

Cue Restructuring in English U-fronting: The Role of Phonetics and Phonology in Inducing Featural Representations

Ildikó Emese Szabó
New York University

1 Introduction

The representation of contrasts has long fascinated phonologists. Since the earliest day of formal phonology it has been suggested that speakers represent every phonemic distinction in their language as a distinction in the values of at least one feature (e.g. Jakobson et al., 1951; Chomsky & Halle, 1968; Hall, 2001). Experimental research supports the idea that speakers generalize over sound classes and use some of the canonical features for conceptualizing patterns. However, most of this research happens to test the cognitive salience of sound classes that are supported by both phonetic similarity and phonological patterns (phonetically natural and phonologically active classes, respectively; Mielke, 2008). Some examples include studies on [+/- voice] for English obstruents (Phillips et al., 2000; Nielsen, 2011; Cristià & Peperkamp, 2012) and on [+/-continuant] sounds in English and (Saffran & Thiessen, 2003; Cristià & Seidl, 2008). In this paper, I look for evidence of a feature-level change of a class that is phonetically natural, but phonologically inactive: English backness. Probing for the representation of classes missing either phonetic or phonological support can tell us more about how important these types of information are, as well as the necessary and sufficient information for learning a feature.

There is very little evidence for backness to be phonologically active in English. Aside from marginal trigram patterns—e.g. in a CwV string, the vowel cannot be back or be containing a back vocoid (Davis & Hammond, 1995; Jagers, under review)—there appear to be no patterns relying on a phonological [+/-back] distinction. Furthermore, experimental data do not point in the direction of speakers learning patterns on [back] either. In a study Smith (2016) tested the productivity of the noun-verb *food–feed* pattern on attested (/u/–/i/) and unattested vowel pairs (e.g. /æ/–/a/). While speakers demonstrated a preference for /i/ in verbs and /u/ in nouns, the preference did not carry over to unattested pairs. Moreover, in Finley & Badecker’s (2009) artificial grammar learning experiment, English speakers showed partial success in learning backness harmony for either low (/a/ and /æ/) or high and mid vowels (/e, o, i/ and /u/). Participants generalized from low vowels to high and mid vowels, but participants trained on high and mid stimuli did not generalize the pattern to low vowels. While the training data could have been interpreted as either backness or rounding harmony, participants seem to have interpreted it as rounding harmony, avoiding or disprefering reference to a class defined via backness.

In this paper, I show that English in the Northeast region of the United States exhibits u-fronting (Section 2). This phenomenon results in cue restructuring for the /u–i/ contrast—i.e. there has been a change to what cues speakers pay attention to in distinguishing the two sounds—and I investigate whether this cue restructuring happens only for the given contrast, or on a larger, featural level. In order to find out if this change also affects other front lax, back tense pairs, I show results from a perception experiment (Section 3), asking two questions. First, which cues are used to distinguish between /u–i/, and, second, whether those cues are useful for the unchanging /o–ɛ/ distinction.

The case of English u-fronting allows us to address larger issues. As the experiment is testing whether the backness distinction is being redefined phonetically for all front lax, back tense pairs on a featural level, results can provide evidence for the psychological reality of a [+back] feature. Thus, they bear significance

* I would like to thank Gillian Gallagher, Maria Gouskova, Laurel MacKenzie, and Scott Seyfarth as well as participants of the NYU Phonetics and Experimental Phonology Lab and the AMP 2017 for their helpful comments and suggestions on this work.

on the representation of the front-back distinction in English. Moreover, since backness is only marginally used in English phonology—i.e. [+/-back] is not crucial for describing phonotactic patterns and alternations in English—we can isolate the contribution of phonetics and phonology in inducing featural representations through the example of English backness. Therefore, this work can contribute to the topic of distinctive features—whether features are learned based on phonetic information only, irrespective of the phonological activity of the given group.

2 U-fronting

The fronting of /u/ has been documented in a host of dialects and areas of the English-speaking world. The F2 of /u/ is rising, in, among other varieties, British English (Gimson, 1970), in California English (Metcalf, 1972; Hinton et al., 1987), in Māori English and New Zealand English (Maclagan et al., 2009), and in the South of the US (Bailey, 1997). In this section I present results of an acoustic analysis comparing data from the Nationwide Speech Corpus (NSC: Clopper & Pisoni, 2006) and Lang & Davidson (2016), which reveals acoustic changes in the English vowel system in the Northeast US.

2.1 Nationwide Speech Corpus (Clopper & Pisoni, 2006) The first corpus, the NSC, contains recordings obtained in a wordlist reading task of hVd minimal pairs. I analyzed tokens of 8 US English sounds, front vowels /i, ɪ, e, æ/, and back vowels /u, ʊ, o, ɑ/—or their equivalents—in six different regions in the Nationwide Speech Project Corpus in hVd words, such as *head* /hɛd/, *heed* /hi:d/, *hood* /hʊd/. For every region, the corpus contains recordings from 10 speakers (5 males and 5 females), who were in the age range of 18-25 at the time of the recording, thus they represent the same generation. Every speaker read every word out five times, yielding 50 recordings per vowel per region. With the 6 regions (Mid-Atlantic, Midland, New England, North, South, West), the corpus contains a total of 2,400 tokens altogether. 11 of these were mislabeled, these were excluded—the dataset used in this paper consisted of 2389 tokens. I obtained F1, F2 and F3 formant measurements at the 50% time point of the vowel (for duration, see Clopper et al., 2005). The automatic measurements were hand-checked, and the data were normalized with formant-intrinsic Nearey transformation (Nearey, 1978).

