Processing variant forms in spoken word recognition: The role of variant frequency

Cynthia M. Connine, Larissa J. Ranbom, and David J. Patterson
State University of New York, Binghamton, New York

Recognition of a spoken word phonological variant—schwa vowel deletion (e.g., *corporate → corp'rate*)—was investigated in vowel detection (absent/present) and syllable number judgment (two or three syllables) tasks. Variant frequency corpus analyses (Patterson, LoCasto, & Connine, 2003) were used to select words with either high or low schwa vowel deletion rates. Speech continua were created for each word in which schwa vowel length was manipulated (unambiguous schwa-present and schwa-absent endpoints, along with intermediate ambiguous tokens). Matched control nonwords were created with identical schwa vowel continua and surrounding segments. The low-deletion-rate words showed more three-syllable judgments than did the high-deletion-rate words. Matched control nonwords did not differ as a function of deletion rate. Experiments 2 and 3 showed a lexical decision reaction time advantage for more frequent surface forms, as compared with infrequent ones, for schwa-deleted (Experiment 2) and schwa-present (Experiment 3) stimuli. The results are discussed in terms of representations of variant forms of words based on variant frequency.

Researchers in the field of spoken language processing have long grappled with the issue of how variability in the speech signal is accommodated by the listener. The picture that has emerged from a wide variety of research domains is that listeners appear to encode the variability that is experienced in the speech they hear. In the processing of speech sounds, Miller (2001) has shown that listeners encode variability within a phonetic category, and Norris, McQueen, and Cutler (2003) have demonstrated that listeners can tune existing categories to accommodate new dialects. Similar issues have been investigated for spoken word recognition, and the focus has been on how listeners accommodate variability among talkers (Nygaard & Pisoni, 1995). One general finding has been that individual-speaker information appears to be represented in memory and accessed during spoken word recognition (Goldinger, 1998; Mullennix, 1997). More recently, Kisilevsky et al. (2003) have shown that information specific to a mother's voice is encoded in utero. Other research has demonstrated that emotional tone of voice (Nygaard & Lunders, 2002), intonation (Church & Schacter, 1994), and vocal effort (Nygaard, Burt, & Queen, 2000) are also represented when spoken language is processed. The present research posed a similar question for the processing of phonological variants (alternative pronunciations of spoken words). Specifically, we investigated the question of whether listeners represent phonological variants in lexical memory on the basis of form frequency occurrence (phonological variant frequency) and utilize that information when processing spoken words.

Phonological variant frequency refers to the experienced production frequency of a particular phonological variant, relative to an alternative. The notion of variant frequency differs from typically experienced lexical frequency metrics (simple frequency counts) that have been used to predict performance in a wide variety of word recognition tasks (Balota, 1994). The robust and ubiquitous nature of lexical frequency effects has shaped theoretical assumptions about the representation of lexical form (Lively, Pisoni, & Goldinger, 1994). Occurrence frequency has also been implicated in meaning activation of ambiguous words (Binder & Rayner, 1998). In the research domains of word recognition and retrieval of lexical meaning, the empirical results show that higher occurrence frequency (operationalized by large- or small-scale analyses of language use) is correlated with faster and/or less errorful processing.

The central role for frequency in predicting lexical processing emphasizes the notion that experience is an important organizing dimension of lexical knowledge. The present experiments extended this insight to the experienced occurrence of phonological variation. Specifically, we investigated the role of variant form frequency in the processing of a phonological variant called *schwa vowel deletion*. Linguistic accounts specify two environments, pre- and poststress, in which schwa vowel deletion can occur (Zwicky, 1972; see also Oshika, Zue, Weeks, Neu, & Aurbach, 1975). In the prestress environment, a schwa vowel may be deleted in a word-initial syllable if it is preceded by at least one consonant and followed by a stressed syllable (e.g., *suppose → */spouz/). In the poststress environment, three-syllable words may have a schwa vowel in the second syllable deleted if it is preceded by a syllable

---

C. M. Connine, connine@binghamton.edu

containing a stressed vowel (*corporate* → /korprət/). In the present experiments, we focused on the latter environment. Our approach was to consult variant frequency statistics from a conversational database (Patterson, LoCasto, & Connine, 2003) and use the statistics to generate behavioral predictions for perceptual judgments.

Corpus statistics have been successfully utilized for making predictions concerning how listeners process another variant, American English flapping. The flap variant can occur when a voiceless alveolar stop follows a stressed vowel (e.g., the word *pretty* may be realized as [prɪɾɪ]). In their corpus analysis, Patterson and Connine (2001) extracted productions of potentially flapped words from a conversational database of American English (Godfrey, Holliman, & McDaniel, 1992) and found that 96% of the tokens ($N = 2,172$) consisted of unambiguous, flapped productions, with the remaining 4% distributed across [t] and glottal stop productions. Connine (2004) used the flap variant frequency statistics to investigate the hypothesis that listeners recognize flap productions via their surface form. In a phoneme identification task, listeners were asked to identify the initial segment ("b" or "p") of word–nonword continua (e.g., *pretty–bretty*) that was embedded in either a flap or a [t] variant carrier word (e.g., *preDy–breDy* or *preTTy–breTTy*). The results showed a variant frequency effect: More identification responses forming a real word were found when the to-be-identified speech sound occurred in the more frequently experienced flap carrier. Connine argued that these results suggest that listeners recognize a flap via a preexisting representation in lexical memory.

