Schwa has an articulatory target in American English

Uriel Cohen Priva, uriel_cohen_priva@brown.edu
Emily Strand, emily_strand@brown.edu

Abstract

Is American English schwa's position determined solely by coarticulatory pressures? There is currently disagreement between articulation-based and acoustic-based studies (Browman & Goldstein, 1994; Flemming, 2009). In two acoustic corpus studies using the Switchboard and Buckeye corpora, we find that vowels head toward a central high vowel position when subject to increased coarticulatory pressures, rather than toward the position occupied by American English schwa. Even lexical schwa vowels shift to lower F1 values when their duration is relatively short. Our findings are consistent with schwa occupying a perceptually but not articulatorily neutral position. As such, they bear on vowel neutralization patterns, and suggest that unstressed syllables may convey specific information: that they are not stressed.

1 Introduction

1.1 Overview

Phonological theories tend to regard schwa as the "neutral position" (e.g. Harris, 2005; Flemming, 2009). Given this, schwa is defined by the properties it lacks (i.e. it is featureless, Oostendorp, 1995): It is not front, high, low, back, or round. In phonetic terms, however, it is not completely clear what properties or lack thereof should define schwa. One alternative is that schwa is *articulatorily* neutral and serves no role except being the in-between vowel (Oostendorp, 2002). In this case, schwa's specific articulatory properties would be fully determined by the context in which it occurs (Kondo, 1994), or the most convenient place between every combination of preceding and following contexts. Another alternative is that schwa is *perceptually* neutral: It does not occupy the acoustic space of other vowels that are specified as having particular perceptual targets (as predicted by Dispersion Theory, Flemming, 2004). Given that such differences may require the approaching of particular acoustic targets, a perceptually-neutral schwa may behave as if it does have an articulatory target (Hall, 2003). In this case, rather than merely serving as the best articulatory path between preceding and following contexts, schwa would do more, or serve additional functions, namely signaling that it is not some other vowel or that the syllable in which it occurs is not stressed (along the lines of an argument made for consonant lenition in Katz, 2016; Katz & Fricke, 2018). Therefore, it is necessary to determine (a) where the articulatorily-neutral position in American English is (in acoustic terms), and (b) whether that position is similar to that of American English schwa.

The question seems easy enough to answer. From an articulatory standpoint, we may be able to determine whether C_1-schwa-C_2 transitions involve gestures not dictated by C_1 or C_2. However, articulatory studies (Browman & Goldstein, 1994) and acoustic studies (Flemming, 2009) that have taken this approach arrived at conflicting conclusions. Here, we use a different approach to investigate the articulatorily-neutral position and its

relation to schwa in American English. We use significantly more data than previous studies, and we rely on the observation made by Lindblom (1963): shorter articulations correspond to increased coarticulation pressures. Thus, with respect to vowels, those produced in fast speech or those that take less time to produce would reflect stronger coarticulation pressures. Such variation would help identify the coarticulation pressures that vowels are subject to, and whether that resulting position is where we also find schwa.

1.2   Background

Browman & Goldstein (1994) provide evidence supporting the position that schwa has an actual articulatory target in American English. They find that the articulation of American English schwa requires a lower tongue position than coarticulation alone would predict (the neutral position). Although Flemming (2009) similarly finds that schwa has higher F1 values than the neutral position, he argues against schwa having an articulatory target. Flemming argues that schwa results from assimilation to context, which he interprets as a lack of mid-central quality. His analysis argues for two schwa vowels in American English: (1) a true mid vowel that arises from the coarticulation between low vowels and neighboring consonants, e.g. the final vowel in *Rosa's*, and (2) a more variable vowel that results from more extreme reduction, e.g. the final vowel in *Roses* (in Section 1 and in more detail in Flemming & Johnson, 2007). However, despite the difference, he stresses that neither has an actual articulatory target:

> "both result from assimilation to context, so there is no support for the notion of vowel reduction as approximation to a mid central quality. The different outcomes represent different degrees of assimilation to context."

Flemming (2009, Section 3) provides several possible reasons for why schwa would appear to warrant a lower tongue position and higher-than-expected F1 values. First, the data in Browman & Goldstein (1994) may have been interpreted as word-final, where coarticulation pressures are not as strong and underlying vowel quality may have been lower (e.g. the famous *Rosa's roses* minimal-pair contrast). Second, he proposes that assimilation to consonants may not always involve higher tongue position (and lowered F1 values) and that the observed higher F1 values could have resulted from coarticulation pressures rather than schwa having an articulatory target. However, he acknowledges that with the absence of more evidence, it would be difficult to positively conclude that schwa lacks an articulatory target.

Davidson (2006), who builds on Davidson (2005) and Davidson & Stone (2004), provides some support for the idea that schwa does more than just serve as an in-between vowel. Native speakers of American English were required to pronounce words with onset clusters that do not exist in American English. One strategy they used was to insert a minimal vowel to break the clusters, but that vowel seemed to have a different articulatory target than underlying schwas, suggesting that schwa does more than serve as an in-between vowel. One caveat is that it is possible for languages to use epenthetic vowels that do have specific articulatory target, so these findings could be interpreted as evidence only to schwa not being the epenthetic vowel in American English.

