Article

Language
and Speech

Language *and* Speech
2014, Vol. 57(1) 42–67

DOI: 10.1177/0023830913479105
las.sagepub.com
SAGE

Phonological Variant Recognition: Representations and Rules

Eleni Pinnow
Binghamton University, USA; University of Wisconsin-Superior, USA

Cynthia M Connine
Binghamton University, USA

Abstract
The current research explores the role of lexical representations and processing in the recognition of phonological variants. Two alternative approaches for variant recognition are considered: a representational approach that posits frequency-graded lexical representations for variant forms and inferential processes that mediate between the spoken variant and the lexical representation. In a lexical decision task (Experiment 1) and in a phoneme identification task (Experiment 2) using real words, low-frequency variants, but not high-frequency variants, show improved recognition rates following additional experience with the variants. This knowledge generalized to novel variant forms. Experiment 3 replicated these results using an artificial lexicon and showed that recognition of low-frequency variants was influenced by similarity to a high-frequency variant form. Similarity to a high-frequency variant alone, however, was insufficient to explain recognition of the infrequent variants (Experiments 4 and 5). The results support a hybrid account of variant recognition that relies on both multiple frequency-graded representations and inference processes.

Keywords
spoken word recognition, phonological variation, lexical representation

I Introduction

At the heart of spoken word recognition research is the question of how variability in the speech signal is accommodated during processing. At the lexical level, words have alternative pronunciations that are systematic based on local relationships between words, phonemes or phonetic features. Variation also can occur in an irregular way, such as mispronunciations or background noise.

Corresponding author:
Eleni Pinnow, Department of Human Behavior, Justice and Diversity, University of Wisconsin-Superior, Belknap & Catlin, Superior, WI 54880, USA.
Email: epinnow@uwsuper.edu

Despite these multiple sources of variation, listeners are generally able to pick out meaningful units (words) and map the speech onto stored lexical representations. The questions addressed in the current research are two-fold: firstly, how does experience with pronunciation variants influence the structure of lexical representations? Secondly, how do pronunciation variants interface with those representations?

Two primary accounts for how pronunciation variants are recognized have been proposed. In inference-based or processing accounts, a single, abstract lexical representation resides in lexical memory; pronunciation variants that differ from the lexical representation undergo an inference process in which the mismatching segments are transformed to match the lexical representation (Gaskell & Marslen-Wilson, 1998). Creating a representation from the input accommodates variability introduced by pronunciation variants. The inference-based approach requires that the necessary conditions for the pronunciation variant are available. For example, in regressive place assimilation a word ending in a coronal segment takes on a bilabial place of articulation only when the following word begins with a bilabial. For the inference rule to be utilized in perception, the bilabial environment must be available: /lim/ can be recognized as an acceptable variant form of lean if it precedes bacon, but not if it precedes ham. Lexical status also influences the application of inferential rules. In the place assimilation environment (e.g., /freIt beəʳ ɜːʳ/–/freIp beəʳ ɜːʳ/), Gaskell and Marslen-Wilson (1998) found that listeners reported hearing an underlying /t/ even when the assimilation occurs in a non-word (e.g., /preIt beəʳ ɜːʳ/), although the effect was smaller than when the assimilation occurs in a word. This suggests that both the licensing context environment as well as the lexical status of the assimilated stimulus influences the inferential process. In a related view, Pierrehumbert (2006b) argues that inferential processes are based on the distribution of the variant forms in the language such that large sets of variants with the same structure more readily lead to generalization of inferential process. In addition, variants that are more certain to undergo a change are more likely to be processed by inference thus reducing the possibility of error. These two hypotheses complement Gaskell and Marslen-Wilson's proposed inferential processes and offer an explicit statement of how listeners might develop these processes.

A second class of accounts for processing pronunciation variants assumes that systematic variation is encoded in lexical representations. In episodic models (Goldinger, 1998; Johnson, 1997), all instances of a spoken word are stored in the lexicon and include both systematic and idiosyncratic information. The content of the activated instances (what words are activated) and the intensity (strength of activation) determines if the speech input will be recognized as a known word. In episodic models, similarity between the speech input and the previously encountered instances is crucial for word recognition, because the speech input acts as a memory probe that activates similar-sounding instances. In other representational models (Connine, 2004; Connine & Pinnow, 2006; Ranbom & Connine, 2007), multiple representations for a given word are present but these are assumed to be abstract, and weighted by frequency of occurrence of a given variant, for a given word. In representational accounts, no specialized inference mechanisms are required (see also McLennan, Luce, & Charles-Luce, 2003; Pierrehumbert, 2006a; Sumner & Samuel, 2005). In both episodic and abstract representational schemes, frequency matters in recognizing pronunciation variants, either via the presence of multiple episodic traces or via multiple frequency-weighted abstract representations.

The pronunciation variant investigated in the current experiments is schwa vowel deletion. The schwa is a short unstressed vowel (such as the first vowel in *belong*) that can be deleted in some environments. Schwa deletion varies in frequency of occurrence across different words; some words occur frequently in their schwa-deleted form, such as *catholic* (/kæθəlɪk/–/kæθlɪk/) and others occur infrequently in their schwa-deleted form, such as *belong* (/bəlɔŋ/-/blɔŋ/). In a corpus analysis of conversational speech between native speakers of American English (Switchboard corpus), Patterson, LoCasto, and Connine (2003) analyzed schwa deletion frequency in two- and three-syllable words. One major finding was that three-syllable words had overall higher rates of schwa deletion than two-syllable words (55% versus 6.7%). In addition, the three-syllable words showed a wide range of deletion frequency (range: 0–100%). In contrast to the three-syllable words, two-syllable words showed a much more uniformly low deletion rate (range: 0–18%); only one two-syllable word fell outside of this range: *suppose* (average deletion rate: 45.6%).

Investigations of how schwa-deleted words are processed point to some commonalities across deleted variant forms in that, overall, schwa-deleted stimuli are rated as less acceptable than their schwa-present counterparts (LoCasto & Connine, 2002). Both infrequent (two-syllable) and frequent (three-syllable) schwa-deleted pronunciations showed this pattern, but the difference in acceptability ratings across the schwa-deleted/schwa-present forms was greater for infrequent schwa-deleted words. Two-syllable schwa-deleted forms were also less likely to be labeled as real words and were responded to more slowly in a lexical decision task than the three-syllable schwa-deleted forms.

While the preceding studies suggest a general processing cost for words produced with missing information, other research suggests that the mere absence of information is not always predictive of poor performance. Connine, Ranbom, and Patterson (2008) selected three-syllable words with high deletion (greater than 50%) and low deletion (less than 50%) rates. Listeners were presented with speech continua in which the duration of a schwa vowel was manipulated and their task was to determine whether a schwa vowel was present or absent. Listeners made similar judgments for acoustically matched non-word continua. An influence of deletion rate was revealed in three aspects of the data. Firstly, low deletion rate words showed more schwa vowel-present judgments as compared with high deletion rate words. Secondly, control non-word carriers with the same physical schwa vowel information (and surrounding segments) as their high and low deletion rate word counterparts did not differ from one another. This indicates that the deletion rate effect for words was not a consequence of idiosyncratic properties of the schwa vowel or its environs. Thirdly, both high and low deletion rate words showed more vowel-present judgments relative to their non-word counterparts, but the difference was larger for the low deletion rate words.

The variant-frequency effects found in the preceding studies provide a starting point for our research. The experiments investigated the role of experience with a specific variant form of a word on later recognition through two questions. Firstly, does experience with a variant form of a word influence later recognition, and does the effect of experience depend on pre-existing variant frequency? A second question investigated in the experiments was the role of experience with the general category of pronunciation variants. Does experience with a group of words that are heard as a specific pronunciation variant generalize to processing a new group of words with the same pronunciation variant properties? Experiment 1 used lexical decision to examine the influence of variant frequency of schwa-deleted pronunciations (high-frequency deletion or low-frequency deletion) and exposure to variants. The exposure conditions included three learning groups: a control condition in which listeners performed a lexical decision without additional experience; a repeat condition in which listeners heard the schwa-deleted forms in a training phase and then performed a lexical decision task on those same forms; and a transfer condition in which listeners heard a set of schwa-deleted forms in training and then performed a lexical decision task on a new set of schwa-deleted forms (transfer condition). We anticipate, based on previous work, that high-frequency schwa-deleted variant forms will be recognized easily in all three conditions. Experiment 1 sought to extend these previous findings by examining the nature of additional experience. Does experience with a schwa-deleted variant form enhance its subsequent recognition? Does experience with a set of words presented in a schwa-deleted variant form transfer to a new set of words presented in schwa-deleted variant form? Do these effects depend on pre-existing variant frequency?

2 Experiment 1

2.1 Method

2.1.1 Participants. A total of 75 undergraduates enrolled in psychology courses at Binghamton University participated for course credit. All participants were native speakers of English and reported no hearing disorders.