One potential limitation of Connine's (2004) results is that the frequency structure of flaps has a single, dominant production; there were no words from the Patterson and Connine (2001) corpus with frequency statistics showing more [t] than flap productions. The frequency distribution of the flap variant precludes a test of a variant frequency effect favoring a [t] carrier—that is, a reversal of the variant effect for a set of words with a predominant [t] production. This raises the possibility that the variant frequency effect for flaps arises from a generalized preference for flapped pronunciations that is not based on information about individual words. A phonological variant with a range of frequency rates among words would permit an experimental design with predictions based on the frequency statistics of different groups of words. A variant frequency effect for groups of words within the same variant class but with opposing frequency structures would provide important supporting evidence for variant-frequency-based lexical representations. A second potential limitation of Connine's results is that the variant in question concerned a change in a highly frequent consonant (/t/), and it is important to determine whether the variant frequency effects generalize to other segments. Finally, schwa vowel deletion involves a number of changes in the structure of the resulting production. It is important to determine whether deletion variants operate in the same way as variants that involve changed segments. Schwa vowel deletion variants extend previous results in a way that addresses these issues.

As was previously mentioned, the distribution of schwa vowel deletion has a less dominant structure. Patterson et al. (2003) obtained occurrence frequencies for schwa vowel deletion from conversations between strangers over the telephone (Godfrey et al., 1992) and between friends in a face-to-face story-retelling task (see Patterson et al., 2003, for further details). One finding was that schwa vowels in prestress environments (e.g., *suppose* → *s'pose*) were rarely deleted (16%, or 718 of 4,333 tokens). A second finding from the Patterson et al. corpus was that a poststress environment (e.g., *corporate* → *corp'rate*) showed relatively high deletion rates (55%; 1,228 of 2,214 tokens). Corpus analyses also revealed two distinct groups of words within the poststress environment that were distinguished by deletion rates. One word group (high deletion rate) occurred frequently as the deleted form (more than 50%), and a second word group (low deletion rate) occurred frequently as the undeleted form (less than 50%).

In the present research, we utilized the corpus statistics for poststress environment words to select high- and low-deletion-rate words. The two word groups were used to investigate the hypothesis that experience with a particular variant form (high- vs. low-deletion-rate words) shapes the way that word is lexically represented. In Experiment 1, we used a version of the paradigm developed by Ganong (1980), in which speech continua are presented for an identification response. Speech continua were created in which schwa vowel duration of the medial syllable was manipulated to produce a perceptual change from two to three syllables. The resulting continua consisted of a clear three-syllable token (the longest schwa vowel duration) to a clear two-syllable token (the shortest vowel duration). Intermediate duration schwa vowel tokens were perceptually ambiguous with respect to the presence/absence of the medial vowel. Experiment 1 required a judgment of the number of syllables (two or three).

We hypothesized that a phonological variant frequency effect would emerge: Words frequently encountered as a three-syllable variant (on the basis of corpus statistics) would be more likely to be judged as consisting of three syllables than would words frequently encountered as a two-syllable variant. This predicts more three-syllable judgments for the low-deletion-rate words than for the high-deletion-rate words. In making these predictions, it should be noted that typical lexical influences in this paradigm (Connine & Clifton, 1987; Ganong, 1980; Pitt & Samuel, 1993) are confined to ambiguous speech tokens, and we anticipated a similar pattern for deletion rate influences. Unambiguous endpoints should be identified on the basis of the acoustic–phonetic information, irrespective of carrier type.

A second aspect of the design ensured that differences in the processing of high- and low-deletion words were variant frequency based and not due to uncontrolled differences in the schwa vowel and its environs. This is an important consideration because perception of schwa vowels could be influenced by a number of aspects of the stimuli. Van Donselaar, Kuijpers, and Cutler (1999) have demonstrated that one apparent consequence of schwa vowels is to increase the perceptibility of surrounding

segments. In addition, Pitt (1998) has shown that phonotactic constraints (legal vs. illegal sequences of segments) influence the perceptibility of schwa vowels in a syllable-counting task. A final consideration is that the segments surrounding the schwa can influence the amplitude, duration, and spectral composition of the schwa (see Stevens, 1998), and the gestural overlap of coarticulated segments may obscure acoustic markers of a schwa vowel (Manuel, 1992). In order to address these concerns in Experiment 1, a nonword carrier was constructed for each word stimulus. Recordings of the stimuli were cross-spliced so that the same medial syllable (schwa vowel plus preceding and following consonants) appeared in a word and nonword pair. Thus, information in the schwa-bearing syllable was physically identical for a word and its nonword counterpart. Any difference due to variant frequency in the low- and high-deletion word carriers should be minimized for the low- and high-deletion nonword counterparts. This predicts that deletion rate should not influence reports of the medial vowel in nonwords: high- and low-deletion nonwords should not differ in syllable number judgments.

Inclusion of word and nonword carriers will also permit an examination of a second aspect of the variant frequency effect (more judgments corresponding to the structure of the most frequently encountered variant form) as a function of deletion rate. If the vowel-bearing variant is represented lexically for low-deletion-rate stimuli, the lexical effect (more three-syllable judgments for words than for nonwords) should be larger in low- than in high-deletion-rate stimuli.

EXPERIMENT 1

Method

Participants. A total of 31 native speakers of American English with normal hearing participated in the experiment and received credit toward a psychology class.