Although a fair amount of research has been conducted on schwa in terms of vowel reduction, none of it has

incorporated large amounts of corpus data, which seems ideal for resolving this open question. In large-enough corpora, every vowel occurs in many different contexts rather than in a limited set of carefully constructed contexts. In this paper we use two large corpora: (1) the Switchboard corpus (Godfrey & Holliman, 1997; Harkins, Feinstein, Lindsey, Martin, & Winter, 2003), which provides word-level annotations from hundreds of speakers all over the United States who participated in thousands of telephone conversations, and (2) the smaller Buckeye corpus (Pitt et al., 2007), which provides segment-level annotations and higher-quality recordings of 40 speakers interviewed in Columbus, Ohio.

An important advantage of corpora is the variability of language production across different individuals, times, and contexts. Since our focus is on casual spontaneous speech, such a rich source gives way to the analysis of reduced-duration vowels (Bergem & Beinum, 1989; Turnbull, 2019), which in terms of coarticulation, Lindblom (1963) and Moon & Lindblom (1994) argue plays a pivotal role. This is evident from the research done by Sproat & Fujimura (1993), such that reduced-duration /ɫ/ is less velarized in general given that speakers, due to coarticulation pressures, aren't able to retract and lower their tongues fast enough to produce the more velarized variant. With respect to vowels, this means that as vowels shorten, they are expected to be subject to the same coarticulation pressures that targetless schwa would be subject to. For the question at hand, increased coarticulation at reduced duration makes two related predictions. First, the average position of schwa in acoustic space should not change as it gets shorter: If schwa is indeed targetless, vowels that are lexically specified as schwa should not shift in the F1/F2 space as their duration gets shorter. Alternatively, if schwa does have a specified higher F1 value/lower tongue position, then we would expect it to have a higher F1 value when its duration is longer and a lower F1 value when its duration is shorter. Second, other vowels should shift toward schwa as their duration gets shorter (Fourakis, 1991): Reduced-duration low and mid vowels should have lower F1 values, and F1 values of high vowels should either undergo no change or drop to the height of the lax high vowels. Crucially, if schwa is indeed targetless, lax high vowels should approximate to schwa as their duration gets shorter, wherever that position is. Regardless of whether schwa does or does not have an articulatory target, all vowels are expected to center (have more schwa-like F2 values) as they become shorter (Kaźmierski, 2015), which could serve as the basis for evaluating the predictive power of the model.

2 Study 1

2.1 Introduction

The goal of this study is to identify the trajectory of vowels in acoustic space as speech rate increases, where coarticulatory pressures are expected to grow, in order to identify the acoustic location that corresponds to the average articulatorily neutral position. In other words, if vowels are expected to reflect stronger coarticulatory pressures in faster speech rate, where are they heading? Our expectation is that they would center in F2 terms (Mooshammer & Geng, 2008; Stevens & House, 1963), and assume F1 values that resemble schwa (if schwa is in the articulatorily neutral position) or high vowels (due to the proximity of consonant articulation) (Kondo, 1994). In terms of testing this, our focus is on the mean F1/F2 values, not on actual values. Our logic is that if

schwa has a target that is driven by coarticulatory pressures, it would reflect the sum of the pressures that it is subject to. We therefore try to remove the influence that individual contexts have on each vowel, and observe the mean location instead. Our collection method, described below, makes it difficult to collect information on schwa itself, so we focus on vowels that have F1 positions close to that of schwa. We ask whether /ɪ/ and /ʊ/ move to lower or higher F1 values when speech rate increases.

2.2 Methods and materials

2.2.1 The Switchboard Corpus

The Switchboard Corpus (Godfrey & Holliman, 1997) is a large collection of telephone conversations, containing associated data for speaker identification. Speakers were randomly paired with one another and given topics to converse about. In total, the corpus has over 500 speakers engaged in over 2000 conversations. Each side of the conversation is a distinct recording, so speakers' measurements can be reliably obtained. Word boundaries were based on manually corrected word annotations produced at MS State (Harkins et al., 2003), which made it possible to extract individual words. However, segment-level annotations were not available for the Switcboard corpus. Therefore, we employed the following procedure: We extracted the sound files of all monosyllabic words that contained only a monophthong and obstruents. We used Praat to extract F0 values (`To Pitch: 0.001, 75, 300`) and formant values (`To Formant (burg)...  0.01 3 3500 0.025 50`). The maximal frequency (3500) is due to the low sampling frequency in telephone communication. We excluded formant values whose timestamp was not within 10ms of a pitch measurement. We then classified the modes of F1 and F2 for each vowel as the actual values rather than specifying the midpoint of each vowel, given that such information was not available. To verify the validity of the formant mode approach, we employed it in the Buckeye corpus (Pitt et al., 2007), which does provide segment-level timestamps (see Study 2 section 3.2 below). We found that the F1/F2 modes are at Pearson r=.89 and .91 correlation with the F1/F2 values for the midpoint of each vowel, respectively. Underlying representations were assumed to be those provided by the CMU dictionary (Weide, 2008). The resulting number of tokens for each vowel as well as the three most frequent words are in Table 1. The formant values of all the words we applies this procedure to are available at https://github.com/ucpresearch/moredata (note that these include vowels we excluded based on the listed criteria).