Data were analyzed in a mixed-effect logit model using the `lme4` package, predicting the value of [back] for each vowel from its Nearey-transformed F1, F2, and F3 values with random intercepts by subject. For all three heights, the backness of the vowel correlated with F2 ($\beta = -19.63$, $p < 0.01$), and to a lesser extent with F1 ($\beta = -2.97$, $p < 0.01$), but not with F3 ($p = 0.45$). I ran a separate model for high vowels (the ones directly affected by u-fronting in the later corpus). This model corroborates the findings of the model with all three vowel heights: F1 is significant ($\beta = -10.46$, $p = 0.03$), F2 is significant ($\beta = -31.83$, $p < 0.01$), but F3 is not a significant predictor of backness for high vowels ($p = 0.45$). A separate analysis revealed that F2 and F3 are correlated with each other (F3 predicted by F2: $\beta = 3.175516$, $p < 0.01$, F2 predicted by F3: $\beta = 0.1974135$, $p < 0.01$), suggesting that F3 might also be a correlate of backness, but its effect is masked by F2.

Even though F2 and F3 are correlated for all regions in the NSC, F2 and F3 are not equally consistent in distinguishing backness. While F3 values of certain back-front pairs can overlap, Figure 1 illustrates the average F2 values of the examined sounds on a scale.* It shows that F2 is a reliable cue for distinguishing members of all back-front pairs, as we can draw a clear boundary between back and front sounds, based on which a [+/-back] feature can be learned. With respect to the Northeast region, this shows a stage where speakers can acquire a [back] feature based on phonetic information on F2 alone. It is questionable how useful or necessary F1 and F3 cues are in this case.

* Figure 1 is based on all normalized vowel measurements, irrespective of region. Visual proportions in the figure are adjusted.

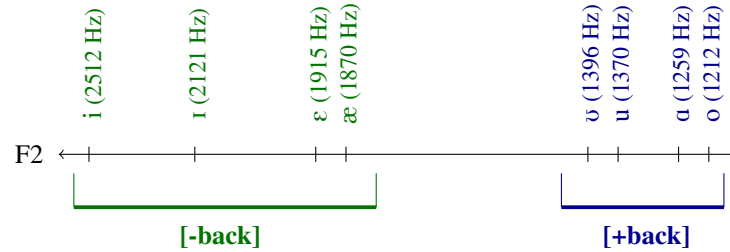


Figure 1: F2 values for [+back] and [-back] in English
Data from NSC (Clopper & Pisoni, 2006)

2.2 Lang & Davidson (2016) The dataset of Lang & Davidson (2016) paints a different picture. The dataset contains vowels segmented from recordings of 3 female and 2 male speakers from the Northeast United States. Monosyllabic target words were embedded in carrier sentences, which allowed for more casual production of speech. The same set of vowels was analyzed in this case too (/i, ɪ, ε, æ, u, ʊ, o, ɑ/).

In contrast to the NSC, the Lang & Davidson (2016) data revealed an overlap in the F2 values of the front-back pair /ɪ–u/. This is a result of simultaneous /u/-fronting (an increase in F2; Labov et al., 2005; Labov, 2008), and /ɪ/-backing (a decrease in F2; Dinkin, 2011). Although /u/ tokens show large dispersion, some /u/ tokens have a higher F2 than any /ɪ/, /ε/, or /æ/ tokens, therefore F2 is not a reliable cue for these vowel pairs. In lieu of F2 cues, the /ɪ–u/ contrast in this dataset is differentiated by F1 cues (as well as in F3 and duration).

At the same time, other front-back pairs still seem to maintain an F2 distinction. This is enough to show a statistically significant effect, between [back] and all 4 predictors in a mixed-effect logit model on all front and all back vowels: F1 ($\beta = -13.09$, $p < 0.01$), F2 ($\beta = -28.21$, $p < 0.01$), F3 ($\beta = -35.96$, $p < 0.01$), and duration ($\beta = 33.49$, $p < 0.01$). Similarly to the NSC results, F2 is a significant predictor of F3 ($\beta = 0.21303$, $p < 0.001$), and F3 is also a significant predictor of F2 ($\beta = 1.5815$, $p < 0.001$) in this data set. This indicates that F2 and F3 are still correlated, and thus neither is a significantly better cue for backness than the other.

Pairwise `lsmeans` comparisons in R indeed show that there is more variation in what cues contrasts use. The data confirm that F2 is no longer contrastive for all front-back pairs: /ε–u/ ($p = 0.40$) and the crucial pair /ɪ–u/ are not distinguished in terms of F2 ($p = 0.73$), therefore F2 is no longer a reliable cue of backness across the whole inventory. The /ɪ–u/ F2-overlap and the lack of phonetic consistency across vowel pairs makes the learning of a [back] feature difficult (Figure 2, visual proportions adjusted) based on F2—in contrast to the easily drawable boundary between [+back] and [-back] in the NSC corpus (Figure 1). Some cues are outright contradictory to a possible boundary: /u/ is fronter than some [-back] vowels (i.e. /ε/ and /æ/).

