favoring [.] and favoring [ʔ], neither of these predictions holds. Instead, there is an RT advantage for [t] realizations of 57 ms and 54 ms, respectively. In the context favoring [ɾ], the pattern of RTs reverses across conditions, being an average of 28 ms faster to words with [ɾ] than [t].

Despite the sizeable differences in RTs, statistical results were mixed. For the Favors [ʔ] and Favors [ɾ] contexts, realization did not improve model fit significantly (Favors [ɾ]: $b=-.0135$, $p=.724$; Favors [ʔ]: $b=.0549$, $p=.1544$), whether alone or with other factors added to the model (e.g., variant frequency, word duration);. In contrast, for the Favors [.] context, the best fitting model included all three factors (realization: $b=.0784$, $p=.045$; variant frequency: $b=-.1209$, $p=.007$; word duration: $b=.28$, $p=.001$). The effect of variant frequency shows that RTs were faster to variants that underwent deletion more often, a finding reported by Connine et al, 2008. Not surprisingly, word duration was also reliable in the other two contexts, showing that response time slowed as stimulus duration increased (Favors [ɾ]: $b=.2642$, $p=.02$; Favors [ʔ]: $b=.2706$, $p=.001$), but variant frequency was not significant in either context.

One reason the effects of realization were weak in two of the favored conditions is item variability. Although the majority (>70%) of items trended in the same direction, there were unusually large reversals for a few items, which contributed an inordinate amount of variance to the data. For example, there was a 144 ms RT slowdown for the flapped variant of *letter* over its citation form. At the other extreme, there was a 178 ms speed up for the flapped variant of *meeting*. Similarly wide swings in effect magnitude are present in the Favors [ʔ] data, suggesting that properties of words besides their phonological context and manner of reduction have a significant influence on classification speed.

Although the RT data are equivocal, they trend like those in past studies. In the case of deleted variants, Ranbom and Connine (2007; Pitt, 2009) found that RTs were slower to deleted variants (which Ranbom and Connine refer to as 'nasal flaps') relative to the canonical [t] form, which is what was found in the present experiment. In the case of (oral) flapped variants, [ɾ], the reverse pattern was obtained here, with RTs being faster to words spoken with a flap. Tucker and Warner (2007) report the same result, and Connine (2004) found a response bias in classifying the flapped variant in a phoneme identification task. Taken together, these findings show that there is consistency across studies for a particular type of variation, and that all forms of variation are not processed identically. The RT results also extend our understanding to word-medial glottal variants, suggesting that they pattern like deleted variants.

General Discussion

The goal of the present study was to explore further the simple yet powerful idea that recognition of pronunciation variants can be explained in part by the frequency with which the listener hears a variant spoken. The results of the two experiments begin to clarify the nature of this relationship. The results of our production study (Experiment 1) demonstrate that the distribution of pronunciation variants is tightly restricted, with particular variants dominating certain phonological environments and rarely occurring in others. Lexical decision data (Experiment 2) show that listeners are exquisitely sensitive to how /t/ is

realized in a particular word. Only realizations that are common, as determined by the counts in Experiment 1, are consistently classified as words. Forms of /t/ reduction that are rare, even though the same allophone is recognized clearly in other contexts, lead to successful recognition much less often. That better recognition with higher variant frequency was found across multiple phonological contexts demonstrates that this is a stable finding, lending support to the exposure-frequency hypothesis.5

The data also suggest that the link between exposure frequency and recognition is not simple. Listeners categorize the canonical pronunciation almost perfectly, even though it is apparently rarely spoken. Although less conclusive statistically across contexts, RTs were faster to the much less frequent [t] realization of words in the Favors [.] and Favors [ʔ] contexts. This advantage for the canonical pronunciation has been reported by others (Ernestus & Baayen, 2007; McLennan et al., 2003, 2005; Pitt, 2009; Tucker & Warner, 2007), and suggests that factors other than variant frequency affect the speed of recognition.

There are two reasons why [t] might be processed differently. Ranbom and Connine (2007) suggest that the lexical representation of written forms of words influences encoding of the spoken form, to the point of facilitating recognition when /t/ is pronounced canonically. Another possibility is that the distinctiveness provided by [t], in distinguishing it from phonetically similar words, outweighs exposure frequency in some circumstances. That is, successful recognition depends on discriminating words from one another. To the extent that [t] provides clarity (Tucker & Warner, 2007), it is encoded in the lexical representation of the word to aid recognition. Although the current data cannot decide between these two alternatives, they further confirm the presence of the exception and the need for an explanation of it.