2.2.2 Method

We used a word's relative speech rate as a predictor to the degree in which the word is subject to coarticulation pressures. As in Cohen Priva & Gleason (2018; and in Cohen Priva, 2017), speech rate was calculated by dividing the duration of the actual word by the geometric mean duration of all other word tokens of the same type. Thus, if the duration of a token *give* was 300ms, and the mean duration of all tokens *give*, apart from the token being calculated, was 400ms, then the speech rate would have been 300/400=.75 (fast). However, had the token duration been 500ms, the speech rate would have been 500/400=1.2 (slow). As ratios, speech rate values were

4

Table 1: The number of tokens and frequent words of all the monophthongs used in Switchboard

Vowel	Count	Top words
/ɑ/	13580	got, job, bought
/æ/	12252	back, bad, fact
/ʌ/	25411	just, stuff, cut
/ɔ/	4774	thought, talk, saw
/ɛ/	24333	get, guess, said
/ɪ/	12829	kids, big, bit
/i/	10232	see, keep, eat
/ʊ/	11539	good, put, took
/u/	5654	two, food, shoot

log-transformed. We are not claiming that speech rate fully represents changes of vowel duration, but it does represent the functional pressure of interest, which is increased coarticulation.

We used two approaches to find the trajectory of vowels in the F1/F2 space as speech rate gets faster: A binning approach, that has fairly weak assumptions regarding the influence of duration on formant values, and a mixed effects linear regression, used for hypothesis testing, which assumes that the relationship between speech rate and movement along each formant can be described using a cubic polynomial. Both methods are described below.

First, we binned the vowels by speech rate quintiles (roughly 20% of the tokens in each bin) and used a mixed effect linear regression to find the mean F1 and F2 of each bin. We controlled for F0 and log F0 in addition to a number of random intercepts, which included the word, the speaker, as well as the two preceding and the two following phonemes, even if they were in different words, but as long as they were in the same utterance. An example from the data we used can be found in (1). For each vowel, we ran one model for the F1 values and a separate model for the F2 values. The mixed effects regression terms are illustrated in (2), which uses lme4 terms (Bates, Mächler, Bolker, & Walker, 2015). The intercept term is explicitly excluded, so the value of each quintile is the value that minimizes the squared error in that quintile rather than the distance from the first quintile. We use this approach to see whether we should expect the vowels' movement across the F1/F2 space to be roughly monotonic and to provide an illustration of the data, which is provided below under the results sections.

(1)

Phoneme	Speech Rate 5-tile	F1 mode	F2 mode	F0 mode	Speaker	Word
/ʊ/	5	568	1689	189	1020	good
/ɪ/	4	600	1829	151	1044	bit

Preceding word	Following word	Preceding 1	Following 1	Preceding 2	Following 2
a	thing	/g/	/d/	/ə/	/θ/
little	nice	/b/	/t/	/l/	/n/

(2) *F1 Mode ~ 0 + as.factor(Speech Rate 5-tile) +*
     stand.(F0 mode) + stand.(log(F0 mode)) +
     (1 | Speaker) + (1 | Word) +
     (1 | Preceding 1) + (1 | Preceding 2) +
     (1 | Following 1) + (1 | Following 2)

For statistical significance testing, we focused on F1 values, as all vowels are expected to center along the F2 axis (Mooshammer & Geng, 2008; Stevens & House, 1963). We used log F1 as the dependent variable and log relative speech rate as the independent variable of interest. Since we were not sure that the relationship between the two could be described in linear terms, we added quadratic and cubic terms. Together, they support the assumption that the relationship between log relative speech rate and log F1 values can be described using a repeated cubic polynomial. We also used log F0 as a control as well as the same random intercepts found in the binning approach, but we added random slopes for standardized log speech rate to the random intercepts of the immediately preceding and following contexts. This procedure resulted in the formula represented in (3). P-values are provided using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017), which encapsulated lme4, and provides degrees of freedom using Satterthwaite's method. We removed random slopes that had correlations with the intercept of 1, -1, and when the intercept was assigned no variance. Similarly, we removed cubic and quadratic components if they were not statistically significant. All continuous variables were standardized.

(3) *stand.(log(F1 mode)) ~ 1 + poly(stand.(log(Speech Rate)),3) +*
     stand.(log(F0 mode)) +
     (1 | Speaker) + (1 | Word) +
     (1 + stand.(log(Speech Rate)) | Preceding 1) + (1 | Preceding 2) +
     (1 + stand.(log(Speech Rate)) | Following 1) + (1 | Following 2)