Materials and Design. A group of words, all three syllables long with an optional poststress, medial schwa vowel, was selected on the basis of a corpus analysis of variant production frequency for vowel deletion (Patterson et al., 2003). Low-deletion words ($n = 8$) had deletion rates less than 50% ($\bar{X} = 20.5\%$; range, 3.7%–32.4%), and high-deletion words ($n = 11$) had deletion rates greater than 50% ($\bar{X} = 85.4\%$; range, 69.1%–100%). The average lexical frequency for low-deletion words was 98 per million (range, 1–405), and for high-deletion words, it was 70 per million (range, 15–378; Francis & Kučera, 1982).[1] A nonword was created for each word by altering the segments in the first and last syllables and leaving the medial vowel and its surrounding segments identical to the word counterpart (see Appendix A). A final low-deletion word (*cardinal/pardinim*) was selected for a practice block.

The stimuli were recorded by a female native English speaker and were digitized directly onto a computer at 44 kHz (16-bit resolution). Two schwa-vowel-bearing productions of each word and one production of each nonword were selected. The schwa vowel and the speech segments directly prior to and following the schwa were excised from a recording of each word (e.g., the "mer" from *camera*). A series of tokens was generated by shortening the vowel one pitch period at a time until no acoustic or perceptual evidence of a schwa vowel remained. Each schwa vowel token, plus the surrounding segments, was used to replace the medial syllable in the second recording of the word and its nonword counterpart. The resulting continua ranged from 9 to 11 stimulus tokens. Measurements of stimulus duration showed longer durations for vowel-bearing nonwords than

for words (680 vs. 597 msec). Since stimulus duration may influence judgments in an unpredictable fashion, a second set of stimuli (duration-equivalent stimuli) was constructed. In the second set of stimuli, new nonwords were constructed and recorded, and tokens were selected with equivalent durations (word and nonword durations differed by less than 5 msec). A procedure identical to that previously described was followed in order to create schwa vowel continua for the duration-equated stimulus set. The resulting continua consisted of 9–11 stimulus tokens. A pilot study was run in order to select a manageable number of tokens for the experiment. In the pilot study, the original and duration-adjusted stimuli were presented for vowel present/absent judgments. One group of listeners made judgments on the original stimuli, and a second group of listeners made judgments on the duration-adjusted stimuli ($n = 5$ in each participant group). The pilot data were used to select 3 tokens from the region of each continuum perceived to be in the midrange region (referred to as Tokens 2, 3, and 4). Perception of the midrange tokens for listeners in the pilot study ranged from 80% (Token 2) to 35% (Token 4) vowel present. Tokens 1 and 5 contained the longest and shortest vowel lengths, respectively (see Appendix A).

In the experiment, there were two groups of participants; one group heard the original stimuli ($n = 15$), and a second group heard the duration-equated stimuli ($n = 16$). Each yoked word and nonword continuum was organized into a block (e.g., the 5 stimuli making up the schwa-present–schwa-absent continuum for the word *general* and the nonword *shenerom* defined a block). Within each block, the 5 stimuli from a given continuum were presented five times; thus, a block consisted of 50 stimuli. Presentation of the 50 stimuli within a block was randomized. Across the experimental session, presentation of blocks was randomized. A different random order of tokens within a block and a different random order of blocks across the experimental session were used for each subject. Trials were separated by 2 sec of silence.

Procedure. The stimuli were presented binaurally over headphones to each participant in a sound-dampened room. Instructions to the participants introduced the possibility of two syllable-variant productions of three-syllable words and nonwords. The participants were instructed to indicate whether the stimulus they heard was three syllables or two syllables by marking a response sheet with a "3" or a "2," respectively. The experimental session was preceded by a practice block using a low-deletion-rate stimulus (*cardinal/gardinim*) not included in the experiment.

Results and Discussion

Three-syllable judgments for each stimulus token (collapsed across original and duration-matched stimulus groups) as a function of carrier and deletion rate are shown in Figure 1. The listeners used the vowel length of the stimulus to guide their syllable number judgments, since there was a very orderly decrease in the percentage of three-syllable responses as vowel duration decreased. Judgments for endpoint stimuli were nearly identical for words and nonwords. Intermediate tokens showed more three-syllable responses for the low-deletion word carriers than for the high-deletion word carriers. Nonword carriers did not differ, and the lexical effect (word vs. nonword carrier schwa detections) was larger for low- than for high-deletion-rate stimuli.

A four-way ANOVA was conducted on the percentage of three-syllable judgments with stimulus group (original/length-matched), deletion rate (high/low), carrier lexical status (word/nonword), and stimulus token (three intermediate tokens) as factors. All reported effects were significant at $p < .05$ unless otherwise indicated. The main effect of stimulus group was not significant ($F < 1$). The
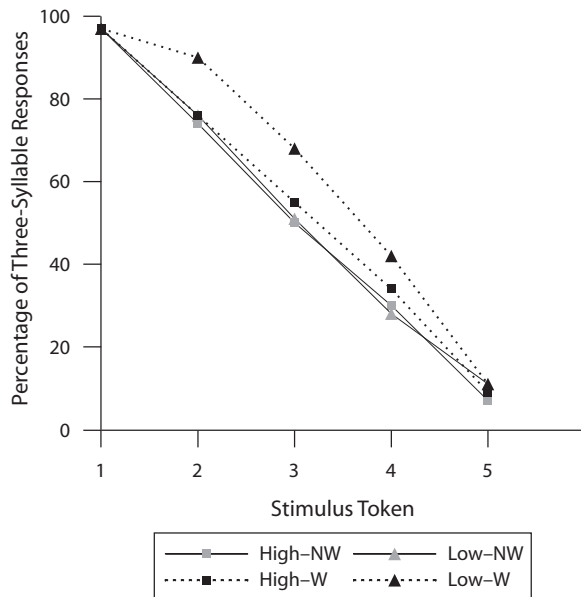
Figure 1. Percentage of three-syllable judgments for each stimulus token as a function of carrier (word [W] and nonword [NW]) and variant deletion rate (high and low) in Experiment 1.

interaction of stimulus group, lexical status, and deletion rate was also not significant ($F < 1$). Stimulus group interacted with lexical status and stimulus token [$F_1(2,58) = 7.9$, $MS_e = 49$; $F_2(2,34) = 8.7$, $MS_e = 29$].