A possible limitation of the current study is that variant recognition was tested in isolation whereas the frequency of producing the variants was estimated in sentences. If the canonical form is produced more often in isolation, the production and categorization data (Tables 2 and 5) might resemble each other more closely. For example, two variants might dominate in the Favors [ʔ] and Favors [ɾ] contexts. If Experiment 2 were carried out with the tokens embedded in sentences, how might the results change? We suspect that the sentential context would make listeners more accepting of *all* pronunciation variation, so that word classification rates in the cells not close to ceiling would increase. This prediction is based in part on listeners' poor ability to detect mispronunciations in words when the altered phoneme occurs in non-word initial positions (Marslen-Wilson & Welch, 1978).

In sum, the current study provides a profile of listener attunement to pronunciation variation across multiple realizations of /t/ in American English. The large amount of data generated replicates consistencies and inconsistencies in the literature, and extends these findings to new surface realizations and phonological contexts. Together, they show that classification aligns well with production frequency, but not when the surface realization is [t]. The challenge going forward is to explain the cause of these anomalies.

Acknowledgments

5To ensure the results were not due to use of the lexical decision task, Experiment 2 was replicated by having listeners type what they heard on a computer keyboard. The proportion of words in each condition spelled correctly (with clear typographical errors corrected) was the dependent measure. The results closely resemble those in Table 2, except that use of the open response set led to lower values when the surface realization was [.] (by an average of 15%) and higher values when the surface realization was [ʔ] (by an average of 13%). These data can be found at http://lpl.psy.ohio-state.edu/publications.html

References

Bates D, Maechler M. Ime4: Linear mixed effects models using S4 classes. 2009

Bybee, J. Phonology and language use. Cambridge: Cambridge University Press; 2001.

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(6): 1084–1089. [PubMed: 15875980]

Connine CM, Ranbom LJ, Patterson DJ. Processing variant forms in spoken word recognition: The role of variant frequency. Perception & Psychophysics. 2008; 70(3):403–411. [PubMed: 18459250]

de Jong K. Stress-related variation in the articulation of coda alveolar stops: flapping revisited. Journal of Phonetics. 1998; 26:283–310.

Ganong WF. Phonetic categorization in auditory perception. Journal of Experimental Psychology: Human Perception and Performance. 1980; 6:110–125. [PubMed: 6444985]

Godfrey, J.; Holliman, E.; McDaniel, J. SWITCHBOARD: Telephone speech corpus for research and development. Proceedings of the IEEE International Conference on Acoustics, Speech, & Signal Processing (IEEE ICASSP-92); San Francisco: IEEE; 1992. p. 517-520.

Goldwater, S. Unpublished PhD Dissertation. Brown University; 2007. Nonparametric Bayesian models of lexical acquisition.

Gomez, R. Statistical learning in infant language development. In: Gaskell, GM., editor. The Oxford Handbook of Psycholinguistics. New York: NY: Oxford University Press; 2007.

Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. Journal of Memory and Language. 2008; 59:434–446. [PubMed: 19884961]

Janse E, Nooteboom SG, Quene H. Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. Language and Cognitive Processes. 2007; 22(2):161–200.

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

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]

McMahon, A. An Introduction to English Phonology. New York: Oxford University Press; 2002.

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

Mitterer H, McQueen JM. Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. Journal of Experimental Psychology: Human Perception and Performance. 2009; 35:244–263. [PubMed: 19170486]

Patterson DJ, Connine CM. A corpus analysis of variant frequency in American English flap production. Phonetica. 2001; 58(4):254–275. [PubMed: 11641632]

Patterson DJ, LoCasto P, Connine CM. A corpus analysis of schwa vowel deletion frequency in American English. Phonetica. 2003; 60:45–68. [PubMed: 12802088]

Pierrehumbert J. Phonetic diversity, statistical learning, and acquisition of phonology. Language and Speech. 2003; 46(2–3):115–154. [PubMed: 14748442]

Pitt M. The strength and time course of lexical activation of pronunciation variants. Journal of Experimental Psychology: Human Perception and Performance. 2009; 35:896–910. [PubMed: 19485698]

R core development team. R: A language and environment for statistical computing. Vienna, Austria: 2009. Retrieved from http://www.R-project.org on May 24, 2010

Ranbom LJ, Connine CM. Lexical representation of phonological variation in spoken word recognition. Journal of Memory and Language. 2007; 57:273–298.