2.3   Results and discussion

The results for the binning approach are presented in Figure 1 and in Table 8 in the appendix. It is evident that while all vowels centralize in the F2 space, all vowels, with the possible exception of high tense vowels /i/ and /u/, shift to lower F1 values when speech rate increases. Table 2 shows the fixed effects of the polynomial results for the four high vowels. The standard deviations of the random effects are in Table 3. Quadratic and cubic terms were not statistically significant and were removed. For /ɪ/, /ʊ/, and to a lesser extent, /u/, F1 values were correlated with speech rate, such that faster speech was associated with lower F1 values. For /i/, however, a statistically significant (yet small) negative correlation was found, which means that faster speech rate correlated with *higher* F1 values.
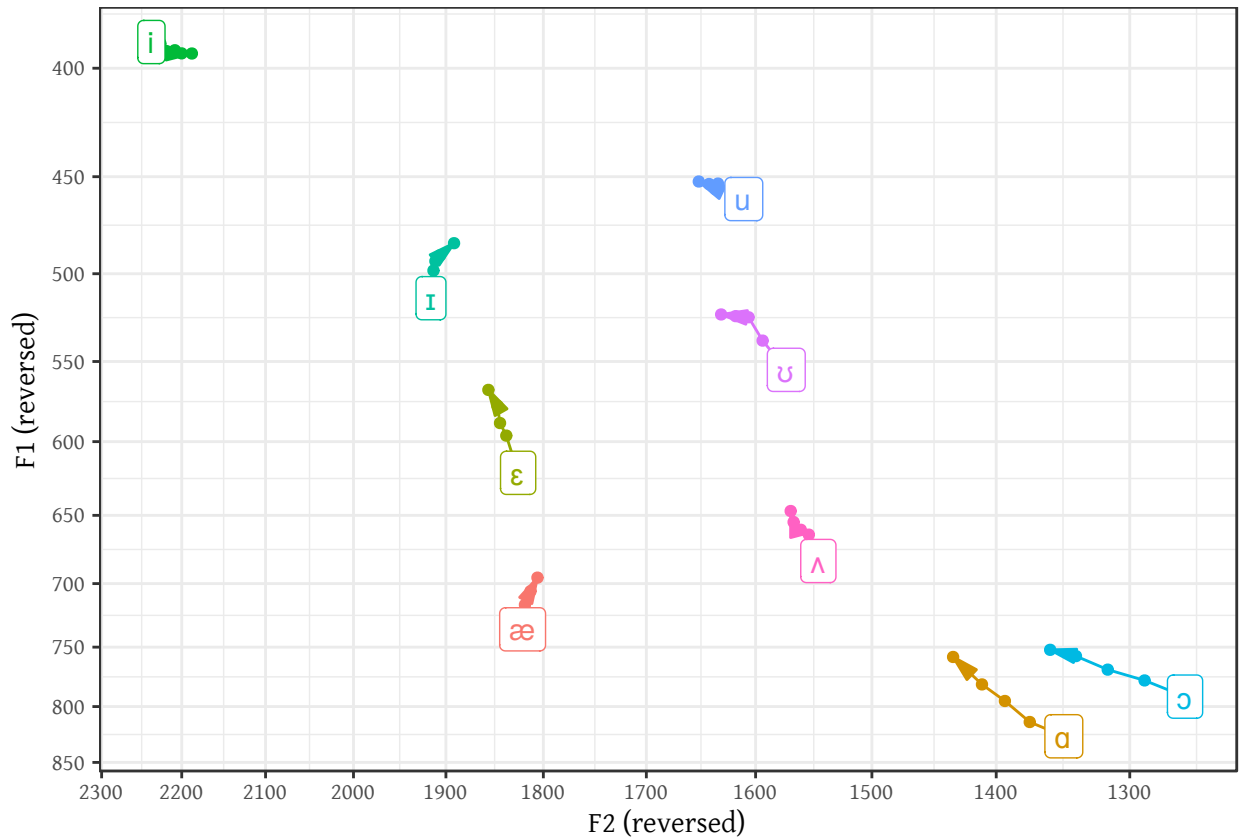
Figure 1: The relationship between American English non-rhotic monophthongs and speech rate in Switchboard. Both axes are flipped per convention and log-transformed, but showing their original values. The vowel label marks the geometric mean F1 and F2 values for the fifth quintile (longest duration). Each successive point marks the next quintile, with the arrow pointing at the first quintile (shortest duration). All vowels seem to center along the F2 axis when their duration gets shorter; all but the tense high vowels seem to move to lower F1 values.

Table 2: Study 1 models. In all four studies, F0 values are correlated with F1 values. (a) Shows the single model results for /ɪ/, in which speech rate is correlated with F1 values. (b) Shows the single model results for /ʊ/, in which speech rate is correlated with F1 values. (c) Shows the single model results for /i/, in which speech rate is *negatively* correlated with F1 values (reduced duration is associated with higher F1), suggesting that the maximally coarticulatory position does not have F1 values that are as low as those of /i/. (d) Shows the single model results for /u/, which patterns like the lax vowels, with a positive (but weak) correlation between speech rate and F1 values. For both /i/ and /u/ the correlation is smaller than for lax high vowels.

(a) /ɪ/

	β	SE	df	t	p
Intercept	0.14	0.11	12.3	1.3	0.231
Sp. rate (std.)	0.14	0.02	10.7	5.9	<0.001
F0 (std. log)	0.08	0.01	7778.6	6.4	<0.001

(b) /ʊ/

	β	SE	df	t	p
Intercept	0.26	0.14	9.3	1.8	0.098
Sp. rate (std.)	0.16	0.03	5.4	5.0	0.003
F0 (std. log)	0.05	0.01	7188.3	4.4	<0.001

(c) /i/

	β	SE	df	t	p
Intercept	-0.16	0.06	71.7	-2.7	0.009
Sp. rate (std.)	-0.03	0.01	5181.4	-2.8	0.005
F0 (std. log)	0.15	0.01	5727.2	10.0	<0.001

(d) /u/

	β	SE	df	t	p
Intercept	0.32	0.14	17.3	2.2	0.038
Sp. rate (std.)	0.04	0.03	16.3	1.4	0.182
F0 (std. log)	0.26	0.02	2405.8	12.7	<0.001

Table 3: The SD of random effects for intercepts and slopes in Study 1 models, using the Switchboard corpus. If the slope was removed, the corresponding line is blank. Overall, immediate contexts were associated with higher variance than non-immediate contexts, which is to be expected.