Importantly, the interaction of lexical status and deletion rate was significant [$F_1(1,29) = 15$, $MS_e = 374$; $F_2(1,17) = 28$, $MS_e = 169$]. The interaction can be seen in three aspects of the data: The high-deletion words showed fewer three-syllable judgments than did the low-deletion words (55% vs. 66%); the high- and low-deletion nonwords did not differ (51.5% vs. 51.6%); and the lexical effect was smaller for high- (4%) than for low- (11%) deletion stimuli. Thus, words typically heard as three syllables were more likely to be reported as such, as compared with words typically heard as two syllables. The similar data for the high- and low-deletion nonwords suggest that the influence of variant frequency was not due to some idiosyncratic properties of the region surrounding the deletion site.

Other significant effects included main effects of lexical status [more three-syllable judgments for word, relative to nonword, carriers; 61% vs. 51%, respectively; $F_1(1,29) = 22$, $MS_e = 374$; $F_2(1,18) = 30$, $MS_e = 207$] and stimulus token [the number of three-syllable judgments increased as schwa vowel length increased; $F_1(2,58) = 537$, $MS_e = 118$; $F_2(2,34) = 71$, $MS_e = 531$]. As was previously mentioned, the identification functions were exceedingly orderly and provide strong evidence that the task is sensitive to changes in the acoustic properties of the stimuli (see McQueen, 1997).

Two additional effects were statistically significant. The main effect of deletion rate (more three-syllable judgments in the low- than in the high-deletion-rate stimuli; 59% vs. 53%) was significant in the participant analysis

[$F_1(1,29) = 12.03$, $MS_e = 275$], as was the interaction of deletion rate and stimulus token [$F_1(2,58) = 4.6$, $MS_e = 51$]; three-syllable judgments for the low- versus high-deletion-rate stimuli varied in size across token (Token 2, 83% vs. 73%; Token 3, 59% vs. 52%; Token 4, 35% vs. 32%).

The data pattern is consistent with predictions based on corpus statistics; judgments of syllable number for ambiguous tokens were a function of the frequency with which a given variant was experienced. Two additional experiments were conducted to provide converging evidence, using a lexical decision reaction time task for high- and low-frequency schwa-deleted (Experiment 2) and schwa-present (Experiment 3) variants. Another goal of Experiments 2 and 3 was to investigate a potential contribution from the segments surrounding the deletion site—specifically, whether the nature of the surrounding cluster contributes to the processing of the variants (see Pitt, 1998). An effect of segmental environment for stimulus sets equated for variant frequency may suggest a prelexical locus for schwa perception (Gow, 2003; see also the General Discussion section). Selection of stimuli included sets of high-deletion-rate stimuli in which the segments formed a cluster that might function as a legal onset (e.g., the /kl/ sequence formed from *chocolate → choclate*) or an illegal cluster (e.g., the /mr/ sequence formed from *camera → cam'ra*). The extent to which processing is influenced by the segmental environment of the schwa will be revealed by differences in the high-deletion-rate stimulus sets. All of the low-deletion-rate stimuli were words in which the deletion site resulted in an illegal cluster.[2] A comparison of the illegal cluster high- and low-deletion-rate stimulus groups would provide a replication of the variant frequency effect found in Experiment 1, using a new task.

EXPERIMENT 2

Method

Participants. A total of 39 native speakers of American English with normal hearing participated in the experiment and received credit toward a psychology class.

Materials. Two sets of eight high-deletion words were selected; in one set (high-deletion legal cluster), deletion of the schwa resulted in a legal cluster, and in the second set (high-deletion illegal), schwa deletion resulted in an illegal cluster. A third set of eight low-deletion words was selected in which schwa deletion resulted in an illegal cluster (see Appendix B). An additional set of 24 filler pseudowords was created by altering one or more segments of optional schwa-deleted three-syllable words not used in the experiment (average duration = 585 msec). Schwa-deleted and schwa-bearing tokens of each stimulus were recorded by a female native English speaker and were digitized directly onto a computer at 44 kHz (16-bit resolution). The schwa-deleted nature of the token was verified by visual inspection of the waveform and by perceptual judgments of two phonetically trained listeners.

In order to confirm the alleged schwa status of the schwa-bearing and schwa-deleted tokens, 24 additional participants (none of whom participated in Experiment 1, 2, or 3) were presented the 48 stimulus tokens (24 two-syllable tokens from Experiment 2 and 24 three-syllable tokens from Experiment 3) in a syllable-counting task (two vs. three syllables). The participants were instructed that they would hear pronunciations of words that differed in the presence/absence of the medial schwa. Instructions included an example of the two-

Table 1
Reaction Time (in Milliseconds) and Accuracy (%) for Schwa-Deleted Variants As a Function of Stimulus Type in Experiment 2

Stimulus Type	Reaction Time	Accuracy
High Deletion Rate		
Legal	190	98
Illegal	205	95
Low Deletion Rate		
Illegal	307	82

versus three-syllable pronunciations and emphasized accuracy. Each token was presented five times (randomized across the experiment), and feedback on each token was provided after each response. Overall, accuracy rates were quite high and showed clear evidence for the perceptibility of the intended pronunciation for both the three-syllable (96%, 97%, and 97% for high legal, high illegal, and low illegal, respectively) and the two-syllable (96%, 94%, and 98% for high legal, high illegal, and low illegal, respectively) tokens.