Raymond W, Dautricourt R, Hume E. Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. Language Variation and Change. 2006; 18:55–97.

Saffran JR. Statistical language learning: Mechanisms and constraints. Current Directions in Psychological Science. 2003; 12:110–114.

Shockey, L. Sound Patterns of Spoken English. Cambridge: Blackwell; 2003.

Tucker, BV.; Warner, N. Inhibition of processing due to reduction of the American English flap. Proceedings of the 16th International Congress of Phonetic Sciences; Saarbruecken. 2007. p. 1949-1952.

Zue V, Laferriere M. Acoustic study of medial /t,d/ in American English. Journal of Acoustical Society of America. 1979; 66(4):1039–1050.

Table 1

Means and standard deviations (in parentheses) of percentages of allophonic variant productions as a function of phonological context for Experiment 1.

| | | Phonological context | | | |
		Favors [.]	Favors [t]	Favors [ɾ]	Favors [ʔ]
Surface realization (%)	[.]	**46 (32)**	12 (11)	5 (6)	4 (3)
	[t]	54 (32)	**86 (11)**	6 (6)	6 (8)
	[ɾ]	0 (1)	0 (1)	**88 (7)**	1 (1)
	[ʔ]	0 (1)	0 (1)	1 (1)	**90 (9)**

Table 2

Means and standard deviations (in parentheses) of percentages of allophonic variant usage as a function of phonological context for the subset of lexical items with $n \geq 14$ from the Buckeye Corpus; these items are marked with asterisks in the Appendix.

		Phonological context			
		Favors [ɾ]	Favors [t]	Favors [ɹ]	Favors [ʔ]
Surface realization (%)	[ɾ]	**93 (2)**	3 (4)	9 (11)	16 (NA)
	[t]	4 (4)	**96 (3)**	8 (9)	18 (NA)
	[ɹ]	3 (3)	1 (1)	**81 (17)**	0 (NA)
	[ʔ]	0 (0)	0 (0)	2 (7)	**65 (NA)**

Table 3

Closure and VOT duration (in ms) for surface realizations of word-medial [t] and [ɾ] in Experiment 2 stimuli.

| | | Closure Duration (ms) | | | VOT (ms) | | | |
| | | Phonological context | | | Phonological context | | | |
Surface realization		Favors [.]	Favors [t]	Favors [ɾ]	Favors [2]	Favors [.]	Favors [t]	Favors [ɾ]	Favors [2]
	[t]	47 (13)	51 (12)	54 (10)	56 (13)	65 (11)	43 (20)	57 (16)	56 (12)
	[ɾ]	25 (8)	29 (11)	28 (5)	32 (9)	13 (5)	13 (7)	11 (3)	13 (5)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 4

Duration of intermittent waveform irregularity and silence for surface realizations of word-medial [ɾ] in Experiment 2 stimuli.

		Irregularity + Silence (ms)		
		Phonological context		
		Favors [.]	Favors [t]	Favors [ɾ]	Favors [ʔ]
Surface realization	[ɾ]	95 (21)	93 (20)	99 (17)	106 (16)

Table 5

Mean percentages of "word" classifications for the four surface realizations in each phonological context. Letter subscripts indicate which conditions were statistically different from one another in each phonological context.

		Phonological context			
		Favors [ɾ]	Favors [t]	Favors [ɹ]	Favors [ʔ]
Surface realization (%)	[ɾ]	**77**_a	18_a	19_a	20_a
	[t]	97_b	**98**_b	97_c	92_b
	[ɹ]	45_c	86_c	**94**_c	21_a
	[ʔ]	64_a	18_a	47_b	**94**_b

Table 6

Mean reaction time to "word" classifications in three phonological contexts as a function of whether the surface realization of /t/ was [t] or the realization favored in that context. Standard deviations are in parentheses.