2.1.2 Materials and design. The naturally occurring differences in schwa-deletion frequency based on corpus analyses were used to create three sets of stimuli (from Patterson et al., 2003). Fourteen three-syllable words had high schwa-deletion rates (average deletion rates greater than 75%), 16 three-syllable words had low schwa-deletion rates (average deletion rates less than 40%) and 18 two-syllable words had very low deletion rates (average deletion rates less than 15%). (The two-syllable stimulus *suppose* was not included because of its higher deletion rate.) A complete list of stimuli used in Experiment 1 can be found in Appendix A. The stimulus groups did not differ in lexical frequency per million based on log-transformed frequency counts, $F(2, 45) = 1.4$, $p = .24$. A set of non-transformed frequencies is provided for reference: Francis and Kucera (1982): three-syllable high: mean = 80, $SD = 31$, range 1–405; three-syllable low: mean = 48, $SD = 90$, range = 1–378; two-syllable very low: mean = 63, $SD = 82$, range = 1–336; $F < 1$; Celex (1993): three-syllable high: mean = 941, $SD = 1695$, range = 16–1659; three-syllable low: mean = 285, $SD = 262$, range = 16–303; two-syllable very low: mean = 535, $SD = 208$, range = 1–3639.

In order to create the training and lexical decision lists, the 48 experimental stimuli were divided into two sets of 24 items; each set contained equal numbers of three-syllable high, three-syllable low and two-syllable low stimuli. In the repeat condition, the set of stimuli heard in training were also presented in the lexical decision task (set 1 was presented in training and set 1 was presented at test or set 2 was presented in training and set 2 was presented for lexical decision). In the transfer condition, one set of stimuli was presented in training and the second set was presented in the lexical decision task (set 1 was presented in training and set 2 was presented for lexical decision or set 2 was presented in training and set 1 was presented for lexical decision). In the control condition, a single set of stimuli was presented (set 1 or set 2 were presented for lexical decision). Both sets functioned as a training and lexical decision set (or as a control set), but were counterbalanced such that each set functioned as a training and test set across equal numbers of participants.

For the lexical decision task, a set of 24 word fillers were selected with the following properties: 15 two-syllable filler words had a medial consonant cluster that matched the medial clusters of the variant versions of the three-syllable stimuli; nine one-syllable filler words had complex word onsets that matched the onset clusters of the schwa-deleted variant versions of the two-syllable stimuli. The word fillers and experimental words had similar Francis and Kucera (1982) lexical frequencies (mean = 95, mean = 72, filler and experimental, respectively). A set of 48 non-word fillers were created that consisted of 30 two-syllable stimuli with medial segments that matched the schwa-deleted form of the three-syllable variants and 18 one-syllable stimuli with a complex onset that matched the schwa-deleted form of the two-syllable words.

All stimuli were recorded by a native speaker of American English (CMC) in a soundproof recording booth and were digitized at 44 kHz (16-bit resolution). The average stimulus durations were 580, 584 and 516 ms for the three-syllable high, three-syllable low and two-syllable very low, respectively.

2.1.3 Procedure. The experiment was divided into two sections: a training phase and the lexical decision phase. In the training phase, the schwa-deleted variant words were presented binaurally over headphones. The intended word was indicated by presentation of a correctly spelled orthographic version of the word following the auditory version. During training, instructions emphasized careful listening to each word and noted that some words may have unusual pronunciations.

A lexical decision task immediately followed the training phase. Participants were instructed to make a word/non-word judgment quickly and accurately by pressing an appropriately labeled response key. During the test, participants heard 48 non-words, 24 filler words and 24 schwa-deleted variants.

Participants were randomly assigned to control, transfer or repeat conditions, and to one of the list/test combinations. In the repeat condition, participants heard the same list of schwa-deleted variants during the training and lexical decision phases. In the transfer condition, one list functioned as training and the second list, containing novel schwa-deleted variants, functioned as the test. List function was counterbalanced across participants. A control group only completed the lexical decision phase on the schwa-deleted variants. Participants and conditions were distributed equally across lists.

2.2 Results and discussion

Word responses for schwa-deleted stimuli were scored as correct. Reaction time was measured from stimulus offset and outlier responses (slower than 1750 ms) were excluded (<2% of the data). A significance level of $p < .05$ was used for all analyses. Average accuracy rates and reaction time for correct 'word' responses as a function of stimulus type and learning group are shown in Table 1. Accuracy rates for filler stimuli did not differ across the conditions (non-words: 90.3%, 88.5%, 90.6%; words: 93.8%, 94.7%, 94.8%, repeat, transfer and control, respectively, all F values < 1). Visual inspection of the accuracy rates for fillers revealed a positive skew; consequently, the data were log transformed prior to analysis.

A 3 (condition: repeat, transfer or control) × 3 (stimulus type: three-syllable high-deletion, three-syllable low-deletion, two-syllable very low-deletion) analysis of variance (ANOVA) on the log-transformed percent correct showed a main effect of Condition, $F1$ (2, 72) = 14.51 MSE = .021; $F2$ (2, 21) = 9.81; MSE = .026. The repeat condition had the highest overall accuracy (85.2%), followed by the transfer (81.1%) and control (72.3%) groups. A main effect of stimulus type indicated

Table 1. Percent 'word' judgments presented as a function of experimental group and variant type. Average reaction times (ms) for 'word' judgments, measured from the offset, are shown in parentheses.

	Experimental group		
	Repeat	Transfer	Control
Three-syllable high deletion	97.3 (241)	95.3 (330)	97.6 (283)
Three-syllable low deletion	97.7 (313)	93.1 (390)	89.6 (379)
Two-syllable very low deletion	60.7 (516)	55.0 (564)	29.7 (453)

that accuracy increased as variant frequency increased, *F1* (2, 72) = 185.26; *MSE* = .010; *F2* (2, 21) = 23.40; *MSE* = .013.

A series of planned comparisons assessed the repetition effect (repetition versus control) and the transfer effect (transfer versus control). Repetition of the stimuli from training to test facilitated processing, but only for the low-frequency variants. The repetition effect was significant for three-syllable low, *F1* (1, 72) = 6.7, *MSE* = .009; *F2* (1, 21) = 6.3, *MSE* = .011, and two-syllable very low, *F1* (1, 72) = 14.3, *MSE* = .019; *F2* (1, 21) = 10.1, *MSE* = .02, but not for three-syllable high variants, $F < 1$.

The generalization effect was also influenced by pre-existing variant frequency; two-syllable very low stimuli benefitted from the generalized exposure, *F1* (1, 72) = 13.4, *MSE* = .021, *F2* (21) = 6.0, *MSE* = .014. The trend for three-syllable low stimuli (3.5%) was not significant, *F1* = 1.42, $p = .23$; *F2* = 2.32, $p = .14$, and no effect was found for three-syllable high stimuli, $F < 1$.

A parallel ANOVA conducted on reaction time showed a significant main effect for stimulus type, *F1* (2, 72) = 98.9, *MSE* = 9754; *F2* (2, 21) = 30.5, *MSE* = 12,579; lexical decisions were faster to more frequently occurring variant forms. The main effect of condition was marginally significant in the subjects analysis, *F1* (2, 72) = 2.8, $p = .06$, *MSE* = 41,904; *F2* (2, 42) = 2.8, *MSE* = 11,191, and the interaction of group and stimulus type was significant in the subjects analysis, *F1* (4, 144) = 2.8; *MSE* = 9677; *F2* = 1.8, *MSE* = 81442, $p = .16$. Numerically, the repeat condition was faster than the control for the three-syllable high and low deletion stimuli; the repetition effect was significant across items for the low, *F1* (1, 72) = 3.24, $p = .07$, *MSE* = 14396; *F2* (1, 21) = 11.82, *MSE* = 1662, and high-frequency variants, *F1* (1, 72) = 2.43, *MSE* = 8076, $p < .1$ .12; *F2* (1, 21) = 4.25, *MSE* = 1724, $p = .055$. The numerical trend was reversed for the two-syllable low deletion variants, but this was not significant, *F1* (1, 72) = 1.79, *MSE* = 24,395; *F2* (1, 21) = 1.49, *MSE* = 33,243.

The transfer condition was slower than the control for all variants but was significant only for the three-syllable high variants across subjects and marginally significant across items, *F1* (1, 72) = 8.2, *MSE* = 13,469; *F2* (1, 21) = 5.04, *MSE* = 1351, $p = .06$. No transfer effect was found for the three-syllable or two-syllable low deletion variants, *F1* (1, 72) = 3.48, *MSE* = 16,740, $p = .07$; *F2* (1, 21) < 1; *F1* < 1, *F2* (1, 21) = 1.23, *MSE* = 26,560.

A final analysis considered in more detail the influence of the training phase. Of particular interest was whether the influence of experience was confined to the schwa-deleted variants at test, or extended to a more general willingness to accept stimuli as words. In order to address this issue, a measure of discriminability, *d′*, was calculated. In Signal Detection Theory, *d′* is a measure of sensitivity that compares correct responses (hits) to false alarms. Higher *d′* values suggest that the correct responses can easily be discriminated from the false alarms (Macmillan & Creelman, 2005). The value of *d′* was calculated comparing responses to the schwa-deleted stimuli to non-word fillers for each condition. A measure of bias, *C*, was also calculated: positive values for *C* indicate a conservative bias, a decreased willingness to label a stimulus as a ‘word’; negative values indicate a bias towards a liberal labeling of words (Macmillan & Creelman, 2005). Values for *d′* and *C* by stimulus type and learning group can be found in Table 2.