Procedure. The set of 24 word and 24 nonword stimuli were presented binaurally over headphones to groups of 3 or fewer participants in a sound-dampened room. An experimental session lasted approximately 5 min. A different randomized presentation was used for each participant group. The participants were instructed to indicate whether the stimulus they heard was a real word or a pseudoword by pressing an appropriately marked response box.

Results and Discussion

Table 1 displays reaction time (measured from stimulus offset) and accuracy as a function of stimulus type (high-deletion-rate legal cluster, high-deletion-rate illegal cluster, and low-deletion-rate illegal cluster). Reaction times longer than 1,200 msec were removed from the analysis (fewer than 2% of the data). A planned comparison for the high-deletion-rate versus low-deletion-rate illegal clusters was significant [$t_1(38) = 7.4, p < .01; t_2(14) = -2.7, p < .05$]. The high-deletion-rate stimuli (legal and illegal clusters) did not differ [$t_1(38) = -1.3, p = .17; t_2 = -0.55$]. A set of planned comparisons for accuracy rates was significant for the high- versus low-deletion-rate illegal stimuli [$t_1(38) = -4.3, p < .01; t_2(14) = 2.57, p < .05$] and was significant across participants, but not across items, for the high-legal and -illegal stimuli [$t_1(38) = 2.8, p < .01; t_2(14) = 1.38, p = .2$]. Nonword stimuli were responded to with an accuracy of 88% and a reaction time of 365 msec.

A reaction time advantage for more frequently heard variants was found: High-deletion-rate stimuli were responded to more quickly than were low-deletion-rate stimuli for schwa-deleted tokens. In addition, the legality of the cluster surrounding the deletion site did not influence the speed of responding for words equated on deletion rate. Experiment 3 paralleled the design of Experiment 2, except that schwa-bearing tokens were used. If deletion rate influences word recognition, reaction time should be shorter for the low-deletion stimuli and equivalent for stimulus groups matched on variant frequency.

EXPERIMENT 3

Method

Participants. A total of 33 native speakers of American English with normal hearing participated in the experiment and received credit toward a psychology class.

Materials. A schwa-bearing token of each stimulus was recorded by a female native English speaker and was digitized directly onto a computer at 44 kHz (16-bit resolution; see Experiment 2). As in Experiment 2, the status of the schwa was assessed by visual inspection of the waveform and by the judgment of two phonetically trained listeners. The nonword fillers used in Experiment 2 were recorded as schwa-bearing tokens and were used as fillers (average duration = 698 msec).

Procedure. The procedure was identical to that in Experiment 2.

Results and Discussion

Table 2 displays reaction time (measured from stimulus offset) and accuracy as a function of schwa-bearing stimulus type (high-deletion-rate legal cluster, high-deletion-rate illegal cluster, and low-deletion-rate illegal cluster). A planned comparison for the high-deletion-rate versus low-deletion-rate illegal clusters was significant by participants, but not by items [$t_1(32) = -3.2, p < .01; t_2 = 0.96$]. The high-deletion-rate stimuli (legal and illegal clusters) did not differ ($t_1 = 0.04; t_2 = 1.1$).

The critical comparison, high-deletion- versus low-deletion-rate illegal stimuli, was not significant across items. The small number of items available with the appropriate properties likely limited the power of the items analysis. However, in light of the importance of this comparison and to more fully explore the relationship between reaction time and deletion rate, a random regression approach was used as a complementary analysis (Lorch & Myers, 1990). In this analysis, a separate regression model was fit for each participant's individual data, with deletion rate (from corpus statistics) and reaction time (for individual-item responses) as variables. A t test for single means (tested against zero) on the beta values ($\bar{X} = .13$) was significant [$t(32) = 3.7, SD = 0.19, p < .01$]. Comparable analyses performed separately for each group of stimuli were significant [high legal, $t(32) = -1.94, SD = 0.37, p = .06$; high illegal, $t(32) = 6.4, SD = 0.32, p < .01$; low legal, $t(32) = -3.1, SD = 0.38, p < .01$]. The regression analyses clearly do not replace an items analysis but suggest that there is a reasonable relationship between the corpus estimate of item deletion rate and word processing (reaction time).

A set of planned comparisons for accuracy rates was significant across participants (but not for items) for the high- versus low-deletion illegal stimuli [$t_1(32) = 2.88, p < .01; t_2(14) = 1.1, p = .27$] and for the high-legal versus illegal stimuli [$t_1(32) = -3.7, p < .01; t_2(14) = 1.1, p = .27$]. Nonword stimuli were responded to with an accuracy of 85% and a reaction time of 306 msec.

As did Experiment 2, Experiment 3 demonstrated that deletion frequency influenced processing time; lexical

Table 2
Reaction Time (in Milliseconds) and Accuracy (%) for Schwa-Present Variants As a Function of Stimulus Type in Experiment 3

Stimulus Type	Reaction Time	Accuracy
High Deletion Rate		
Legal	217	97
Illegal	216	93
Low Deletion Rate		
Illegal	187	97

decision reaction times for schwa-present surface forms were shorter for the low-deletion-rate than for the high-deletion-rate words. Comparing reaction times across Experiments 2 and 3 suggests a similar conclusion; for a given word set, the higher frequency surface form was responded to more quickly than was the low-frequency surface form. A post hoc cross-experiment ANOVA with deletion rate (low vs. high, collapsed across the legal and illegal word sets) and experiment (2 vs. 3) as variables showed a significant interaction [$F_1(1,72) = 100$, $MS_e = 1,706$, $p < .01$].