	/ɪ/	/ʊ/	/i/	/u/
Preceding 1 intercept	0.000	0.338	0.094	0.406
Preceding 1 speech rate slope		0.068		
Preceding 2 intercept	0.032	0.085	0.054	0.041
Following 1 intercept	0.324	0.060	0.110	0.217
Following 1 speech rate slope	0.065			0.088
Following 2 intercept	0.145	0.208	0.063	0.064
Word intercept	0.173	0.043	0.122	0.340

These findings suggest that the position in which coarticulation pressures are strongest is more similar to that of central high vowels, /ɨ/ or /ɨ/, and not of the mid-central position often associated with schwa (where the International Phonetic Alphabet places the symbol /ə/). That /i/ does not shift to lower F1 values may mean that the articulatorily neutral position is not as high as that of /i/.

We are not arguing that fast speech rate and the coarticulatory pressures associated with fast speech cause vowels to move to lower F1 positions. Perhaps vowels shift to lower F1 values for other reasons, resulting in faster speech rate. Or, perhaps both effects (lower F1 values and faster speech rate) result simultaneously from a common factor. The correlation between faster speech rate and lower F1 values does suggest that closer articulation is more compatible with restricted-duration articulation, in which higher coarticulation is observed. To elucidate this point, it would be worthwhile to compare schwa to other vowels of the same duration. This is the focus of Study 2.

3 Study 2

3.1 Introduction

Study 1 leaves several questions open. Schwa is generally understood as occupying a position more similar to that of a mid-vowel (Carr, 2019; Giegerich, 1992; McMahon, 2001), but Study 1 could not measure schwa performances directly. Does schwa occupy the articulatorily neutral position or a mid-vowel position? A target-less schwa approach would expect schwa not to move along the F1 axis when it gets shorter, as its articulation would be dictated by coarticulation-pressures regardless of its duration. If schwa does have a lowered tongue position/higher F1 target, then its F1 value should get higher the shorter its duration gets, though perhaps not as high as /i/. Second, a target-less schwa approach would expect all vowels to approach schwa's F1 value as they get shorter. In contrast, if schwa does have an articulatory target then other vowels, as a result of articulation, should approach a lower-than-schwa's F1 position when their duration gets shorter. Specifically, the difference between the two behaviors could be evident among the lax high vowels, as their initial F1 value is similar to or

higher than that of schwa.

All of these questions could be resolved if the results of Study 1 are replicated using the smaller Buckeye corpus, which does provide segment-level alignments, making it possible to extract the formant values of schwa directly. We therefore focus on schwa and American English's two high lax vowels /ɪ/ and /ʊ/. For completeness, we also provide the results for /i/ and /u/, as we provided them for Switchboard. For all five vowels we fitted per-vowel models (as described in Section 2.2.2). We also fitted two comparative models to compare the two high lax vowels with schwa.

Support for the target-less schwa would be found if: (a) Schwa's F1 value does not change when its duration changes, (b) vowels with F1 values similar to schwa approach schwa's location in the F1 space. Support for the alternative hypothesis, such that schwa does have an articulatory target, would be found if: (a) Schwa's F1 value changes when its duration changes, specifically to a lower F1 value if the observation in Browman & Goldstein (1994) is correct, and (b) high lax vowels with F1 values similar to schwa similarly shift to lower F1 values when their duration gets shorter.

3.2   Methods and materials

We use the Buckeye Corpus of Conversational Speech (Pitt et al., 2007), which provides data collected from The Ohio State University, where 40 speakers conversed freely with an interviewer. The corpus provides several values for each word, including its duration, part of speech, underlying form, and actual pronunciation. For each word, underlying and surface segments were aligned using a procedure detailed in Cohen Priva (2015) and Cohen Priva & Gleason (n.d.). The goal of the procedure was to align underlying dictionary forms with their surface realizations, as transcribed in the corpus. For instance, if the word *backs* /bæks/ surfaced as [bɜz], the procedure would align /b/ with [b], /æ/ with [ɜ], /s/ with [z], and regard /k/ as deleted. To implement such a procedure, an algorithm was trained on the entire corpus to deduce which correspondences and deletions were more likely than others. Given that we wanted to focus exclusively on (unstressed) schwa in relation to vowels marked as having primary stress, the underlying representations provided by Buckeye were not used, as Buckeye does not provide information related to stress. Instead, we replaced Buckeye's underlying forms with their CMU equivalents (Weide, 2008). Another reason to employ a different underlying representation is that it is impossible to distinguish between schwa and /ʌ/ using Buckeye's underlying representations.[1] The formant values of all the underlying vowels in the Buckeye corpus, as calculated using our procedure, are available at https://github.com/ucpresearch/moredata (note that these include vowels we excluded based on the following criteria).