An analysis using the *d′* and *C* values mirrored the analysis of the percent correct and reaction time data. A 3 (condition: repeat, transfer or control) × 3 (stimulus type: three-syllable high-deletion, three-syllable low-deletion, two-syllable very low-deletion) ANOVA on *d′* values by subject revealed a main effect of stimulus type, *F* (2, 72) = 14.48, *MSE* = 1.110. Paired comparisons revealed that three-syllable stimuli were more readily discernible from non-word fillers than the two-syllable stimuli (three-syllable high deletion compared to two-syllable, *t* (72) = 19.313, *SD* = 1.56, and that three-syllable low deletion stimuli were more readily discernible from non-word

Table 2. *d'* is presented as a function of experimental group and variant type; *C* is presented in parentheses. Average reaction times (ms) for 'word' judgments, measured from the offset, are shown in parentheses.

	Experimental group		
	Repeat	Transfer	Control
Three-syllable high deletion	5.11 (−1.10)	5.23 (−1.11)	5.48 (−1.04)
Three-syllable low deletion	5.13 (−1.11)	4.94 (−.97)	4.08 (−.34)
Two-syllable very low deletion	2.05 (.43)	1.86 (.57)	1.14 (1.13)

fillers than two-syllable low deletion stimuli, *t* (72) = 12.19, *SD* = 2.14). There was no difference between the three-syllable stimulus types, *t* (72) =1.21, *SD* = 3.24, and no effect of learning group, *F* (2, 72) = .998, *MSE* = 4.744. The corresponding ANOVA analyzing *C* found a main effect of stimulus type, *F* (1, 72) = 42.395, *MSE* = .277, showing that the three-syllable stimuli were responded to with a more liberal bias than the two-syllable stimuli. There was no effect of learning group, *F* (2, 72) = 1.372, *MSE* = 4.739.

Recognition of schwa-deleted variants depended on pre-existing variant frequency, with better recognition for high than for low-frequency variants. The effect of experience with a schwa-deleted variant of a word also depended on the pre-existing variant frequency: hearing a schwa-deleted variant in training benefitted later recognition at test more so for low than for high-frequency variants. These findings suggest that schwa-deleted variants are represented in a frequency-sensitive fashion. The transfer condition did not generally influence processing at test but there was a trend for the two-syllable low-frequency variants. Given the trend in a transfer effect, caution is in order in interpreting these results, but experience with a set of variants that transfers to a new set suggests abstraction of general knowledge about a given variant form that can later be applied to assist recognition of similar forms. Experiments 3, 4 and 5 address this issue further.

Importantly, an analysis of the discriminability of the schwa-deleted stimuli also showed an effect of pre-existing variant frequency. Regardless of the learning group, stimuli with higher frequency deletions (three-syllable stimuli) were more readily discriminated from fillers than lower frequency stimuli. This lends further support to the claim that pre-existing variant frequency influences subsequent recognition. In addition, the relatively low discrimination shown for the two-syllable stimuli indicates that participants were not treating all recently heard stimuli as words – rather the two-syllable words were difficult to distinguish from the novel non-word fillers.

Experiment 2 addresses a potential concern in using the training and test paradigm: in the training phase, participants were told that some of the stimuli may have unusual pronunciations but are in fact words. Subsequently, at test, the stimuli are presented in a task that mirrors what they were told about the stimuli during training (e.g., the schwa-deleted stimuli are words). Thus, while Experiment 1 clearly demonstrates an effect of variant frequency, the task directions may have inadvertently influenced the results and produced a general acceptance of schwa-deleted variants as words. The *d'* analysis did not support this interpretation of the results, but a replication using a new task would provide further support for the role of experience in processing variants.

In order to address this concern, Experiment 2 investigated the influence of experience on variant processing using a paradigm in which the task at training was divorced from the task at test. As in Experiment 1, a training phase familiarized participants with the two- and three-syllable low variant-frequency forms. In the test phase, speech continua were presented in a phoneme identification task. One set of continua varied from a variant endpoint to a non-word endpoint (e.g., *crimnal–grimnal*, /krImnʌl/–/grImnʌl/). A second set of continua served as a control and varied from a non-word endpoint to a non-word endpoint (*cribnal–gribnal*, /krIbnʌl/–/grIbnʌl/). In the phoneme identification task, no explicit judgment about the lexical status of a schwa-deleted variant is required. We predict, consistent with the results from Experiment 1, that more phoneme identifications consistent with the variant form relative to the control should be made in the repeat and transfer conditions.

3 Experiment 2

3.1 Method

3.1.1 Participants. Seventy-eight undergraduates at Binghamton University, all native English speakers with normal hearing, participated for course credit. Participants had not previously participated in Experiment 1.

3.1.2 Materials and design. Six two-syllable, low schwa-deletion words used in Experiment 1 (average = 6.7% deletion rate) were selected with a voiced (*believe, belong, balloon*) or voiceless onset (*police, pollute, parole*). A set of three-syllable low schwa-deletion words with a voiceless onset was also selected (average = 27.5% deletion rate) (*criminal, cardinal, cabinet*). For each variant form of a word, a voicing change in the initial segment created a non-word (see Appendix B). Three-syllable voiced onset stimuli were not included because the voiceless-to-voiced change formed real words (word–non-word pairs were not available). Variant forms that began with a voiced onset (/g/ or /b/) were termed 'voiced bias' and variant forms that began with a voiceless onset (/k/ or /p/) were termed 'voiceless bias.' The non-word and variant word pairs were recorded by the same speaker used in the previous experiment.

To create the voicing continua from the recordings of the variant word–non-word pairs, the acoustic energy of the voiced and voiceless onsets was manipulated (cf., Connine & Clifton, 1987). Using this method, voice-onset time as a cue to voicing status is approximated by combining information from the voiced and voiceless segments. Voicing continua were created by excising successive portions of periodic energy from the voiced stimulus, starting from the onset of the release. These portions were replaced with equal length acoustic portions of aperiodic energy taking from the voiceless stimulus, starting from the onset of the release. All splice points were made at zero crossings in step sizes of 3–5 ms and the stimuli created in this fashion had no audible disfluencies. Twenty tokens were created progressing from variant word to non-word endpoint.

For each variant word–non-word continuum, a companion non-word–non-word continuum was created. The non-word–non-word continua were created by altering the final segment in the original variant word–non-word continua. Specifically, a new set of non-word stimuli were created from the variant words by changing the final segment (e.g., *cabnef* from *cabnet*). Each of the changed offset non-word stimuli was recorded and the final segment was excised. The excised final segment was used to replace the final segment in the original variant word-to-non-word continua (e.g., the final /t/ in the *cabnet–gabnet* continuum was replaced with the final /f/ from the '*cabnef*' recording to create a *cabnef–gabnef* continuum). The resulting non-word–non-word control continuum was identical to its variant word–non-word counterpart except for the final segment; this ensured that the critical acoustic information was the same for the variant–non-word continuum and its non-word–non-word continuum counterpart. In this way, differences in identification of the initial segment between the variant word–non-word and non-word–non-word conditions can be attributed to the lexical status of the variant word.

To accommodate the transfer condition, a separate set of variants was selected for the training phase (*collect, correct, peruse, select, supply, supreme, gasoline, broccoli, catholic*). The variant form of each item was recorded by the same speaker as the previous stimuli using the same methods. Note that this new set of variant forms was never presented at test. In addition, the segmental contents of the transfer training stimuli (the consonant clusters created by schwa deletion) were diverse and, with the exception of *peruse*, these consonant clusters differed from those encountered in the test stimuli. In the transfer condition, the diversity of onset clusters from training to test ensured that participants could not learn a rule about schwa vowel deletion based only on local segmental dependencies.

In order to select stimuli from the midrange of the continua for presentation in the experiment, a pilot study was conducted with five listeners, using the entire set of stimuli. Three midrange stimuli (identical across variant bias and non-word continua) that received a mixture of voiced and voiceless responses (the voicing status was ambiguous) were selected.

3.1.3 Procedure. Participants were randomly assigned to one of three groups: repeat, transfer and control. A training task, similar to Experiment 1, preceded the phoneme identification task. Auditory presentation of the variant (the variant endpoint token from each continuum) was combined with an orthographic rendering along with a classification of lexical status (word or non-word). Participants were asked to listen carefully to each variant, read the orthographic rendering and note the lexical status. Each variant endpoint was presented five times.