		Phonological context		
		Favors [.]	**Favors [ɾ]**	**Favors [ʔ]**
Surface realizati on (%)	Favored realization	942 (221)	796 (200)	902 (199)
	[t]	885 (179)	824 (154)	848 (204)

Appendix

Phonological Context: Favors [t]

*She visited a **Baptist** church on Sunday. (mother)*

*He had a **blister** on his finger. (hammer)*

*He had a **captive** audience during the show. (magic)*

*She finished **faster** than she planned. (school)*

*He will turn **fifty*** years old this year. (March)*

*She heard that **laughter** is sometimes best. (medicine)*

*Jeremy **lifted** the weights in gym. (iron)*

*She was a **master** at convincing them. (truth)*

*We used to watch **Mister** Rogers on television. (school)*

*He pulled a **pistol** out of his belt. (robbery)*

*The walls were made of **plaster** and clay. (house)*

*She had a **poster** in her room. (cat)*

*He was scheduled on the **roster** to play. (game)*

*Be sure to make **safety** come first. (driving)*

*The college **semester** seems longer. (December)*

*She gave her **sister*** a present. (Christmas)*

*You need a **system*** for keeping track. (homework)*

*The **Western** world values hard work. (employees)*

*She visited the **doctor*** when she was sick. (cold)*

*Change the **filter** before you call someone. (repairs)*

*They learned a lot in **history** class. (world)*

*She asked the **minister** for advice. (problems)*

Phonological Context: Favors [ɾ]

*Being early is **better*** than being late. (class)*

*Have some **butter** with your bread. (fresh)*

*She toured the **city*** in a bus. (Thursday)*

*He asked his **daughter*** to pick up her room. (clothes)*

*This car gets **forty*** miles to the gallon. (highway)*

*He was **getting*** cold by the window. (draft)*

*He came **later*** than expected. (gathering)*

*He needed to write a **letter** to his friend. (short)*

*The house is a **little*** further up the road. (tree)*

*It doesn't **matter*** whether you go or not. (conference)*

*There was a **meeting** on Wednesday. (long)*

*We went to a **party*** at our friend's house. (Saturday)*

*It was a **pity** you lost the game. (football)*

*You can drive **pretty*** far on a tank of gas. (car)*

*She gave him **thirty*** days to leave. (office)*

*His pet **turtle** lives in a box. (cardboard)*

*Pour the **water*** into the glass. (tap)*

Phonological Context: Favors [ʔ]

*He was **beaten** in a game. (chess)*

*He had been **bitten** earlier in the week. (spider)*

*He told the principal a **blatant** lie. (accident)*

*A couple of **buttons** were missing. (shirt)*

*You can **certainly*** get to the school. (time)*

*He prefers **cotton** shirts and pants. (rayon)*

*Sit by the **fountain** and cool off. (shade)*

*She had **gotten** a raise. (work)*

*The brown **kitten** was very small. (helpless)*

*He took **Latin** in college. (years)*

*Put on your **mittens** before going out. (cold)*

*They climbed the **mountain** in February. (snow)*

*The apple was **rotten** through. (brown)*

*Write a **sentence** from memory. (word)*

*He decided to **shorten** the stay. (family)*

*You should **straighten** up your room. (visit)*

*We didn't **sweeten** the dough enough. (cookie)*

*He tried to **whiten** his teeth. (bleach)*

*She has an **apartment** near the cleaners. (dry)*

*There can be **lightning** during storms. (thunder)*

*She needed a **partner** for the contest. (dance)*

*He was a **witness** in the trial. (criminal)*

Phonological Context: Favors [.]

*There was a **bounty** on his head. (million)*

*She likes to be the **center*** of attention. (family)*

*The kitchen **counter** was a mess. (dinner)*

*The catcher patrolled the **county** regularly. (dog)*

*He saw the **dentist** for a checkup. (monthly)*

*We had to **enter** the sweepstakes. (prize)*

*He writes **fantasy** and science fiction. (books)*

*He was very **gentle** with the baby. (newborn)*

*He had no **incentive** to find a job. (steady)*

*They asked him to **interview** on Friday. (position)*

*She had trouble **mounting** the frame. (wall)*

*He paid **ninety*** dollars for the show. (cash)*

*We had **plenty** of food left over. (dinner)*

*The sign **pointed** down the road. (town)*

*She got a **rental** car for the trip. (business)*

*I have about **twenty** dollars. (pocket)*

*She thought he **wanted*** to go shopping. (clothes)*

Phonological Context: Favors [t]	Phonological Context: Favors [ʔ]
*He became a **writer*** after college. (sports)*	*It's cloudy in the **winter** months. (outside)*
*The students **hated** going to school. (morning)*	*She was **contented** in the relationship. (boyfriend)*
*This is **native*** to our country. (fruit)*	*She had an **encounter** on the train. (police)*
*The library sent a **notice*** about our books. (overdue)*	*There was damage to the **frontal** lobe. (brain)*
*You should have **voted*** in the last election. (president)*	*There was no **parental** guidance. (concert)*