We excluded all word-final schwas as well as schwas in words containing apostrophes (e.g. *Rosa's*) in order to control for and eliminate word-final schwas within complex words. We also excluded penultimate schwas in

---

[1]In the CMU dictionary, /ʌ/ is marked as ah1 or ah2, depending on whether it has primary or secondary stress, respectively. Schwa is marked using ah0, i.e. as unstressed /ʌ/ (which does not exist in American English). Buckeye drops the stress marks (0, 1, 2), which conflates the two vowels. For instance, in *Columbus*, both the second and third vowels are specified as ah, when in reality, the former is actually /ʌ/ and the latter is a schwa.

words that ended with *-ed* or *-es* to remove epenthetic vowels (e.g. in *roses* and *wanted*).[2] We further excluded all vowels that were followed underlyingly by nasals, liquids or glides to rule out possible effects of nasalization as well as coalescence (Cho, Kim, & Kim, 2017; Gick & Wilson, 2006; Proctor & Walker, 2012). This was particularly necessary because the CMU dictionary often lists words that have syllablic /l/ and syllabic nasals as having schwa + liquid and schwa + nasal sequences, e.g. *people* (/pɪpəl/), *button* (/bʌtən/). The resulting number of tokens for each vowel, as well as the three most frequent words are found in Table 4.

To extract formant values we first split the sound files into individual words and then used Praat (Boersma & Weenink, 2008) to extract phonetic properties, including pitch (`To Pitch: 0.001, 75, 300`) and formant values (`To Formant (burg)... 0.01 3 3500 0.025 50`). To maintain consistency with Switchboard, which has a lower sampling frequency than Buckeye, we used a lower maximum frequency than the default. For each vowel, we calculated a cubic polynomial regression using mel-transformed formant values. Then, we used the predicted value at the midpoint of each vowel (transformed back to Hz) as the actual value. The goal of using a regression rather than raw values was to overcome isolated mis-measurements of formant values. For F0 values, we used the mode of the distribution of mel-transformed F0 values across the vowel (transformed back to Hz values).

Table 4: The number of tokens and frequent words of all the monophthongs used in the Buckeye corpus'

Vowel	Count	Top words
/ɑ/	4161	lot, got, probably
/æ/	4337	back, actually, bad
/ə/	5459	columbus, society, university
/ʌ/	5802	just, much, stuff
/ɔ/	1208	thought, talk, talking
/ɛ/	8351	get, said, never
/ɪ/	5621	kids, little, different
/i/	4949	people, even, need
/ʊ/	1462	good, look, put
/u/	1799	used, usually, moved

In Study 2a, we followed the same statistical modeling approach described above for the Switchboard corpus; however, to check whether the results would hold given a different predictor of coarticulation, we used vowel duration directly rather than speech rate. In all five models, we removed random slopes that had correlations with the intercept of 1, -1, and when the intercept was assigned no variance. Similarly, we removed cubic and quadratic components if they were not statistically significant. All continuous variables were standardized.

In Study 2b, we fitted two comparative models, in which we contrasted the behavior of a vowel relative to schwa, in order to discern whether the lax high vowels behave differently from schwa. For the comparative models, we added segment identity (with schwa as a baseline) as a fixed predictor, its interaction term with log segment duration, and a per-segment random slope of the immediately preceding and the immediately following context,

---

[2]The common CMU representation for epenthetic vowels in such words is an unstressed /ɪ/, not a schwa.

resulting in (4) for single vowel models and in (5) for comparative models.

(4) *stand.(log(F1 mode)) ~ 1 + poly(stand.(log(Dur.)),3) +*
stand.(log(F0 mode)) +
(1 | Speaker) + (1 | Word) +
(1 + stand.(log(Dur.)) | Preceding 1) + (1 | Preceding 2) +
(1 + stand.(log(Dur.)) | Following 1) + (1 | Following 2)

(5) *stand.(log(Mid-F1)) ~ 1 +*
*Seg-Identity * poly(stand.(log(Dur.)),3) +*
stand.(log(F0 mode)) +
(1 | Speaker) + (1 | Word) +
(1 + stand.(log(Dur.)) + Seg-Identity | Preceding 1) +
(1 | Preceding 2) +
(1 + stand.(log(Dur.)) + Seg-Identity | Following 1) +
(1 | Following 2)

3.2.1 Study 2a results

Figure 2 uses the binning quintile models, demonstrating that all vowels seem to center in the F2 space, as would be expected in all vowel reduction models. Lower F1 values are associated with lower duration for all vowels except those that are classified as tense and high. Specifically, such trends are also observed for schwa and the lax high vowels. This would support the claim that schwa does have a higher F1 target than what coarticulation pressures alone would predict.

In all five continuous models, the quadratic and cubic components were not statistically significant, and were removed from the models. The relationship described below is therefore linear. However, we do not argue that the relationship is actually linear; we are merely interested in the existence of a trend and its direction (i.e. as a correlation test, but controlling for context and other factors). The full single vowel models are provided in Table 5. The standard deviations of the random effects are in Table 6. We report the crucial variables in the text below.