In the subsequent phoneme identification task, the three ambiguous stimuli, along with the unambiguous endpoints, were presented. Presentation was blocked such that the variant–non-word and non-word–non-word continua were presented in a single block. A block for a given variant contained the five stimuli from each continuum, repeated six times in random order. Stimuli were presented across two days of testing, blocked by variant frequency. Thus, participants made binary choices in the phoneme identification task each day for a given contrast/variant frequency (day 1: low-frequency variants, /b/ or /*p*/, day 2: high-frequency variants, /g/ or /k/). The response box included only the phoneme labels that were relevant to the current day of testing (/b/-/*p*/ *or* /g/-/k/). The experiment was spread across two days, with the low-frequency variants presented first to avoid a potential transfer of knowledge about schwa-vowel deletion gained from hearing the high-frequency variants. Continuum order was randomized across participants on each test day.

The repetition group heard the set of schwa-deleted variant endpoints in training and the corresponding set of continua for phoneme identification. The transfer group heard one set of schwa-deleted variants at training and a different set of continua for phoneme identification. The control group completed the phoneme identification task without prior experience. All groups heard the same set of stimuli at test.

Instructions for the phoneme identification task directed participants to indicate the stimulus-initial phoneme by pressing an appropriately labeled response key; accuracy was emphasized.

3.2 Results and discussion

The percentage voiceless responses are presented in Table 2 averaged across the three intermediate stimuli as a function of continuum type (variant versus non-word) and training condition (see Appendix A for the complete data set). Table 3 also displays the size of the variant effect (change, the difference between the variant–non-word and non-word–non-word continua).

A series of ANOVAs were conducted to evaluate specific hypotheses for voiced and voiceless bias two-syllable or voiceless bias three-syllable words. In each analysis, a training condition

Table 3. Percentage average voiceless responses presented as a function of training condition (control, repeat, transfer), variant type (two syllable, three syllable) and continuum type (variant–non-word, non-word–non-word). The change column indicates the difference in percentage voiceless responses from the variant–non-word to the non-word–non-word continua.

	Control condition		
	Variant–non-word	Non-word–non-word	Change
Two-syllable			
Voiced bias	45	48	+3
Voiceless bias	61	47	−14
Three-syllable			
Voiceless bias	69	63	−6
	Repeat condition		
	Variant–non-word	Non-word–non-word	Change
Two-syllable			
Voiced bias	33	44	+11
Voiceless bias	67	48	−19
Three-syllable			
Voiceless bias	63	54	−9
	Transfer condition		
	Variant–non-word	Non-word–non-word	Change
Two-syllable			
Voiced bias	43	51	+8
Voiceless bias	61	49	−12
Three-syllable			
Voiceless bias	74	68	−6

(repeat or transfer) was compared to the control condition as a function of continuum type. A 2 (continuum: variant versus non-word) × 2 (training: repeat versus control) ANOVA for the voiced bias stimuli revealed a main effect of variant–non-word, F (1, 53) = 27, MSE = 161; variant continua showed more voiced responses than control continua. Similarly, there were more voiceless responses in the control condition than in the repeat condition, F (1, 53) = 132, MSE = 172. In addition, there was an interaction of lexical status and training condition, voiced: F (1, 53) = 6.7, MSE = 161.5. In a corresponding 2 (continuum: variant versus non-word) × 2 (training: repeat versus control) ANOVA on the voiceless bias stimuli, the two-syllable very low variant-frequency stimuli showed no significant effect of training, F (1, 53) = 1.71, MSE = 182. The only significant effect was for continuum, F (1, 53) = 38, MSE = 139, indicating that the variant–non-word continuum had more voiceless responses than the non-word–non-word continuum. There was no significant interaction between continuum and training, $F < 1$.

A 2 (continuum: variant versus non-word) × 2 (training: transfer versus control) ANOVA for the voiced bias stimuli examined the influence of experience with a novel set of variant forms on subsequent variant recognition. The voiced bias two-syllable very low variant-frequency stimuli showed a marginally larger variant effect in the transfer group relative to the control, F (1, 48) = 3.34, p = .07, MSE = 407. This transfer effect was not seen in the voiceless bias two-syllable variants, $F < 1$. Both the voiced bias and voiceless bias stimuli showed a main effect of continuum, voiced two-syllable stimuli: $F(1, 48) = 22$, $MSE = 121$; voiceless two-syllable stimuli: $F(1, 52) = 22$, $MSE = 116$.

In a 2 (continuum: variant versus non-word) × 2 (training: transfer versus control) ANOVA on the three-syllable stimuli there were no significant effects in the three-syllable stimuli when comparing the control group to the transfer group, $F < 1$. However, there was a main effect of continuum (three-syllable stimuli: $F(2, 98) = 24.4$, $MSE = 151$). The interaction between continuum and training condition was not significant, $F < 1$.

Similar to Experiment 1, experience influenced processing of variant forms (as evidenced by the variant effect), but the effect of experience was modulated by the pre-existing variant frequency. Together, Experiments 1 and 2 offer converging evidence that variant frequency influences processing of schwa-deleted variants and that additional experience with a schwa-deleted variant enhances recognition of later encounters with the variant. The effects of variant frequency and experience suggest that variant frequency is lexically represented. The evidence for abstraction of knowledge about a class of variant forms for later processing is somewhat supported by the transfer effects, but these effects were less robust and more variable.

Experiments 1 and 2 provided some new evidence about lexical representation and abstraction in processing low-frequency variants, but some caution is in order in the interpretation of the results. The availability of high- and low-frequency variants in English required that the variant-frequency manipulation was to some extent confounded with word length. This raises the possibility that the different pattern of responses based on variant frequency may be due to differences in word length: high-frequency variants provided a source of lexical constraint not available to low-frequency variants and this may have contributed to the variant-frequency effect. Conversely, lexical constraints may have decreased sensitivity to variant frequency and muted the impact of experience. The limitations of using real words in Experiments 1 and 2 are addressed in the next three experiments.

4 Experiment 3

Experiments 3, 4 and 5 addressed the limitation of relying on real words by investigating the role of variant frequency and experience using an artificial lexicon. Artificial lexica have been used successfully in the field of spoken word recognition to address a number of issues. For example, Magnuson, Tanenhaus, Aslin, and Dahan (2003) used an artificial lexicon to investigate lexical frequency in spoken word recognition. Following training in which pseudowords were presented either one time (low frequency) or seven times (high frequency), high-frequency pseudowords were more accurately recognized than low-frequency pseudowords. Tanenhaus, Magnuson, Dahan, and Chambers (2000) showed that learned pseudowords activated pseudoword competitors, mirroring effects found for real words. In a related set of studies, Gaskell and Dumay (2003) examined how newly learned words interacted with existing words in the lexicon. In their experiment, pseudowords were learned that had onset overlap with a real word (e.g., *cathedruke–cathedral* /kəθidruk/–/kəθidrul). With extended training, Gaskell and Dumay found that a learned pseudoword inhibited recognition of an onset overlap real word (but see Magnuson et al., 2003). Subsequent work demonstrated that learned pseudowords show competition effects that last up to eight months after testing (Tamminen & Gaskell, 2008), and are more easily confused with cohort competitors that share some segmental properties (Creel, Aslin, & Tanenhaus, 2006). The authors suggest that a process of condensation or restructuring occurs during lexicalization involving the specification of phonetic detail (see also Salasoo, Shiffrin, & Feustal, 1985). Taken together, these studies suggest that artificial lexicons are processed and represented similarly to words in the pre-existing lexicon.

The goal of Experiment 3 was to determine whether variant-frequency effects observed for real words would be found in an artificial lexicon. Variant frequency during learning was manipulated by varying the presentation frequency of a full and schwa-deleted form for a given pseudoword. An orthographic version accompanied presentation of the full and schwa-deleted form to indicate these were alternative pronunciations of the same word. The learning phase was followed by a test phase in which the variant forms were presented in a lexical decision task.

4.1 Method

4.1.1 Participants. Thirty-one Binghamton University students received partial course credit in exchange for their participation. They reported normal hearing and were native speakers of English.

4.1.2 Materials. Thirty-two pseudowords were constructed that contained a possible schwa deletion; all pseudowords were trisyllabic prior to schwa deletion and were reduced to two syllables in their schwa-deleted form (see Appendix C). The schwa-deletion site was in the second syllable; the stress pattern for the pseudowords (pre-deletion) was strong/weak/strong with primary stress on the first syllable. For each experimental pseudoword, an orthographic version was created for presentation during training. The orthographic version contained a vowel representing a schwa, indicated orthographically by vowels that were present in the written version of the schwa-deleted words examined in Patterson et al.'s (2003) corpus analysis (a, o, u, e).

A matching set of 32 pseudowords was constructed to serve as fillers during the auditory lexical decision test phase. Half of the fillers were three syllables and half were two syllables. Three-syllable fillers include a schwa in the medial syllable surrounded by consonants that could be combined to form a phonotactically legal cluster. Two-syllable fillers had a medial, phonotactically legal consonant cluster. The two- and three-syllable fillers were constructed to match the stress patterns of the pseudowords.

All pseudowords were recorded by a female native speaker of American English in a sound-dampened recording booth and digitized at 44 kHz (16-bit resolution).