GENERAL DISCUSSION

In three experiments, we examined whether frequency of a surface form (schwa deleted or schwa intact) influenced syllable judgments (Experiment 1) and lexical decisions (Experiments 2 and 3). Experiment 1 used a syllable-counting task and showed more three-syllable judgments for low-deletion than for high-deletion words. Control nonword carriers with the same physical schwa vowel information as their word counterparts did not differ as a function of deletion rate. Experiments 2 and 3 showed a lexical decision reaction time advantage for more frequent surface forms than for infrequent surface forms for schwa-deleted (Experiment 2) and schwa-present (Experiment 3) stimuli. The results support a strong influence of variant form frequency and suggest that form frequency is an important organizing principle for lexical representation.

The results of Experiment 1 showed a smaller lexical effect for the high-deletion words than for the low-deletion words and are consistent with a form frequency effect. The effect of deletion frequency was confined to the mid-continua tokens in which there was some acoustic evidence for a schwa vowel. This is consistent with previous findings, using the phoneme identification task, in Ganong (1980) and others (see, e.g., Connine & Clifton, 1987; McQueen, 1997; Pitt & Samuel, 1993) showing that lexical influences are confined to the category boundary region. These findings suggest that lexical representations do not overturn perceptual evidence, rather, listeners integrate the available speech information with lexical knowledge. However, why does lexical context provide some benefit (relative to the control) for schwa perception in stimuli that are rarely heard as schwa-present forms? One possibility is that lexical representations encode more than a single form and the strength of the alternative form is determined by experienced frequency. For schwa-deleted variants, a high-deletion word would have a dominant schwa-deleted form, along with a weak schwa-present form. Processing of a schwa-deleted form will be most strongly influenced by the dominant form, but lexical activation dynamics (based on similarity activation) would demand that the alternative lexical form is weakly activated. This account attributes the smaller lexical effect for low-deletion-rate stimuli to a weakly represented schwa-present representation that coexists in lexical memory with a more dominant schwa-deleted form.

Research on processing phonological variation has raised a number of issues related to the interpretation of our results. The first issue concerns the role of prelexical processes in the recognizing of variants. Gow and Im (2004; see also Mitterer & Blomert, 2003) have argued that the processing of place assimilation variants relies on general perceptual mechanisms devoted to grouping of phonetic features. In support of this view, a number of studies have demonstrated that language experience is not a limiting factor in the processing of assimilation variants (Mitterer, Csepe, Honbolygo, & Blomert, 2006). The research presented here did not specifically address the ability of non-English speakers' perception of the variants, but our interpretation of the results implicates a lexical contribution to the observed effects. Although we do not deny a potential for prelexical influences in schwa vowel perception, the careful word/nonword comparison in Experiment 1 suggests that any such prelexical influences are weak, relative to a listener's lexical knowledge of schwa variant frequency.

A second issue concerns the extent to which lexical representations operate to *restore* a schwa, given little (the midrange tokens in Experiment 1) or clear (the endpoint stimuli in Experiment 1 and the unambiguous tokens in Experiments 2 and 3) evidence for or against a schwa. Some researchers have argued that the processing of highly reduced speech missing any acoustic–phonetic evidence for missing segments can be characterized as a restoration process (see Ernestus, Baayen, & Schreuder, 2002; Mitterer & Ernestus, 2006). In support of this view, phoneme-monitoring responses indicate perception of a missing segment. Classic demonstrations of phoneme restoration have shown that there is perceptual equivalence for complete (noise-added) and incomplete (noise-replaced) speech tokens (Samuel, 1981). In contrast to these findings, listeners did not report schwa vowels in the absence of speech evidence (Experiment 1) and did not process schwa-present and schwa-absent tokens equivalently (Experiments 2 and 3). The present results are consistent with other research suggesting that an important limiting factor in the perception of missing information is that it must co-occur with some acoustic–phonetic evidence for the missing event. For example, Samuel has shown that phonemic restoration depends on the similarity between the missing speech and noise overlay. Deelman and Connine (2001) have shown that lexical activation for words missing a consonant release depends on the presence of vocalic information to support the missing information. Although we cannot rule out the presence of hidden schwa gestures (Manuel, 1992), the schwa-deleted tokens in Experiment 1 (endpoints) and Experiment 2 do not appear to contain acoustic markers for a deleted schwa, and listeners appear to perceive these tokens as intended. This raises the question as to why other investigations have shown that listeners report hearing missing information in highly reduced stimuli in the absence of any acoustic–phonetic markers (Kemps, Ernestus, Schreuder, & Baayen, 2004). The apparent inconsistency may be reconciled by considering the way in which listeners approach a given task and, in particular, their reliance on lexical versus prelexical knowledge (Cutler, Mehler, Norris, & Segui, 1987). Perception of completely missing segments may arise

primarily from a lexical locus for the response (see also Ranbom & Connine, 2007b).