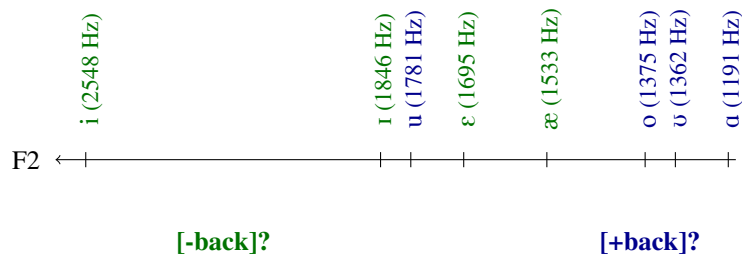


Figure 2: F2 values of [+back] and [-back] in English
Data from Lang & Davidson (2016)

F3 is not a perfectly reliable cue either—not all front-back pairs are distinguished along F3. These pairs are /æ–ɑ/ ($p = 1.00$), /æ–o/ ($p = 0.12$), /æ–ʊ/ ($p = 0.51$), and /ε–ɑ/ ($p = 0.68$). Unfortunately, in this dataset, F1 does not contrast all front-back pairs either: not contrasted pairs are /i–u/ ($p = 0.99$), /ɪ–o/ ($p = 0.84$), and

/ɪ–ʊ/ ($p = 0.3864$). Note, however, that all vowel pairs are contrasted along *some* of these dimensions—e.g. /i–u/ is not contrasted by F1, but it is by F2 and F3, /æ–a/ is not contrasted by F3, but it is by F1 and F2. This indicates that u-fronting in the Northeast resulted in a dialect which while still maintains every previous segmental contrast, but in which backness is cued by a number of contrast-specific cues, instead of a uniform cueing by F2.

2.3 Summary The Nationwide Speech Corpus data (Clopper & Pisoni, 2006) and the data from Lang & Davidson (2016) crucially differ from each other. In the NSC data, F2 was a consistent predictor of backness, thus it is possible to acquire a [back] feature from the acoustic data. In the NSC data, there is a cutoff point in F2 that divides all back vowels from all front vowels. However, the Lang & Davidson (2016) dataset shows an interesting case of near-merger due to u-fronting. The front–back /ɪ–u/ pair is not cued by F2 as reliably as other front–back pairs, but it has a bigger than usual F1 distinction that speakers can use. Instead of relying on a handful of consistent cues to distinguish the group of back vowels from front ones, speakers use a number of contrast-specific cues to distinguish backness (F1, F2, F3, duration).

Acoustic data reveal that /ɪ/ and /u/ differ in F1, F3 and duration. At the same time, /ɛ/ and /o/ maintain a robust F2 difference, but they also exhibit a smaller F1 difference as well as an F3 and duration distinction—i.e. the /ɛ–o/ contrast shares all of the acoustic cues of the /ɪ–u/ contrast (plus F2). Acoustically, the cues for /ɛ–o/ are a superset of those of the /ɪ–u/ contrast. The question is whether native speakers of English make use of the acoustic facts—whether the consistent acoustic distinction of F1 between front lax and back tense vowels truly translates into a perceptual one. We can rephrase this issue on a larger scale. We know that in order to maintain a certain contrast, phonetic cues need to be restructured; the question is whether this happens at the level of individual sound categories or at the level of classes determined by feature values. This question can be addressed by looking at the perception of featurally analogous contrasts, /ɪ–u/ and /ɛ–o/, side by side.

3 The Experiment

In this section I present the design and results from an experiment, which seeks answers to the questions related to whether the cue restructuring following u-fronting is happening on a segmental or a featural level. I look at two questions. First, what cues current Northeast speakers are sensitive to for disambiguating between /u/ and /ɪ/, namely whether F1, F2, F3 and duration are useful.[†] Second, whether the cues useful for /ɪ–u/ are also useful for disambiguating between /o/ and /ɛ/, a contrast which is also distinguished by the features [back] and [lax], but is not directly affected by u-fronting. In other words, we expect F1 to be important for high vowels, the question is whether it is also important for mid vowels, or whether F2 is still the primary cue to the /ɛ–o/ contrast.

3.1 Methods I ran a four-alternative forced choice identification experiment with monolingual native speakers of English on synthesized stimuli. In each trial, subjects heard a single mid or high vowel through noise-cancelling headphones, then they needed to identify the sound as coming from either *lick* (/ɪ/), *leak* (/i/), *look* (/ʊ/), or *Luke* (/u/) for high vowels, and *heck* (/ɛ/), *sake* (/e/), *sock* (/ɒ/), or *soak* (/o/) for mid vowels.

3.1.1 Subjects Subjects were monolingual native speakers of English from the Northeast (from Maine, New Hampshire, Vermont, Massachusetts, Rhode Island or Connecticut) or New York City (not the state). Originally, 37 subjects participated in the experiment, but 10 were disqualified because of substantial knowledge of another language, having lived elsewhere extensively, or other reasons (e.g. falling asleep during the experiment). In the end, this yielded 27 participants. Every participant judged all 256 tokens, which yielded a total of 6912 responses. Further break-down of the stimuli is presented in Section 3.1.2.

[†] Even though F2 trajectory has been linked to backness as well, this is infeasible to execute in this case. Rising F2 contour is associated with front vowels, e.g. /i/, whereas a falling F2 contour is associated with back vowels, e.g. /u/ (cf. Chládková, 2014). However, these results only apply for long (tense) vowels, and lax vowels might be too short to observe (lack of) formant trajectories. Therefore, all vowels in the experiment are monophthongs (steady formants throughout the vowel). Follow-up experiments might be able to expand the set of potential cues.