During training, participants heard each of the 32 pseudowords eight times, resulting in 256 trials. Half of the pseudowords were presented as high-frequency schwa deletions and half of the pseudowords were presented as low-frequency schwa deletions. For high variant-frequency pseudowords, the schwa-deleted variant form was presented seven times (87.5% of repetitions) and the full schwa-present form was presented once (12.5%). For low variant-frequency pseudowords, the schwa-deleted variant form was presented once (12.5%) and the full schwa-present form was presented seven times (87.5%).

Two training lists were created. In each list, half of the items were presented as low-frequency schwa-deleted/high-frequency full form and the other half of the items were presented as high-frequency schwa-deleted/low-frequency full form. Item and variant frequency were counterbalanced across the two lists: an item that was presented as high-frequency full form/low-frequency schwa-deleted form in list one, was presented as low-frequency full form/high-frequency schwa-deleted form in list two.

Two test lists were created. Within a list, half of the pseudowords that were presented as a high variant-frequency schwa-deleted form/low-frequency full form during training were presented in their low-frequency full form during the test; the other half of the high variant-frequency schwa-deletion forms were presented in their high-frequency schwa-deleted variant form. The same pattern held for pseudowords that were presented as the low variant-frequency schwa-deleted form/ high variant-frequency full form during training: at test, half were presented in their low-frequency schwa-deleted form and half in their high-frequency full form during the lexical decision phase. Across lists full and schwa-deleted forms were counterbalanced such that a list included only the full or schwa-deleted form of a given pseudoword (no participant heard both forms of a pseudoword at test).

To summarize, the training lists included both schwa-deleted and full forms of each of the pseudowords; within a list, half of the pseudowords were presented as high-frequency schwa-deleted forms, while the other half of the pseudowords were presented as low-frequency schwa-deleted forms. The test lists included only one variant form of each of the pseudowords, counterbalanced for variant frequency (half high-frequency schwa-deleted variants, half low-frequency schwa-deleted variants); within each variant-frequency group, half of the stimuli were presented as the schwa-deleted form and half of the stimuli were presented as the full form.

4.1.3 Procedure. Participants were tested in groups of three or fewer in a sound-dampened chamber. All pseudowords were presented binaurally over headphones. A break was provided every 50 trials. Prior to the training and test phases, short practice sessions were presented to ensure participants understood the tasks. The entire experiment was completed in approximately 30 minutes.

A training phase preceded the test phase. On each trial, an auditory stimulus was presented simultaneous with an orthographic version of the stimulus on a computer screen. Participants were told that the experiment involved learning a list of new words and were instructed to listen carefully to each stimulus presented over the headphones and to read the written version of the word to confirm what they heard. Participants were told to press a key labeled ‘yes’ to acknowledge that they have heard the pseudowords and read the orthographic version. In the test phase, participants performed a lexical decision task. Participants were told to make lexical decisions as quickly and accurately as possible based on what they learned in the training session. If the stimulus that participants heard over the headphones is a real ‘word,’ based on the pseudowords learned during training, participants were told to press a key labeled ‘word.’ If what they heard was a non-word, they were told to press the ‘non-word’ key.

4.2 Results and discussion

Outlier reaction times (slower than 2500 ms from stimulus offset) were excluded from the analysis (less than 1% of the data). Participants with accuracy rates lower than 60% on the auditory fillers were also excluded from the analysis (five participants). Average accuracy rates and reaction times for the auditory lexical decision task are reported in Table 4; Table 5 shows the corresponding values for d' and C. Reaction times reflect only correct responses. Fillers were responded to with 80.0% accuracy (three-syllable fillers: 81.1%, and two-syllable fillers 78.9%). As in Experiment 1, the accuracy data had a positive skew, thus analyses are of the percent correct are based on log-transformed data.

A 2 (variant frequency: high frequency, low frequency) × 2 (form: full, variant) ANOVA of the percent correct data revealed that high-frequency pseudowords were recognized more accurately than low-frequency pseudowords, $F1$ (1, 26) = 6.469, MSE = .019; $F2$ (1, 31) = 9.37, MSE = .022. Pseudowords with a schwa present were also recognized more accurately than pseudowords without a schwa, $F1$ (1, 26) = 6.890 MSE = .013; $F2$ (1, 31) = 7.25, MSE = .014. There was no

Table 4. Average percent correct (word responses) and reaction times for learned pseudowords as a function of form frequency (high: seven repetitions/low: one repetition) and pseudoword form (full/variant) for Experiment 3.

	Pseudoword form	
Frequency	Full	Schwa-deleted variant
High (7 repetitions)	89.0 (286)	77.0 (349)
Low (1 repetition)	73.9 (407)	69.4 (415)

Table 5. *d*' values for learned pseudowords as a function of form frequency (high: seven repetitions/low: one repetition) and pseudoword form (full/variant) for Experiment 3. *C* is presented in parentheses.

	Pseudoword form	
Frequency	Full	Schwa-deleted variant
High (7 repetitions)	3.81 (−.86)	2.63 (−.25)
Low (1 repetition)	2.38 (−.28)	2.04 (−.01)

interaction between pseudoword frequency and schwa presence, *F1* (1, 26) = 1.279, *MSE* = .009; *F2* (1, 31) = 1.53, *MSE* = .011. A post hoc analysis in which items were categorized based on the grapheme used to represent the schwa vowel demonstrated that the specific orthographic rendering of the schwa did not influence accuracy rates (pseudowords with a potential schwa vowel deletion were learned equally well, regardless of the grapheme that represented the schwa, *F* < 1).

A corresponding 2 × 2 ANOVA analyzed the reaction time data. A main effect of frequency showed that high-frequency pseudowords were responded to faster than low-frequency pseudowords, *F1* (1, 25) = 14.274, *MSE* = 15,838; *F2* (1, 31) = 6.554, *MSE* = 35,110. There was no effect of pseudoword form on reaction time, *F1* (1, 25) = 1.562, *MSE* = 20,956; *F2* (1, 31) = 1.905, *MSE* = 21,981. Pseudoword frequency did not interact with schwa presence, *F1* (1, 25) = 1.278, *MSE* = 15,278; *F2* (1, 31) < 1.

As in Experiment 1, measures of sensitivity and bias were calculated based on the hit rates for different forms (full versus variant) and frequencies (high versus low) compared to false alarm rates for novel stimuli for each participant (see Table 5). A 2 (variant frequency: high frequency, low frequency) × 2 (form: full, variant) ANOVA analyzed the discriminability of the learned pseudowords from novel fillers. There was a significant effect of frequency, with high-frequency pseudowords more readily discerned from fillers than low-frequency pseudowords, *F* (1, 26) = 11.99, *MSE* = 2.18. In addition, the full forms were also more discriminable from the fillers, *F* (1, 26) = 8.12, *MSE* = 1.848. In a corresponding ANOVA examining effects of bias, *C*, the high-frequency pseudowords showed a stronger bias than the low-frequency pseudowords, *F* (1, 26) = 7.31, *MSE* = .608. Similarly, the full forms showed more bias than the variant forms, *F* (1, 26) = 13.21, *MSE* = .385.

Note also that high-frequency forms were more readily discerned from non-words than low-frequency forms. It is important to note that respondents showed liberal tendencies in their responses. Taken together, this suggests that the learned forms (both full and variant) were distinguishable from novel pseudowords.

For both full and deleted forms, recognition was better for the higher frequency variant. The use of pseudowords permitted other pre-existing lexical properties that might be confounded with variant frequency to be carefully controlled, yet the results are remarkably consistent with what was found for real words in Experiment 1. One striking aspect of the data, however, is the relatively high recognition rates for the low-frequency forms – these variant forms were heard only once during training, yet accuracy rates were quite high (69.4% and 73.9%, deleted and full forms, respectively). The high recognition rates for the low-frequency forms is unexpected by a frequency-weighted multiple representation scheme (given the very low frequency of these forms). In considering possible explanations for the relatively high accuracy rates for the low-frequency variants, we focused on the role of the high-frequency form in contributing to recognition of its low-frequency counterpart. There is a great deal of phonological overlap between the full and variant forms – this overlap may facilitate the learning, and/or subsequent recognition, of the low-frequency forms. One possible contribution of a representation of the high-frequency variant to recognition of low-frequency variants is that it may function as a recognition route for the low-frequency variant based on similarity; on this view, the similarity of the low-frequency variant to its high-frequency counterpart provides a route to recognition and there is no role for a representation of the low-frequency variant. The subsequent experiments address the question of whether low-frequency variants are simply recognized via similarity to their high-frequency counterparts. A first step in addressing this question is to determine whether the availability of a high-frequency, similar-sounding variant influences recognition of a low-frequency form. Do low-frequency variant forms benefit from a similar-sounding high-frequency counterpart compared with a low-frequency form with no similar-sounding high-frequency counterpart?