Schwa-bearing surface forms (Experiment 3) showed a muted variant frequency effect, relative to the schwa-absent counterparts, which suggests that the full forms were less influenced by form frequency (this can also be seen by comparing performance across Experiments 2 and 3). The reasons for this are not clear, but some work by van Donselaar et al. (1999) has provided some intriguing possibilities. Van Donselaar et al. investigated optional schwa insertion (epenthesis) in the processing of Dutch words and found that schwa-bearing tokens facilitated responding. They suggested that the presence of a schwa vowel is beneficial because its presence provides a perceptual advantage for the segments surrounding the schwa. An overall perceptual advantage for the schwa-bearing tokens may be in force and would be consistent with results in the more general literature on reduction in conversational speech (e.g., Pollack & Pickett, 1963, and, more recently, Ernestus et al., 2002). Processing of a full-form production of any variant may benefit from a richer acoustic–phonetic signal and may mute the influences of other factors (LoCasto & Connine, 2002; Sumner & Samuel, 2005), and/or there may be a benefit because the full form is typically consistent with an orthographic form (see Ranbom & Connine, 2007b). However, our results demonstrate that a carefully controlled, naturally occurring reduction, schwa vowel deletion, must be considered in light of its occurrence frequency for a specific word.

The results presented here are consistent with the claim that the frequency of phonologically related productions is encoded and used during word recognition. In classic models of exemplar-based memory representations, multiple episodic representations have been hypothesized to account for prototype effects in memory (Hintzman, 1986) and the representation of indexical speaker information (Goldinger, 1998). If listeners represent individual instances of spoken words—episodic traces for each heard version of a word—presentation of a variant form will activate most strongly those representations consistent with it. The strength of the activation will be related to the number of activated representations; frequent occurrences of a given variant will result in more activation than will infrequent variants. This is an attractive account, in that it provides a single mechanism for the processing of indexical information (speaker and emotion) and variants of spoken words. Alternatively, the representation of linguistic information may be captured more naturally by abstract representations of phonological forms (Ranbom & Connine, 2007a). An important characteristic of this approach is that the representation of a variant in memory will be influenced by its frequency of occurrence. Variant occurrence frequency could be operationalized as differences in activation levels or as the relative amount of input required for activation. For both episodic and abstract representations, processing of phonological variation is not viewed as a special case for theories of spoken word recognition. Lexical representation of phonological variant frequency provides the listener with a lexical memory that is tuned to the experienced spoken language and permits fast and efficient processing.

Finally, the results highlight an important role for variant frequency statistics from conversational speech in the development of spoken word recognition theories. Corpus analyses of variant frequency move beyond simple notions of lexical experience embodied in word frequency counts and provide a window into more subtle aspects of language use (see also Dilley & Pitt, 2007; Pitt, Johnson, Hume, Kiesling, & Raymond, 2005). However, variant frequency counts are only one step in moving theories of spoken word recognition in the direction of ecological validity. It may be that speakers capitalize on the production variability of words in order to convey important dimensions of discourse processing. Previous research has shown that aspects of word production reflect discourse structure. For example, Fowler and Housum (1987) found that speakers reduce a word's intelligibility and duration when it occurs on a second occasion in a discourse. However, Fowler and Housum also found that the acoustic event that was potentially problematic from the point of view of spoken word recognition was utilized in forming a coherent representation of the discourse; reduced intelligibility and duration were used by listeners as an indicant of old versus new information. These results suggest that variability in first versus second mentions of a word provides important information concerning discourse structure. Alternative variant forms may serve a similar function in signaling discourse-relevant information. If so, this would suggest an important linkage between word recognition and discourse processing that may inform theories in both domains.

REFERENCES

BALOTA, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 303-358). San Diego: Academic Press.

BINDER, K. S., & RAYNER, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin & Review*, **5**, 271-276.

CHURCH, B. A., & SCHACTER, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 521-533.

CONNINE, C. M. (2004). 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 & Review*, **11**, 1084-1089.

CONNINE, C. M., & CLIFTON, C., JR. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 291-299.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, **19**, 141-177.

DEELMAN, T., & CONNINE, C. M. (2001). Missing information in spoken word recognition: Non-released stop consonants. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 656-663.

DILLEY, L., & PITT, M. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word

recognition. *Journal of the Acoustical Society of America*, **122**, 2340-2353.

ERNESTUS, M., BAAYEN, H., & SCHREUDER, R. (2002). The recognition of reduced word forms. *Brain & Language*, **81**, 162-173.

FOWLER, C., & HOUSUM, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory & Language*, **26**, 489-504.

FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

GANONG, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 110-125.

GODFREY, J. D., HOLLIMAN, E. C., & MCDANIEL, J. (1992). SWITCH-BOARD: Telephone speech corpus for research and development. In *1992 International Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, pp. 517-520). Los Alamitos, CA: IEEE Computer Society Press.

GOLDINGER, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, **105**, 251-279.

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

GOW, D. W., JR., & IM, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory & Language*, **51**, 279-296.

HINTZMAN, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.

KEMPS, R., ERNESTUS, M., SCHREUDER, R., & BAAYEN, H. R. (2004). Processing reduced word forms: The suffix restoration effect. *Brain & Language*, **90**, 117-127.

KISILEVSKY, B. S., HAINS, S. M. J., LEE, K., XIE, X., HUANG, H., YE, H. H., ET AL. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, **14**, 220-224.

LIVELY, S. E., PISONI, D. B., & GOLDINGER, S. D. (1994). Spoken word recognition: Research and theory. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265-302). San Diego: Academic Press.

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

LORCH, R. F., & MYERS, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 149-157.

MANUEL, S. (1992). Recovery of "deleted" schwa. In O. Engstrand & C. Kylander (Eds.), *Perilus XIV: Papers from the Symposium on Current Phonetic Research Paradigms for Speech Motor Control* (pp. 115-118). Stockholm: University of Stockholm.