Condition	F1	F2	F3	Duration
control	✓	✓	✓	✓
in-between F1	✗	✓	✓	✓
in-between F2	✓	✗	✓	✓
in-between F3	✓	✓	✗	✓
in-between duration	✓	✓	✓	✗
in-between F1 & F2	✗	✗	✓	✓
in-between F1 & F3	✗	✓	✗	✓
in-between F1 & duration	✗	✓	✓	✗
in-between F2 & F3	✓	✗	✗	✓
in-between F2 & duration	✓	✗	✓	✗
in-between F3 & duration	✓	✓	✗	✗
in-between F1, F2 & F3	✗	✗	✗	✓
in-between F1, F2 & duration	✗	✗	✓	✗
in-between F1, F3 & duration	✗	✓	✗	✗
in-between F2, F3 & duration	✓	✗	✗	✗
in-between F1, F2, F3 & duration	✗	✗	✗	✗

Table 1: Conditions of stimuli

✓: cue is available (prototypical); ✗: the cue is unavailable (in-between)

3.1.2 Stimuli Stimuli were synthesized based on acoustic measurements from recordings. A phonetically trained native English speaker from the Northeast was instructed to read out a list of nonce words 10 times, containing a 4-member monosyllabic minimal pair as target words (*zeck* /zɛk/, *zoke* /zok/, *zick* /zɪk/, *zuke* /zuk/) as well as 20 distractors. The vowels in the target words (/ɪ u ɛ o/) were extracted and measured for duration, F1, F2 and F3, yielding 10 measurements for each vowel. The outliers—the lowest and the highest 2 value for each vowel for each measurement—were discarded.

A unique value for all 4 measurements (F1, F2, F3, and duration) was generated for every stimulus, based on data from the recordings including some random noise (+/-50 Hz for formants). Each measurement could be prototypical or in-between, resulting in 2^4 (=16) **conditions** of stimuli (Table 1). As the *control* condition had only prototypical cues, it is expected to be the most identifiable. In-between values were the noisy average of a random front lax and a random back tense value of the given height—e.g. for an in-between F1 for /u/ a randomly choosing /i/ F1 and /u/ F1 value was used. Since in-between values meant that the given cue was unavailable, *in-between for F1, F2, F3 & duration* is predicted to be among the least identifiable conditions, as it has the fewest available cues. Tokens for all conditions were generated for all four vowels, yielding 64 **types** of stimuli—e.g. *control* /ɛ/, *control* /u/, and *in-between-F1* /u/ are 3 different types, but *control* /ɛ/ and *control* /u/ are in the same condition (*control*). 4 tokens were synthesized in Praat for each type—i.e. there were 4 in-between-F1, F3 and duration-/u/ tokens, 4 in-between-F3 /ɛ/ tokens, etc, yielding a total of 256 stimuli.

Table 2 shows the range of the final stimuli. For F1, which is inversely correlated with height, and is also correlated with laxness, the range of the in-between high vowel tokens (454.6–515.9 Hz) overlaps with prototypical /ɪ/ tokens (486.3–612 Hz), but not with prototypical /u/ tokens (417.7–425.7 Hz). For mid tokens, the in-between range (645–670 Hz) overlaps with the ranges of both prototypical vowels (/o/: 613–649.8 Hz, and /ɛ/: 666.7–710.2 Hz). Though the difference between front and back vowels is smaller for mid than for high vowels (16.9 Hz and 60.6 Hz, respectively), both heights have a significant difference in F1 between front and back vowels. Back tense vowels (/o/ and /u/) are articulated higher—i.e. have a lower F1—than their front lax counterparts (/ɛ/ and /ɪ/). F2, a direct correlate of backness, does not pattern the same way for high and mid vowels. For mid vowels, ranges of front, back, and F2-in-between vowels are distinct (1822.4–1907.5, 1725.0–1802.0, and 1618.9–1698.4 Hz). While extreme F2 values of high vowels also show a similarly scalar pattern, there is a much larger overlap between front and back values. The ranges of F3, which is a correlate of lip rounding, differ between front, in-between-F3, and back vowels for both vowel heights to a similar extent. There was a large overlap between lax and tense vowels in duration. Had these values been used for the generation of stimuli, this would not have allowed us to test the usefulness of duration as a cue between front lax and back tense vowels.[‡] In order to ensure a difference between the

[‡] The target words in the recordings featured the vowels in pre-voiceless environment (*ɾ.k*), thus pre-fortis clipping might be responsible for this merger.

	High vowels	Mid vowels
F1 values	/ɪ—u/	/ɛ—o/
front lax	486.3–612.0 Hz	666.7–710.2 Hz
in-between	454.6–515.9 Hz	645.0–670.0 Hz
back tense	417.7–425.7 Hz	613.0–649.8 Hz
F2 values		
front lax	1881.7–1981.6 Hz	1822.4–1907.5 Hz
in-between	1928.9–1989.5 Hz	1725.0–1802.0 Hz
back tense	1912.6–2004.8 Hz	1618.9–1698.4 Hz
F3 values		
front lax	2773.8–2874.0 Hz	2731.3–2795.4 Hz
in-between	2694.8–2772.6 Hz	2669.2–2711.9 Hz
back tense	2594.2–2671.2 Hz	2540.6–2664.9 Hz
Duration values		
front lax	91.31–126.50ms	
in-between	140.53–166.50ms	
back tense	175.66–208.09ms	

Table 2: F1, F2, F3 and duration range of stimuli, high (/ɪ/ and /u/) and mid vowel (/ɛ/ and /o/) separated

duration of tense and lax vowels in prototypical cases, duration was manually adjusted to a 0.71 lax-to-tense ratio (House, 1961; Giegerich, 1992)—to the ranges seen in Table 2. Even so, there were some remarks by participants on the sounds overall being very short, which I will discuss more in Section 3.2.

Statistical analysis of the stimuli reveals that back and front vowels were significantly different from each other along all of these dimensions, albeit the magnitude of this difference varied across the two heights—e.g. the F1 difference between front and back was bigger for high vowels than for mid vowels. However, looking at Table 2 shows that these significant differences are found between even massively overlapping values—e.g. the F2 of front vs. back high vowels.