Experiment 4 addresses these questions by manipulating the availability of a phonologically similar (or dissimilar) form during training. Two sets of pseudowords were included. One set of pseudowords, variant/similar pseudowords, contained the full and variant forms from Experiment 3 and as such were highly similar (differing only in the presence/absence of the schwa vowel). A second set of pseudowords, dissimilar pseudowords, included full and variant forms with no phonological overlap (e.g., *teesatig* and *foprite*; /tisətIg/ and /foprait/). In both variant/similar and dissimilar pseudowords, the frequency of presentation of a given form was manipulated. A high-frequency advantage is expected for both stimulus sets. Of particular interest is whether the presence of a high-frequency variant contributes to recognition of a phonologically similar low-frequency variant. Are low-frequency variants with a similar-sounding high-frequency counterpart recognized better than low-frequency variants without a high-frequency counterpart?

5 Experiment 4

5.1 Method

5.1.1 Participants. Twenty-eight Binghamton University students received partial course credit in exchange for their participation. Participants reported normal hearing and were native speakers of English.

5.1.2 Materials. A set of 16 three-syllable pseudowords from Experiment 3 was selected to serve as the variant/similar stimuli. During training, the variant/similar pseudowords were presented 15 times in their full form (high frequency), and once in their variant schwa-deleted form (low frequency).

A set of 16 dissimilar pseudowords were selected to serve as the dissimilar stimuli; eight three-syllable stimuli served as the high-frequency condition and eight two-syllable stimuli served as the low-frequency condition. During training, the three-syllable pseudowords were presented 15 times (high frequency) and the two-syllable pseudowords were presented once (low frequency).

For the lexical decision phase, the set of pseudoword fillers from Experiment 3 was used to function as ‘non-words.’ Two test lists were created; each list included only one variant/similar form (schwa-presented or schwa-deleted form) along with the dissimilar high and low frequency. In each list, half of the variant/similar stimuli were high and half were low frequency. The lists were counterbalanced across items and participants were distributed equally across lists. The training phase was presented followed by the lexical decision task. All training and test procedures were identical to Experiment 3.

5.2 Results and discussion

As in Experiment 3, outlier reaction times (slower than 2500 ms from stimulus offset) were excluded from the analysis (less than 1% of the data). A single participant was excluded from the analysis for accuracy rates lower than 60% on the fillers. Participants correctly labeled fillers as non-words with 88.4% accuracy (two- and three-syllable fillers had accuracy rates of 92.4% and 84.3%, respectively). As in Experiments 1 and 3, the accuracy data had a positive skew; accordingly, the analyzed percent correct data were log-transformed. Table 6 shows average reaction time and accuracy as a function of frequency and pseudoword type; Table 7 provides *d′* and *C* values by frequency and pseudoword type. Reaction times reflect only correct responses.

A 2 (pseudoword type: variant/similar versus dissimilar) × 2 (frequency: high versus low) ANOVA on accuracy data revealed a main effect of pseudoword type, *F1* (1, 26) = 45.51, *MSE* = .156; *F2* (1, 40) = 33.4, *MSE* = .167: the variant/similar pseudowords were recognized more accurately than dissimilar pseudowords. There was a main effect of frequency: high-frequency pseudowords were recognized more accurately than low-frequency pseudowords, *F1* (1, 26) = 33.735, *MSE* = .192; *F2* (1, 40) = 15.54, *MSE* = .223. There was also an interaction between pseudoword type and frequency, *F1* (1, 26) = 26.69, *MSE* = .193; *F2* (1, 40) = 19.15, *MSE* = .187. Planned comparisons showed that the low-frequency variant/similar pseudowords were recognized more accurately than the low-frequency dissimilar pseudowords, *F1* (1, 26) = 27.88, *MSE* = .11; *F2* (1, 14) = 25.84, *MSE* = .167. High-frequency variant/similar and dissimilar pseudowords did not differ, *F* < 1.

Table 6. Percent correct (word responses) and reaction time for learned pseudowords presented as a function of pseudoword frequency and type for Experiment 4.

Pseudoword type		
Frequency	Variant/Similar	Dissimilar
High	88.1 (350)	84.3 (316)
Low	78.9 (440)	20.7 (247)

Table 7. *d*' for learned pseudowords presented as a function of pseudoword frequency and type for Experiment 4. *C* is presented in parentheses.

Pseudoword type		
Frequency	Variant/Similar	Dissimilar
High	3.89 (−.66)	3.88 (−.57)
Low	2.88 (−.14)	.50 (1.49)

A comparable 2 (pseudoword type: variant versus dissimilar) × 2 (frequency: high versus low) ANOVA on reaction time revealed a main effect of pseudoword type that was significant in the subject, but not the item analysis, *F1* (1, 26) = 6.36, *MSE* = 19573, *F2* (1, 40) = 1.41, *MSE* = 98,261. High-frequency pseudowords were recognized faster than low-frequency pseudowords, *F1* (1, 26) = 3.55, *MSE* = 102,593; *F2* (1, 40) = 6.59, *MSE* = 134,987. Planned comparisons showed no difference between low-frequency variant/similar pseudowords and low-frequency dissimilar pseudowords, *F* < 1; there was also no difference between the two high-frequency forms, *F* < 1. The relatively high variability in the reaction times likely contributed to the non-significant planned comparisons (particularly to the low-frequency forms).

As in previous experiments, measures of sensitivity and bias were calculated that compared the correct word responses to learned pseudowords to the false alarm rate of the novel pseudowords (Table 7). First a 2 (frequency) × 2 (pseudoword type) ANOVA analyzed *d′* values. High-frequency pseudowords were more readily discriminated from fillers than low-frequency pseudowords, *F* (1, 26) = 57.51, *MSE* = 2.10. Variant/similar pseudowords were more easily discerned from fillers than dissimilar pseudowords, *F* (1, 26) = 16.27, *MSE* = 2.05. There was also a significant interaction between frequency and pseudoword type, *F* (1, 26) = 11.766, *MSE* = 2.804. In regards to bias, dissimilar pseudowords showed a larger and more conservative bias than variant/similar pseudowords, *F* (1, 26) = 28.23, *MSE* = .705. Low-frequency pseudowords also showed a larger and more conservative bias than high-frequency pseudowords, *F* (1, 26) = 72.86, *MSE* = .622. The interaction of frequency and pseudoword type in the *C* values was also significant, *F* (1, 26) = 17.51, *MSE* = .920.

Experiment 4 established that hearing a low-frequency variant in the context of a similar-sounding high-frequency counterpart provides an advantage in recognition: low-frequency similar pseudowords were recognized better than the low-frequency dissimilar pseudowords. High-frequency variants did not differ across stimulus sets. Thus, the recognition of an infrequent variant is modulated by the availability of a similar-sounding high-frequency variant form.

One explanation for the pattern of results is that the representation of a low-frequency variant form is facilitated by the presence of a high-frequency form. This maintains the assumption of multiple representations of variants based on experienced frequency but also assumes that the lexicalization of low-frequency forms is facilitated by the availability of the high-frequency variant. On this view, the high-frequency form plays an important role in the formation of a lexical representation for the low-frequency variant. In the low-frequency dissimilar pseudowords, there is no representation of a high-frequency, similar-sounding variant form to facilitate lexicalization.

An alternative possibility is that no low-frequency variant representation exists. Rather, low-frequency variants are recognized solely via their similarity to a high-frequency variant. On this view, the low-frequency variant in the dissimilar set suffers because it has no similar high-frequency variant form and must be recognized solely based on its low-frequency representation.

Experiment 5 was designed to further investigate the role of a high-frequency variant in recognition of the low-frequency variant by isolating the effects of similarity in the absence of experience. Here, we examined false recognitions of a completely novel form given experience with a similar-sounding high-frequency form or a similar-sounding low-frequency form. During the learning phase, high- and low-frequency variants were presented in their full (schwa-present) form, but without their similar-sounding (schwa-deleted variant) counterpart. During the recognition phase, the high- and low-frequency variants were presented along with the never-heard but similar-sounding variant counterparts. Recognition of the novel stimuli must be based on experience with the learned counterparts and, as such, reflects reliance solely on similarity to the experienced pseudowords. Recognition rates for novel forms (false alarms) that are modulated by the experienced frequency of similar-sounding forms (as in Experiment 3) would support a role of similarity. In summary, the experiment addressed the following question: in the absence of any experience, is recognition of a novel form modulated by experience with a similar-sounding, high-frequency form?

6 Experiment 5

6.1 Method

6.1.1 Participants. Twenty-four students at Binghamton University participated for partial course credit; all reported normal hearing and none had participated in the previous four experiments.

6.1.2 Materials, design and procedure. A set of 32 three-syllable pseudowords with a potential medial schwa vowel deletion was taken from Experiment 3. During training, 16 pseudowords were presented 15 times in their full (schwa-present) form and 16 pseudowords were presented once in their full form. The training phase did not include any variant (schwa-deleted) forms of the pseudowords. Pseudowords were presented over headphones; an orthographic version of the pseudoword was presented simultaneously during training; the schwa was orthographically indicated. The procedure for the training session was identical to Experiments 3 and 4.

The lexical decision task followed the training phase and included the 32 previously learned stimuli and 32 lure pseudowords. The lure pseudowords were previously unheard (schwa-deleted) variant forms of the learned three-syllable pseudowords. Thirty-two filler (16 three-syllable, 16 two-syllable) pseudowords from Experiment 3 were used to function as ‘non-words.’ At test, the stimuli were pseudo-randomized in a single list such that the sequential distance between a learned pseudoword and its novel variant form was maximized and learned; novel forms and ‘non-words’ were distributed equally across the list. The test procedure was identical to Experiments 3 and 4.