For schwa and /ɪ/ (using the single vowel models), longer duration predicted greater F1 values (β=0.12, SE=0.01, df=3324.3, t=8.44, p<.001; β=0.12, SE=0.03, df=11.9, t=3.84, p=0.0024, respectively).
The relationship between F1 values and duration for /ʊ/ and /u/ was not significantly different from zero. As in Study 1, longer duration predicted *lower* F1 values for /i/ (β=-0.071, SE=0.01, df=2934.4, t=-4.74, p<.001).

The results are therefore largely consistent with schwa having an articulatory target from which it shifts when its duration gets shorter. Lax high vowels do not approach schwa when they get shorter, though /i/ does to some extent (though it may approach a lax high position, rather than schwa's).
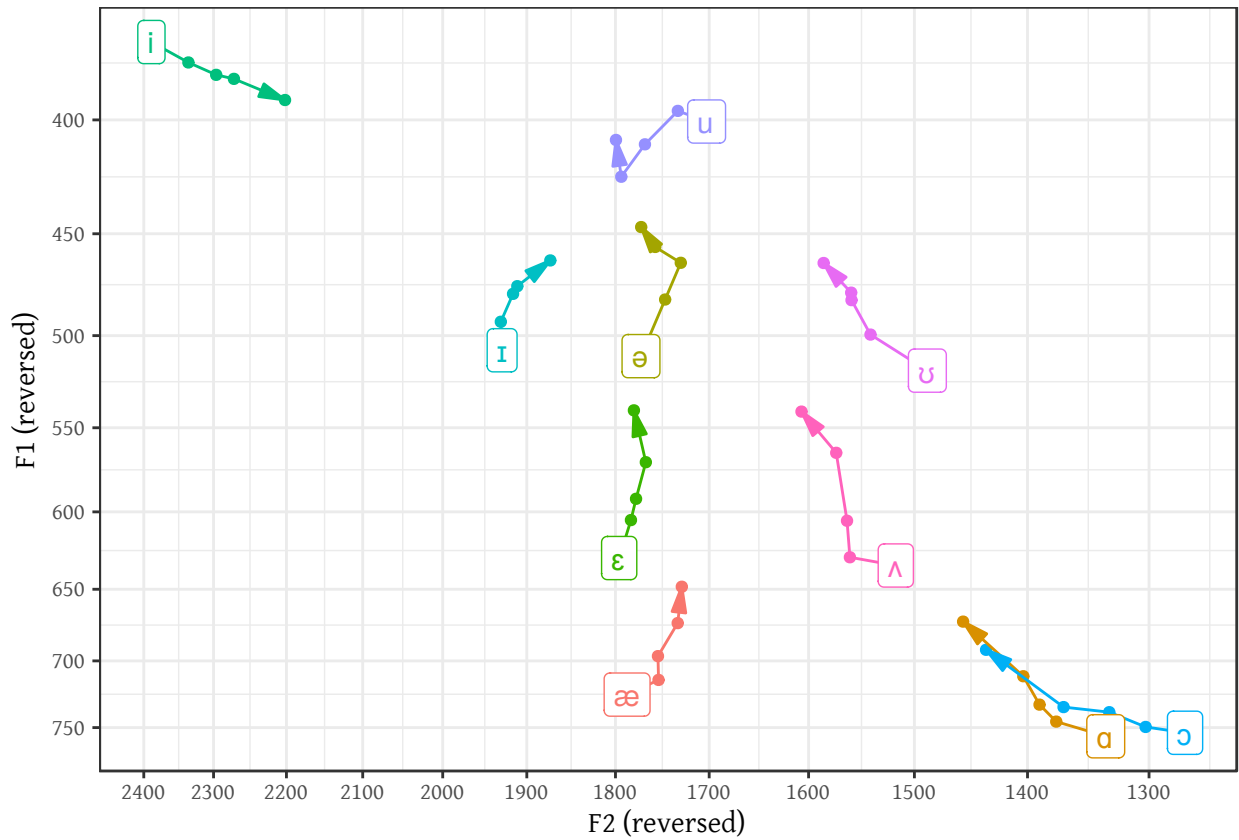
Figure 2: The relationship between American English non-rhotic monophthongs and duration. Both axes are flipped per convention and log-transformed, but showing their original values. The vowel label marks the geometric mean F1 and F2 values for the fifth quintile (longest duration). Each successive point marks the next quintile, with the arrow pointing at the first quintile (shortest duration). All vowels seem to center along the F2 axis when their duration gets shorter; all but the tense high vowels seem to move to lower F1 values. Crucially, schwa moves to lower F1 values too.

Table 5: Study 2 single vowel models. (a) Shows the results for schwa, in which duration is correlated with F1 values (reduced duration is associated with lower F1). (b) Shows the results for /ɪ/, in which duration is correlated with F1 values, similar to the effect we find for schwa, but F0 is also correlated with F1. (c) Shows the results for /ʊ/, in which no significant correlation is found. (d) Shows the results for /i/, which like Study 2, finds a negative correlation between duration and F1 values. (e) Shows the results for /u/, in which no significant correlation is found.