MCQUEEN, J. M. (1997). Phonetic categorisation. In F. Grosjean & U. Frauenfelder (Eds.), *A guide to spoken word recognition paradigms* (pp. 655-664). Hove, U.K.: Psychology Press.

MILLER, J. L. (2001). Mapping from acoustic signal to phonetic category: Internal structure, context effects and speeded categorization. *Language & Cognitive Processes*, **16**, 683-690.

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

MITTERER, H., CSEPE, V., HONBOLYGO, F., & BLOMERT, I. (2006). The recognition of assimilated word forms does not depend on specific language experience. *Cognitive Science*, **30**, 451-479.

MITTERER, H., & ERNESTUS, M. (2006). Listeners report /t/s that speakers lenit: Evidence from /t/ lenition in Dutch. *Journal of Phonetics*, **34**, 73-103.

MULLENNIX, J. W. (1997). On the nature of perceptual adjustments to voice. In K. A. Johnson & J. W. Mullennix (Eds.), *Talker variability and speech processing* (pp. 67-83). San Diego: Academic Press.

NORRIS, D., MCQUEEN, J. M., & CUTLER, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, **47**, 204-238.

NYGAARD, L. C., BURT, S. A., & QUEEN, J. S. (2000). Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1228-1244.

NYGAARD, L. C., & LUNDERS, E. R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition*, **30**, 583-593.

NYGAARD, L. C., & PISONI, D. B. (1995). Speech perception: New directions in research and theory. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 63-90). San Diego: Academic Press.

OSHIKA, B., ZUE, V. W., WEEKS, R. V., NEU, H., & AURBACH, J. (1975). The role of phonological rules in speech understanding research. *IEEE Transactions on Acoustics, Speech, & Signal Processing*, **ASSP-23**, 104-112.

PATTERSON, D., & CONNINE, C. M. (2001). A corpus analysis of variant frequency in American English flap production. *Phonetica*, **58**, 254-275.

PATTERSON, D., LOCASTO, P. C., & CONNINE, C. M. (2003). A corpus analysis of schwa vowel deletion frequency in American English. *Phonetica*, **60**, 45-68.

PITT, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, **60**, 941-951.

PITT, M. A., JOHNSON, K., HUME, E., KIESLING, S., & RAYMOND, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, **45**, 89-95.

PITT, M. A., & SAMUEL, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 1-27.

POLLACK, I., & PICKETT, J. M. (1963). The intelligibility of excerpts from conversation. *Language & Speech*, **6**, 165-171.

RANBOM, L. R., & CONNINE, C. M. (2007a). Lexical representation of phonological variation. *Journal of Memory & Language*, **57**, 273-298.

RANBOM, L. R., & CONNINE, C. M. (2007b, November). *The role of orthography in spoken word recognition*. Paper presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA.

SAMUEL, A. G. (1981). Phoneme restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, **110**, 474-494.

STEVENS, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

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

VAN DONSELAAR, W., KUIJPERS, C., & CUTLER, A. (1999). Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of Memory & Language*, **41**, 59-77.

ZWICKY, A. (1972). Note on a phonological hierarchy in English. In R. Stockwell & R. Macaulay (Eds.), *Linguistic change and generative theory* (pp. 275-301). Bloomington: Indiana University Press.

NOTES

1. The lexical frequency statistics are in the direction opposite to that of the corpus analysis (Patterson et al., 2003), in which higher frequency words had slightly higher deletion rates than did low-frequency words. In the corpus analysis, high-frequency words were slightly more likely to be realized in their vowel-deleted forms, relative to low-frequency words (4.3% difference).

2. No corpus statistics were available for a large enough set of low-variant-frequency legal cluster stimuli.

APPENDIX A
Word/Nonword Pairs and Corpus Deletion Rates, Followed
by Vowel Durations for Stimulus Tokens 2, 3, and 4, Expressed As a
Percentage of Full Vowel Length for Low- and High-Deletion Stimuli

Word/Nonword	Deletion Rate (%)	Stimulus Token		
		2	3	4
Low Deletion				
Calorie/talora	28.6	63.6	51.1	39.7
Criminal/priminam	17.5	40.0	29.0	20.0
Definite/befinap	21.1	67.4	48.3	37.1
Dominant/bominamp	16.7	56.3	44.8	33.3
Gallery/dillero	6.1	55.2	43.8	33.3
General/shenerom	32.4	58.8	49.0	40.2
Mineral/renering	3.7	47.4	37.7	28.1
Salary/falara	21.4	68.0	56.0	46.0
High Deletion				
Average/everode	88.9	54.3	45.8	35.8
Broccoli/droccola	69.1	65.3	53.4	41.2
Camera/tameri	89.3	54.3	43.8	31.4
Catholic/patholat	88.5	59.0	49.2	39.3
Chocolate/shocolipe	96.9	47.2	37.6	28.8
Corporate/pilporeck	91.6	65.4	56.7	48.1
Factory/sactoray	69.8	57.4	47.8	32.2
Family/samilo	94.9	55.7	44.2	34.6
History/sistoro	77	46.0	37.3	28.6
Opera/eperi	90.3	56.5	45.4	35.2
Separate/feparoke	100	60.2	47.6	35.9
Practice Word/Nonword				
Cardinal/pardinim	37.5	57.1	51.7	29.9

APPENDIX B
Stimuli Used in Experiments 2 and 3

High Deletion (Legal)	High Deletion (Illegal)	Low Deletion (Illegal)
broccoli	average	avalanche
chocolate	camera	calorie
history	catholic	salary
opera	family	criminal
gasoline	memory	definite
factory	severance	dominant
separate	ivory	gallery
corporate	grocery	mineral