6.2 Results and discussion

Outlier reaction times (slower than 2500 ms from stimulus offset) were excluded from the analysis (less than 1% of the data). Participants correctly labeled fillers as ‘non-words’ with 83.5% accuracy (87.1% and 80.2% accuracy for two- and three-syllable fillers, respectively). Reaction times reflect only correct responses (or ‘word’ responses in the case of the lure stimuli). As in previous experiments, visual inspection revealed the accuracy data had a positive skew, thus percent correct analyses are based on log-transformed data. Table 8 shows average reaction time and percent ‘word’ responses as a function of frequency and pseudoword type. Table 9 shows d' responses by stimulus type and frequency (with C in parentheses).

Table 8. Percent ‘word’ responses and reaction time for learned pseudowords presented as a function of pseudoword frequency and type for Experiment 5. Note that the novel lure responses are false alarms (incorrectly labeling the form as a learned pseudoword).

Pseudoword type		
Frequency	Learned variant	Novel lure
High	86.8 (283)	50.2 (1342)
Low	45.7 (417)	18.7 (1354)

Table 9. *d'* values for the different pseudoword types and frequencies for Experiment 5. Note that the false alarm rate used for this table was the normal fillers – not the lure stimuli.

Pseudoword type		
Frequency	Learned variant	Novel lure
High	2.83 (−.29)	1.19 (.52)
Low	1.16 (.53)	.09 (1.30)

As the novel lures had never been presented during training and could rightly be considered 'non-words' in contrast to the learned variants, we never directly compared the two stimulus types. Consequently, only stimuli from the same category (learned versus novel) were compared. A paired *t*-test on the learned variants revealed an effect of frequency, *t1* (23) = 8.807, *SD* = .167; *t*2 (30) = 3.471, *MSE* = .134; high-frequency pseudowords were recognized more accurately than low-frequency pseudowords.

A paired *t*-test on false alarms for the novel lures found that high-frequency lures were labeled as words more frequently than the low-frequency lures, *t*1 (23) = 6.675, *SD* = .462; *t*2 (30) = 6.63, *SD* = .498.

A paired *t*-test on reaction time for the learned variants showed that high-frequency variants were recognized faster than low-frequency variants (*t*1 (23) = 3.6, *SD* = 68.3; *t*2 (45) = 5.18, *MSE* = 40.99). High- and low-frequency lures showed no difference in reaction time, $F < 1$.

Two complementary sets of *d'* analyses were conducted to mirror the accuracy analyses. High-frequency pseudowords were more discriminable from fillers than low-frequency pseudowords, *t* (23) = 5.013, *SD* = 1.67. Similarly, the high-frequency lures were more easily discriminated from fillers than the low-frequency lures, *t* (23) = 6.20, *SD* = .87. Analyses based on *C* revealed that high-frequency pseudowords had a more liberal bias than low-frequency pseudowords, *t* (23) = −4.926, *SD* = .167. Similarly, the high-frequency lures showed a more liberal bias than the low-frequency lures, *t* (23) = −7.288, *SD* = .227.

Recognition of never-heard forms is modulated by the frequency of an experienced similar form. This pattern of data is similar to the pattern found in Experiment 3, when both variant forms were experienced during training. This suggests that similarity to a high-frequency form plays a role in recognizing a low-frequency form. However, closer inspection of the data patterns across experiments suggests that experience with the low-frequency form also matters: actually hearing a low-frequency variant (without any experience with a similar-sounding counterpart, Experiment 4) increased accuracy relative to responding to the novel variant (experience only with its high-frequency counterpart, Experiment 5): 78.9% versus 50.2%. A post hoc cross-experiment comparison of the experienced and novel forms confirmed this observation, *F1* (1, 51) = 4.137, *MSE* = 398.7; *F2* (1, 15) = 48.06, *MSE* = 66.2.

The results of Experiment 5 point to two mechanisms for enhanced recognition of the low-frequency variants. One way to recognize low-frequency variants is clearly based on similarity; the high false alarm rates for the novel variant forms point to a role for the high-frequency, similar-sounding counterpart. However, similarity does not account for all of the results: low-frequency variant forms (presented once during training) are recognized better than the variant forms recognized solely via similarity to a high-frequency form. This suggests a role for representation of the low-frequency variant.

7 General discussion

In the introduction, we reviewed two primary approaches for how pronunciation variants are recognized. One approach focuses on the representation of pronunciation variants and assumes that variation is encoded in the lexicon; this approach assumes coding of detailed information in representations that includes details about pronunciation variation. The representation view emphasizes experience with a given variant form as a major factor in recognition. A second approach focuses on inferential, rule-based processes that mediate between the surface form of a pronunciation variant and a lexical representation. The inferential view emphasizes rule-based knowledge about pronunciation variants and their canonical representation. To address the role of representation-based and rule-based accounts of recognition of pronunciation variants, Experiment 1 began by replicating earlier findings showing that processing pronunciation variants depends on variant frequency: more frequent variants are processed faster than less frequent variants. Experiment 1 extended these findings to show that additional experience with a pronunciation variant facilitates later processing and that low-frequency variants benefit more from additional experience than high-frequency variants. In addition, experience with pronunciation variation of a given category provides some benefit in processing a new set of variants of that same type. The transfer from one set of variants to a new set also depended to some extent on variant frequency, as transfer tended to be larger for low than high-frequency variants. Experiment 2 extended these findings to a new task, phoneme identification, and decoupled the task during training (passive listening to words) and testing (phoneme identification). The effects of phoneme identification were largely consistent with Experiment 1: variants repeated from training to test showed a larger variant effect, and low variant-frequency words benefitted more from experience than high variant-frequency words. The transfer effect across items from training to test was more variable, but provided some additional evidence that experience with a variant transfers to other words. Experiments 1 and 2 suggest a role for both representations and rules. The variant-frequency effects in recognition combined with the repetition effects suggest that pronunciation variants are recognized via frequency-weighted representations for the pronunciation variants of a given word. The transfer effects point to more general abstract knowledge about how pronunciation variants are related, which can generalize across forms that fit the requirements of a given pronunciation variant class.

The interpretation of Experiment 1 and 2's findings was somewhat limited by the availability of real word stimuli in English that fit the requirements of manipulating variant frequency. Experiments 3, 4 and 5 used an artificial lexicon to address the limitations of Experiments 1 and 2. Experiment 3 replicated the influence of variant frequency on recognition using learned pseudowords: high variant-frequency forms presented during a training phase were processed more easily than low variant-frequency forms. Striking about the findings from Experiment 3 was that the low-frequency forms were recognized relatively easily given that only a single presentation was provided in training. Experiments 4 and 5 investigated the mechanism underlying recognition of the low-frequency variant and, in particular, the role of the high-frequency variant in processing its low-frequency counterpart. Experiment 4 focused on the role of similarity of the high and low variant-frequency pairs in facilitating recognition of the low-frequency variant: low variant-frequency items learned in the context of a similar-sounding high variant-frequency items were recognized better at test than low variant-frequency items that were learned without a similar-sounding high variant-frequency counterpart. This influence of similarity to a high variant-frequency form suggests that a lexicalized variant is crucial to recognition of an infrequent variant. One implication of the results of Experiment 4 is that there is no role for a representation of the low-frequency variant. Experiment 5 investigated the potential role of a representation for the low-frequency variant; specifically, the experiment asked whether representation of a similar-sounding form contributes to recognition of a completely novel variant. Here, completely novel variant forms were presented at test that were similar to pseudowords presented during training. Recognition of the novel variant forms (false alarms) depended on the frequency of the learned variant: higher false alarm rates were found for novel variants similar to high variant-frequency forms than for low variant-frequency forms. Importantly, however, a single presentation of a variant form during training mattered for later recognition: recognition for low variant-frequency forms learned during training (Experiment 4) was enhanced compared to completely novel forms in which only a similar form was learned during training. This latter finding suggests a role for representation of the low-frequency variant.

The evidence for both representations and rules in processing pronunciation variants supports a hybrid approach for recognition of spoken words. On the one hand, a representation of frequent variants supports recognition of the most commonly encountered form. A representation solution for frequent variants provides a fast and efficient way to recognize pronunciation variants – no ‘recovery’ of an underlying form is required to match a lexical representation. Rather, processes that are used during recognition of spoken words in general support recognizing frequent variants. A representation of infrequent variant forms also plays a role and serves to facilitate recognition even with relatively little experience with that form. These findings suggest that words with alternative variant forms are represented in a frequency-graded fashion in the lexicon and that these representations are malleable based on experience.