3.1.3 Procedure In the experiment, subjects heard a sound file of a vowel in isolation, which they were asked to identify in a forced-choice experiment by answering the question ‘Which word do you think this sound was cut out from?’. Two extra options were added for both vowel height: on top of front lax (e.g. [ɪ]) and back tense (e.g. [u]), participants were able to choose front tense (e.g. [i]) and back tense vowels (e.g. [ʊ]) as a response, resulting in a 4-alternative forced-choice design. If subjects heard a high vowel, they could choose from the words *lick* (/ɪ/), *leak* (/i/), *look* (/ʊ/), and *Luke* (/u/). For mid vowel stimuli, the choices were *heck* (/ɛ/), *sake* (/e/), *sock* (a), and *soak* (/o/). Stimuli were synthesized for only 4 of them (2 high vowels: /ɪ/ and /u/; 2 mid vowels: /ɛ/ and /o/). The vowels [ʊ], [i], [a], and [e] were offered as extra options in the identification task, but were not used for token generation. Subjects had the chance to replay each token once more. They were told that this is a sample from a larger pool of tokens, thus they should not try to pick every option the same number of times.

Responses were encoded in three ways: sound accuracy represented raw data, while backness accuracy and laxness accuracy were included to test the importance of cues for backness and laxness separately. Sound accuracy is a binary variable encoding whether the subject identified the sound correctly. Each token (even in-between ones) has a “correct answer”, irrespective of the ambiguity of the cues, e.g. the correct answer *in-between-F1,F2,F3,duration* u token is “*Luke*”, even if there is a nearly identical, ɪ token with merged F1, F2, F3, duration whose correct answer is “*lick*”.[§] Backness and laxness accuracy are forms of “partial credit”. Backness accuracy encodes whether the subject identified backness. For example, for an /u/ token, the responses “*Luke*” and “*look*” are accurate for backness, even though a “*look*” response is incorrect for laxness (because /u/ is tense, not lax). Laxness accuracy measures whether the subject was right along the tense-lax dimension. In case of an /u/ token, if the subject chose the option “*Luke*” or “*leak*”, then they successfully identified the sound as tense—while potentially making a mistake in backness.

The experiment began with a practice phase, where subjects heard 2 randomly chosen control tokens for

[§] Participants were expected to perform around chance in the *in-between-F1,F2,F3,duration* condition, since there was little information in the token about its identity, and subjects should guess at random.

each vowel (8 sounds altogether). The responses recorded during the practice phase were not included in the results below—the purpose was to familiarize subjects with both the task and synthesized stimuli. In the test phase of the experiment, subjects had the same task. Each subject had to judge each stimulus token exactly once. Tokens were presented in a blocked design—subjects either heard all of the high tokens first, or all of the mid tokens first. Tokens within the same block were randomized for each subject.

3.2 Results The following data represents the by-factor means of accuracy over all participants. In general, people performed poorly on sound, backness, and laxness accuracy, which raise some questions that will be discussed in Section 4, but there are still significant differences in accuracy based on condition.

For sound accuracy, participants performed between 22.22% and 43.52% for high vowels and 25.00% and 45.37% for mid vowels. Remember, this was correctness out of 4 choices, therefore chance is at 25%. As for laxness accuracy, by-condition performance was in the 45.37%–64.81% range for high vowels, and a 48.61%–62.50% range for mid vowels. Backness accuracy result were similar with performance between 48.61% and 70/83%, and between 49.54% and 69.44% for high and mid vowels, respectively. For laxness and backness accuracy, chance was at 50%. While the across-condition averages of mid and high vowels are comparable, the statistical patterns they show are different. In the following, I give a detailed account of by-condition results for high and mid vowels separately, focusing on laxness and backness accuracy data.

The data were analyzed with two binomial mixed-effects logarithmic regression models in R, using the package `lme4` 1.1–12. The two models differed in the dependent variable—the first model used *backness accuracy* data, and the second model used *laxness accuracy*. Both models had *F1*, *F2*, *F3*, *duration*, *vowel height* and *if-first* (whether the item was in the first block or not) as predictors—with all interactions between *F1*, *F2*, *F3*, *duration* and *vowel height*. Random intercepts by *participant* and *item* were fitted.

3.2.1 Laxness Accuracy A general model was run with laxness accuracy as the dependent variable as well. Effects with p -values below $p < 0.05$ are shown in Table 3, but the Bonferroni-corrected threshold for significance is at $p < 0.0016$ again ($0.05/32$ for the 32 tests in the model). Under the corrected threshold only two effects are significant: *duration* and *if-first*.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.2295	0.0603	3.8061	0.0001
F1	-0.1729	0.0601	-2.8766	0.0040
duration	-0.2654	0.0601	-4.4153	<0.0001
if-first	0.2237	0.0499	4.4837	<0.0001

Table 3: Significant predictors of tense–lax accuracy
(e.g. /i, u/ vs. /ɪ, ʊ/)

Participants performed worse in identifying laxness without duration cues in the data ($\beta = -0.26$, $p < 0.0001$). While subjects performed better on items in the first block ($\beta = 0.22$, $p < 0.0001$), this effect probably has to do with the extreme shortness of the stimuli. Since duration is an important cue of laxness, when subjects were exposed to short stimuli for an extended period of time, it is possible that the durations started to skew their responses in the lax direction, producing a fatigue effect. The effect of *F1* comes close, but it does not reach significance under the corrected significance threshold ($p = 0.004$ is above the $p < 0.0016$ threshold, thus not significant).