Our findings also point to a mechanism for recognizing completely novel variant forms: an established representation for a frequent variant form bootstraps recognition of a novel, but similar-sounding variant. This similarity mechanism for recognizing novel forms is consistent with more general processes in spoken word recognition in which recognition of a word involves activation of similar-sounding competitors (Allopenna, Magnuson, & Tanenhaus, 1998). Similarity may facilitate the recognition process of variant forms and it may also serve to expedite the establishment of a lexical representation for the variant form. Thus, similarity may serve as a bridge between a processing and representational explanation of variant recognition.

The identification of a number of underlying mechanisms for processing variants broadens the theoretical landscape for how variants are learned, encoded and later processed. We suggest that the mechanisms identified here are ends on a continuum of strategies that are used to maximize recognition of words. During the early phase of acquiring a new word or word form, episodic traces may be recorded into the lexicon. These episodic traces will allow for maximum differentiation between similar-sounding words, because all phonemes and syllables of the words will be represented and new instances of the words can be holistically compared against the existing exemplars. As the number of exposures to a word increases, episodic traces may begin to support inferential principles to create general rules that encapsulate and are representative of the episodic traces. Optimizing the ways that lexical representations are stored and accessed through mechanisms that evolve given experience provides a potential account for the variant-frequency-based differences. By using these multiple strategies for lexical representation, the lexicon will maximize its potential for recognizing new words and quickly accessing high-frequency words. The increased false alarm rates for the lure stimuli in Experiment 5 seem suggestive of such a generalization to novel stimuli. This generalization may occur because listeners use the phonemes and syllabic structure surrounding the deletion site as a cue to determine if deletion is a viable option. This possibility places the locus of the transfer effect at the syllabic level and, as such, suggests that listeners track co-occurrence probabilities of phonemes and syllables in a way that permits generalization across new and old words. A second and less local source of information available to listeners is the uniform stress pattern that permits schwa vowel deletion – schwa vowel deletion requires a prosodic pattern that is defined by the relational stress of the syllables. Inferential mechanisms paired with episodic representation may support recognition of low-frequency forms or novel forms that fit the general pattern required by a given variant class.

The view presented here is consistent with other suggestions about the relationship between episodic knowledge and more abstract inferential knowledge. Pitt (2009) offers a general account of why both episodic and inferential assumptions are necessary for the recognition of variant forms. Inferential accounts are useful because they capitalize upon systematicities across phonological environments. These regularities are the basis for a rule that can be applied to all words that exhibit the same form of variation (this could also include a novel word that contains the same phonological environment). However, inference alone is insufficient to explain how novel variant forms are recognized as known words. A novel variant in a novel phonological environment will preclude recognition – generalization only occurs once the licensing environment is known. Thus, exposure to the variant form, and subsequent lexical activation of the intended target, is essential for subsequent recognition and for abstraction of a rule. Pitt proposes that lexical involvement, inference and experience all play a crucial role in the recognition of variant forms. Individually each of these aspects is insufficient to fully explain variant recognition, but together they offer a comprehensive approach to the aspects involved in the recognition of variant word forms.

Another excellent discussion of the need for a hybrid approach to speech perception was cogently laid out by Pierrehumbert (2003, 2006a). In complement to the theories and data presented here, Pierrehumbert outlined the need for hybrid models that include multiple levels of representation and the ability to improve recognition based on statistical experience (2006a). In addition, early lexical entries are proposed to occur as a bottom-up process that is initiated by the acoustic information in the speech signal (Pierrehumbert, 2003). With additional language experience the initial lexical entries are more finely tuned in response to the statistical regularities of the language. These statistical regularities are learned by building on the more basic acoustical knowledge that initially formed the lexicon. More nuanced understanding of the language informs listeners' linguistic knowledge and allows them to have greater flexibility in their recognition of spoken language. Thus, initial episodic representations eventually give way to a more mature lexicon that has statistical sensitivities to the structure of the language.

In addition to Pitt and Pierrehumbert, many speech researchers have recently discussed the merits of a hybrid model of spoken word recognition. Sumner and Samuel (2009) espoused a similar view in their discussion of the processing of dialect variations of Standard American English. Experience strongly influenced recognition of variant forms. Building on previous work (Sumner & Samuel, 2005), the authors suggest that variant recognition may occur using both processing (inferential) and representational (episodic) factors. This hybrid view has also been invoked in discussion surrounding second-language (L2) processing (e.g., Cutler & Weber, 2007; Hazan, 2007). The increasing popularity of hybrid models of lexical access speaks to explanatory and predictive power in the spoken word recognition. As previously discussed, the current results add to the literature that proposes joint episodic and abstract referential mechanisms. The question of how the problem of variability in the speech signal is resolved (at least in terms of phonological variants) is likely not a matter of representations *or* rules. Instead, there is a growing body of evidence that calls for a more nuanced approach that considers how representation and rules work synergistically to increase recognition of variant forms.

Acknowledgements

The authors would like to thank Stanislov Sajin and Mike Harman for their helpful comments on this manuscript. We would also like to thank Jen Hay, Paul Warren, and an anonymous reviewer for their feedback on previous drafts of this manuscript. Portions of this paper were submitted as part of Eleni Pinnow's MA Thesis at Binghamton University.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *28*, 419–439.

Baayen, R.H., Piepenbrock, R., & Rlijn, H. (1993). *The CELEX Lexical Data Base on CD-ROM*.

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

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. *Psychological Bulletin and Review*, *11*, 1084–1089.

Connine, C. M., & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review*, *23*, 235–245.

Connine, C. M., Ranbom, L., & Patterson, D. (2008). On the representation of phonological variant frequency in spoken word recognition. *Perception & Psychophysics*, *70*, 403–411.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language*, *54*, 1–19.

Cutler, A., & Weber, A. (2007). Listening experience and phonetic-to-lexical mapping in L2. In J. Trouvain, & W. J. Barry (Eds.), Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007) (pp. 43–48). Dudweiler, Germany: Pirrot.

Francis, W., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.

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

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

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

Hazan, V. (2007). Second language acquisition and exemplar theory. In J. Trouvain, & W. J. Barry (Eds.), Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007) (pp. 43–48). Dudweiler, Germany: Pirrot.

Johnson, K. (1997) Speech perception without speaker normalization: An exemplar model. In Johnson, & Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: 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.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A users guide* (2nd edition). Mahwah, NJ: Erlbaum.

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

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

Patterson, D., LoCasto, P. C., & Connine, C. M. (2003). A corpus analysis of schwa vowel deletion frequency in American English. *Phonetica*, *60*, 45–68.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning and acquisition of phonology. *Language and Speech*, *46*, 115–154.

Pierrehumbert, J. B. (2006a). The next toolkit. *Journal of Phonetics*, *34*, 516–530.

Pierrehumbert, J. B. (2006b). The statistical basis of an unnatural alternation. In L. Goldstein, D. H. Whalen, & C. Best (Eds.), *Laboratory phonology VIII, varieties of phonological competence* (pp. 81–107). Berlin, Germany: Mouton de Gruyter.

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language*, *61*, 19–36.

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

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General*, *114*, 50–77.

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

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*, 487–501.

Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *Quarterly Journal of Experimental Psychology*, *61*, 361–371.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*, 557–580.

Appendix A

Stimuli used in Experiment 1

Three-syllable high-frequency deletion

Severance
Chocolate
Ivory
Corporate
Camera
Opera
Average
Separate
Family
Broccoli
Catholic
History
Factory
Grocery

Three-syllable low-frequency deletion

Envelope
Surgery
Buffalo
Gasoline
Definite
Dominant
Salary
General
Avalanche

Calorie
Criminal
Cardinal
Mineral
Absolute
Gallery
Avenue

Two-syllable very low deletion frequency

Tureen
Pollute
Balloon
Superb
Secure
Collect
Supply
Select
Police
Parole
Corrupt
Peruse
Saloon
Correct
Supreme
Career
Belong
Believe

Appendix B

Stimuli used in Experiment 2

Two-syllable word–non-word continuum endpoints

Believe/Pelieve
Belong/Pelong
Balloon/Palloon
Police/Bolice
Pollute/Bollute
Parole/Barole

Two-syllable non-word–non-word continuum endpoints

Bliege/Pliege
Blove/Plove
Bloos/Ploos
Plije/Blije
Pluge/Bluge
Prote/Brote

Three-syllable word–non-word continuum endpoints

Criminal/Griminal
Cardinal/Gardinal
Cabinet/Gabinet

Three-syllable non-word–non-word continuum endpoints

Cribdal/Gribdal
Carfnal/Garfnal
Caznit/Gaznit

Appendix C

Stimulus of pseudowords used in Experiments 3–5. (Starred pseudowords were presented as variant/similar pseudowords in Experiment 4; unstarred pseudowords were presented as dissimilar yoked pairs.)

Mitorac*
Bepolif
Tobolit*
Dasopee
Ecoloop*
Pesaleeck*
Caparig*
Bicaren*
Sabarat*
Nesatick
Folirine
Jabilive*
Tikirate*
Detirale*
Lopirize*
Sibolive*
Foporite
Tesopitch*
Roculun
Sheborate
Nagarif
Mipaling
Chicarig*
Jusacan*
Casaloon
Petarib
Teesitig
Babalake
Viterack
Mepeloop
Gasecull*
Potarim