No interactions are significant which suggests that the two heights of vowels (high and mid) behave similarly. Inspecting raw data visually, the difference between prototypical and in-between *duration* seems present for both vowel heights. This is illustrated visually in Figure 3.

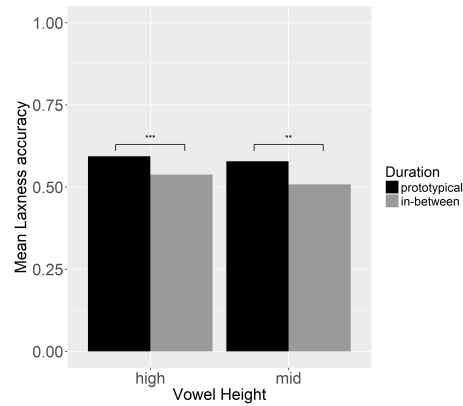


Figure 3: Mean laxness accuracy by height and ambiguity of duration (left) and F1 (right)

3.2.2 Backness Accuracy The first model was fitted on backness accuracy data, Table 4 shows the significant effects under $\alpha=0.05$ —other effects and interactions are not significant. With a 0.05 p-value threshold 4 main effects (*F1*, *F2*, *F3*, *duration*), and 2 interactions (*F1:vowel height*, *F2:vowel height*) are significant. However, since together with the interactions 32 tests were run in the model, the p-value threshold of significance needs to be adjusted with Bonferroni correction to a threshold of $0.05/32=0.0015625$. This means that under the corrected threshold, two effects are no longer significant: *F3* ($p=0.0131$) and *duration* ($p=0.0031$). These effects were likely a result of running many (32) tests.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3990	0.0639	6.2419	<0.0001
F1	-0.3110	0.0527	-5.8969	<0.0001
F2	-0.2613	0.0527	-4.9556	<0.0001
F3	0.1308	0.0527	2.4814	0.0131
duration	-0.1558	0.0527	-2.9552	0.0031
vowel height:F1	0.6227	0.1055	5.9034	<0.0001
vowel height:F2	-0.4818	0.1055	-4.5689	<0.0001

Table 4: Significant predictors of back–front accuracy
(e.g. /u,ʊ/ vs. /i, ɪ/)

The effects of *F1*, *F2*, as well as their interactions with vowel height remained significant with the adjusted threshold. There is an interaction of vowel height and F1 suggesting that F1 is not equally used as a cue for mid and high vowels. Table 4 indicates that the significant main negative effect of F1 ($\beta = -0.31$, $p < 0.0001$) only holds for high vowels—it is canceled for mid vowels ($\beta = 0.62$, $p < 0.0001$). Post-hoc analysis (of running the same models on high and mid vowels separately) corroborated that F1 ambiguity only worsened performance for high vowels (cf. high vowels: $\beta = -0.63$, $p < 0.0001$; mid vowels: $p = 0.9967$). This is plotted in Figure 4.

F2 ambiguity also shows an interaction with vowel height ($\beta = -0.48$, $p < 0.0001$). There is also a main effect of F2 ($\beta = -0.26$, $p < 0.0001$). Compared to the baseline of high vowels, mid vowels are more poorly identified under in-between F2 values. Post-hoc analysis (Bonferroni-corrected threshold of significance: $p < 0.003$) showed that, F2 is significant for mid vowels, but not for high ones (cf. high vowels: $p = 0.81$; mid vowels: $\beta = -0.51$, $p < 0.0001$). While performance on high vowels is not affected by an in-between F2, performance on mid vowels worsens without an F2 cue.

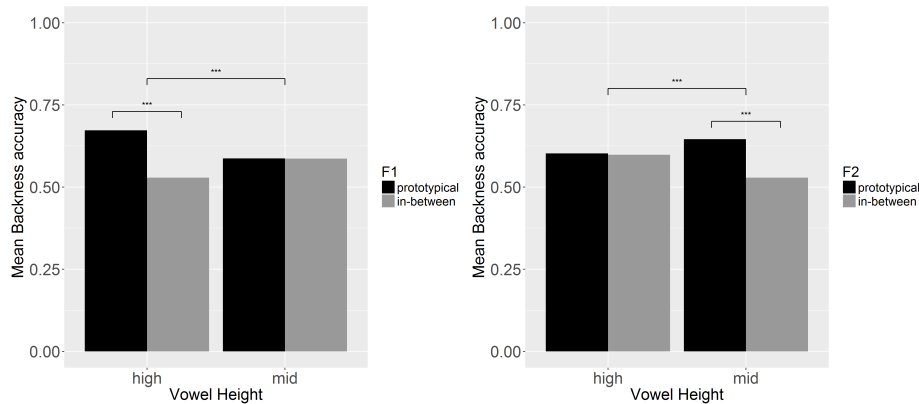


Figure 4: Mean backness accuracy by height and ambiguity of F1 (left) and F2 (right)

3.3 Summary of Results In this section, I presented results from a forced-choice identification experiment, looking at what cues native English speakers use to distinguish between front lax and back tense vowels. I tested two questions: what cues (out of F1, F2, F3, and duration) speakers are sensitive to for the distinction between /u/ and /ɪ/, and whether speakers are sensitive to the same cues for the /o–ɛ/ contrast as for the /u–ɪ/ contrast. While the overall and the baseline accuracy (in the *control* condition) was low, indicating that the task was hard, participants were clearly able to distinguish between /u/ and /ɪ/, as well as between /o/ and /ɛ/ and there were also significant differences in accuracy between conditions.

On the one hand, no difference between high and mid vowels is found in terms of laxness. Short duration is a very strong cue of laxness in English, which is confirmed in the results: neutralizing durational cues worsens the participants’ performance in the identification task, but no formant cues are demonstrably important in the laxness model. These results are not affected by vowel height—i.e. laxness is cued uniformly for vowels of both height.

On the other hand, backness is not cued so uniformly. For high vowels, speakers associate a lower F1 with backness and a higher F1 with frontness, but no such correlation has been found for F2, F3, and duration. However, F1 is not a useful cue of backness for mid vowels, even though there is a consistent F1 difference between /ɛ/ and /o/ tokens in the stimuli, albeit smaller than for high vowels (Table 2). At the same time, /ɛ/ and /o/ were distinguished by F2—but prototypical F2 is only used for mid vowels. However, it is important to note “prototypical” here only means that the token is maximally informative compared to vowels in the recordings. As Table 2 shows, the F2 values of high vowels heavily overlapped in order to best reflect a variety of English in which u-fronting is so drastic that high vowels /u/ and /ɪ/ are not distinguished by F2. Therefore, the lack of usefulness for F2 for high vowels is a result of subjects not being provided with sufficient F2 information to begin with—and therefore they could not make use of it.

To sum up, while laxness was encoded by the same cues for mid and high vowels (duration), participants used different cues to identify mid and high vowels (F1 for high vowels, F2 for mid vowels). No single cue was found to correlate with the English backness distinction on a class level, suggesting that this distinction might not be represented in the grammar in a featural form. Instead, contrasting vowel pairs might be cued at the segmental level.

4 Discussion

In this paper, I looked at a phonetically natural, but phonologically inactive feature ([back] in English) undergoing change. I demonstrated that u-fronting also affects the Northeast US, thereby diminishing the F2 difference between /ɪ/ and /u/, resulting in a need for cue restructuring. I asked the question whether this cue restructuring is happening just at the level of this one segmental contrast, or whether it is extended to the classes of back vs. front sounds. The experiment presented here addresses the central issue in two ways: by looking at English backness in itself, and by comparing the phonetic implementation of backness and laxness in English. Results show that the back tense–front lax contrast is cued differently for high vowels (/u–ɪ/ is cued by F1) and for mid vowels (/o–ɛ/ is cued by F2). In the meantime, laxness is cued uniformly (by duration).

The main conclusion we can draw based on these results is that in this case cue restructuring happens on a segmental level. I could not find evidence for any pressure for class-level cue restructuring: the change in the cuing of the /u–ɪ/ contrast did not result in change in the cuing of the putatively featurally analogous pair /o–ɛ/. In lieu of F2 cues, participants used F1 cues to disambiguate between high vowels. While there was a consistent, albeit smaller, F1 cue available for mid vowels as well, this was not used to determine the backness of mid vowels (only F2 was used for this purpose).

The fact that F1 ambiguity has no effect on identification accuracy for mid vowels can happen for two reasons: the problem is either the magnitude of the cue—i.e. the 15Hz difference in F1 is too small to perceive—or the type of the cue—i.e. F1 is not a dimension speakers associate with the /o–ɛ/ contrast like they do with the /u–ɪ/. While it is true that the 15Hz is both small, and it is harder to perceive in the 600–700 Hz range than the 400–600 Hz range, we can choose between the two reasons by testing whether the perceptual equivalent of a 15 Hz difference in the 600–700 Hz range can be perceived in the 400–600 Hz range—for the /u–ɪ/ contrast. If that is the case, then the observed results must be due to F1 being a “wrong” cue for backness for the /o–ɛ/. Either way, it is true that F1 is not acoustically enhanced for mid vowels, and the lack of importance of F1 for mid vowels can be interpreted as an instance of a putative ([back]) feature being implemented with different cues depending on the contrast.

I have thus found evidence of a at least a segmental representation of contrasts. Speakers are able to distinguish between /u/ and /ɪ/ as well as between /o/ and /ɛ/ using whatever cues are available in the given case (F1 for high vowels, and F2 for mid vowels). The cuing of the two contrasts is not necessarily related in this case: while speakers have detailed phonetic representations of sounds—they are aware of the differences along multiple dimensions within each pair—the contrast is established between individual sounds. With such a representation, no meta-observations about contrasts are possible—for example, the speakers do not represent the /i–u/ contrast as being more similar to the /e–o/ contrast than the /i–ɪ/ contrast.

An alternative is predicted by theories of distinctive features in which distinctions are represented on a featural level—just like voicing is for English obstruents, as evident from Nielsen (2011); Cristià & Peperkamp (2012), et alia. A [+/-back] feature would mean that speakers realize the parallelisms between how the difference between front and back vowels is cued in English, and thus the cuing of such distinction is defined on a featural level—i.e. an /u–ɪ/ contrast is defined as a distinction defined between [+back] vs. [-back] sounds. Under a featural representation, equivalence classes are also accessible, which allows speakers to state similarity relations over contrasts—e.g. that the difference between /u/ and /ɪ/ is phonologically equivalent to the difference between /o/ and /ɛ/. Results from this experiment could have supported a [+/-back] feature if the cue restructuring happened at a featural level, and the two contrasts were cued similarly—i.e. if speakers found /u–ɪ/ cues for backness useful for the /o–ɛ/ contrast as well. However, this experiment could not provide evidence for a featural backness distinction in English.

Comparing backness and laxness casts further doubt on a featural representation of backness. While the cues important for backness differed between /u–ɪ/ and /o–ɛ/, laxness was uniformly cued by duration. This aligns with the phonological (in)activity of the given features—remember [back] is a phonologically inactive feature, [lax] is phonologically active. The latter category has time and time again been demonstrated to correspond to features (mostly because of tense and lax vowels patterning differently word-finally), whereas there is no unambiguous evidence for class behavior for back vs. front classes of vowels. The fact that there is an example of segment-level cue restructuring for backness, but laxness is cued uniformly, points in the direction that they might be represented differently—phonologically inactive distinctions might not correspond to features. This suggests that similarity of phonetic cuing in and of itself might not be enough for inducing features, featural representations might only be learned when the distinction is also supported by different phonological patterning of the two sound classes. Such results cannot corroborate distinctive feature theory, which predicts that (at least some) phonologically inactive classes are represented by features, and thus that (in at least some cases) phonetic information—i.e. similarity—must be sufficient for inducing features.

It is also possible that speakers do acquire a feature [back], it just cannot be detected in the form of feature-level cue restructuring. In this experiment, I tested one way in which class behavior can be attested (the cuing of phonetically natural, phonologically inactive classes under the pressure of a change in progress). This experiment started to tease apart the role of phonetic information in feature learning, but not only is there a need for more experimental work on phonetically natural, phonologically inactive classes in order to learn about the relevance of phonetic information, but also on phonologically active, phonetically unnatural classes

to learn about how much phonological information matters in feature learning. We have little empirical data about what type of information speakers focus on when establishing representations, and how strong evidence they need to represent a set of sounds with a feature, and there is more work to be done on the contribution of different types of information in feature learning. Results like these highlight the need for more experimental and computational work in order to fully understand and isolate the contribution of phonetic and phonological information in feature learning, as well as to develop a comprehensive model of feature-learning.

References

- Bailey, Guy (1997). When did Southern American English begin? *Englishes around the World* 1, 255–275.
- Chládková, Kateřina (2014). *Finding phonological features in perception*.
- Chomsky, Noam & Morris Halle (1968). *The Sound Pattern of English*. Harper & Row Publishers, New York.
- Clopper, Cynthia G & David B Pisoni (2006). The Nationwide Speech Project: A new corpus of American English dialects. *Speech Communication* 48:6, 633–644.
- Clopper, Cynthia G, David B Pisoni & Kenneth De Jong (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America* 118:3, 1661–1676.
- Cristià, Alejandrina & Sharon Peperkamp (2012). Generalizing without encoding specifics: Infants infer phonotactic patterns on sound classes. *Proceedings of the 36th Annual Boston University Conference on Language Development (BUCLD 36)*, 126–138.
- Cristià, Alejandrina & Amanda Seidl (2008). Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development* 4:3, 203–227.
- Davis, Stuart & Michael Hammond (1995). On the status of onglides in American English. *Phonology* 12:2, 159–182.
- Dinkin, Aaron J (2011). Weakening resistance: Progress toward the low back merger in New York State. *Language Variation and Change* 23:03, 315–345.
- Finley, Sara & William Badecker (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61:3, 423–437.
- Giegerich, Heinz J (1992). *English phonology: An introduction*. Cambridge University Press.
- Gimson, Alfred Charles (1970). *An introduction to the pronunciation of English*. Hodder Education.
- Hall, Tracy Alan (2001). *Introduction: Phonological representations and phonetic implementation of distinctive features*, 1–40.
- Hinton, Leanne, Birch Moonwomon, Sue Bremner, Herb Luthin, Mary Van Clay, Jean Lerner & Hazel Corcoran (1987). It's not just the Valley Girls: A study of California English. *Annual Meeting of the Berkeley Linguistics Society*, vol. 13, 117–128.
- House, Arthur S (1961). On vowel duration in English. *The Journal of the Acoustical Society of America* 33:9, 1174–1178.
- Jagers, Zachary (under review). Evidence and characterization of a glide-vowel distinction in American English. Submitted to *Laboratory Phonology*.
- Jakobson, Roman, C Gunnar Fant & Morris Halle (1951). *Preliminaries to Speech Analysis: The distinctive features and their correlates*. Acoustics Laboratory MIT.
- Labov, William (2008). Triggering events. Fitzmaurice, Susan & Donka Minkova (eds.), *Empirical and Analytical Advances in the Study of English Language Change*, Berlin, New York: Mouton de Gruyter, 11–54.
- Labov, William, Sharon Ash & Charles Boberg (2005). *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.
- Lang, Benjamin & Lisa Davidson (2016). Phonetic drift in crowded vowel spaces: Effects of exposure in English learners of French. Poster presented at the 90th Annual Meeting of the Linguistic Society of America, Washington DC, January 7-10.
- Maclagan, Margaret, Catherine I Watson, Ray Harlow, Jeanette King & Peter Keegan (2009). /u/ fronting and /t/ aspiration in Māori and New Zealand English. *Language Variation and Change* 21:2, 175–192.
- Metcalfe, Allan A (1972). Directions of change in southern California English. *Journal of English Linguistics* 6:1, 28–33.
- Mielke, Jeff (2008). *The Emergence of Distinctive Features*. Oxford University Press.
- Nearey, Terrance Michael (1978). *Phonetic feature systems for vowels*, vol. 77. Indiana University Linguistics Club.
- Nielsen, Kuniko (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39:2, 132–142.
- Phillips, Colin, Thomas Pellathy & Alec Marantz (2000). Phonological Feature Representations in Auditory Cortex. *Unpublished manuscript*.
- Saffran, Jenny R & Erik D Thiessen (2003). Pattern Induction by Infant Language Learners. *Developmental Psychology* 39:3, p. 484.
- Smith, Jennifer L (2016). Segmental noun/verb phonotactic differences are productive too. *Proceedings of the Linguistic Society of America* 1:16, 1–15.