(a) Schwa

	β	SE	df	t	p
Intercept	-0.13	0.09	28.6	-1.5	0.155
Duration (std. log)	0.12	0.01	3324.3	8.4	<0.001
F0 (std. log)	0.01	0.02	1915.9	0.7	0.461

(b) /ɪ/

	β	SE	df	t	p
Intercept	0.04	0.09	18.3	0.4	0.672
Duration (std. log)	0.12	0.03	11.9	3.8	0.002
F0 (std. log)	0.04	0.02	2571.7	2.5	0.012

(c) /ʊ/

	β	SE	df	t	p
Intercept	0.02	0.16	4.4	0.1	0.89
Duration (std. log)	0.11	0.11	2.7	1.0	0.401
F0 (std. log)	0.06	0.03	707.4	1.9	0.053

(d) /i/

	β	SE	df	t	p
Intercept	0.20	0.12	21.9	1.6	0.119
Duration (std. log)	-0.07	0.01	2934.4	-4.7	<0.001
F0 (std. log)	0.02	0.02	3414.7	1.2	0.223

(e) /u/

	β	SE	df	t	p
Intercept	0.10	0.12	21.3	0.9	0.39
Duration (std. log)	-0.02	0.03	1153.9	-0.8	0.428
F0 (std. log)	0.04	0.03	294.6	1.4	0.166

Table 6: The SD of random effects for intercepts and slopes in Study 2a models, using the Buckeye corpus. If the slope was removed the corresponding line is blank. Overall, immediate contexts were associated with higher variance than non-immediate contexts, which is to be expected.

	/ə/	/ɪ/	/ʊ/	/i/	/u/
Preceding 1 intercept	0.138	0.266	0.152	0.092	0.255
Preceding 1 speech rate slope					
Preceding 2 intercept	0.104	0.048	0.000	0.041	0.000
Following 1 intercept	0.237	0.175	0.275	0.395	0.305
Following 1 speech rate slope		0.095	0.207		
Following 2 intercept	0.085	0.000	0.069	0.068	0.000
Word intercept	0.200	0.487	0.000	0.221	0.125

3.2.2 Study 2b results

The full comparison models are provided in Table 7. In all five continuous models, the quadratic and cubic components were not statistically significant, and were removed from the models.

In the comparative model that included schwa and /ɪ/, we found the same effect for duration, in which longer duration translated to higher F1 values (β=0.12, SE=0.03, df=11.7, t=4.44, p<.001). This is to be expected since schwa was used as a baseline (the default factor level). The F1 values of /ɪ/ were not significantly different from those of schwa (β=-0.048, SE=0.05, df=15.6, t=-1.03, p=0.32), and the interaction term between segment identity and log duration was not statistically different either (β=-0.021, SE=0.02, df=861, t=-1, p=0.32). The comparative model that included schwa and /ʊ/ similarly found the same effect for duration, as for this model too, schwa is the default factor level (β=0.13, SE=0.03, df=8.5, t=4.08, p=0.0031). The F1 values of /ʊ/ were not significantly different from those of schwa (β=-0.021, SE=0.07, df=338.9, t=-0.29, p=0.77), and the interaction term between segment identity and log duration was not statistically different either (β=-0.055, SE=0.03, df=177.9, t=-1.67, p=0.096; the negative trend is expected, because /ʊ/ did not show a correlation with duration in the single-vowel model).

3.3 Discussion

The fact that F1 is positively correlated with duration for schwa implies that speakers deviate from a position dictated by coarticulation when the duration of schwa is longer, which is in agreement with schwa having an articulatory target. The results for /ɪ/ echo those found in Study 1, and imply that coarticulation pressures push /ɪ/ toward lower F1 values than those of mean-duration /ɪ/. This too suggests that the position dictated by coarticulatory pressures has lower F1 values than the F1 values speakers would use for /ɪ/ when it is not reduced (e.g. in careful articulation). The evidence is less conclusive for /ʊ/, which did shift to lower F1 vowels in fast speech rate in Study 1.

The results for /i/ suggest that it is not the case that all vowels have lower F1 values when their duration get shorter. An articulatorily "true" neutral position is therefore not "as high as possible," but is more similar to

Table 7: Study 2b comparison models. (a) Shows the comparison between schwa and /ɪ/, and which shows that /ɪ/ is not significantly different from schwa. (b) Similarly shows that though there is not a significant correlation between duration and F1 values for /ʊ/, the model does not find significant differences between schwa and /ʊ/.

(a) Schwa vs. /ɪ/

	β	SE	df	t	p
Intercept	0.0008	0.09	31.9	0.0	0.993
phone is /ɪ/	-0.0478	0.05	15.6	-1.0	0.32
Duration (std. log)	0.1203	0.03	11.7	4.4	<0.001
F0 (std. log)	0.0212	0.01	5094.4	1.7	0.097
phone is /ɪ/ : Duration (std. log)	-0.0209	0.02	861.0	-1.0	0.316

(b) Schwa vs. /ʊ/

	β	SE	df	t	p
Intercept	-0.10	0.09	26.0	-1.1	0.288
phone is /ʊ/	-0.02	0.07	338.9	-0.3	0.771
Duration (std. log)	0.13	0.03	8.5	4.1	0.003
F0 (std. log)	0.01	0.02	2653.7	0.9	0.366
phone is /ʊ/ : Duration (std. log)	-0.05	0.03	177.9	-1.7	0.096

that of a lax high vowel. That /u/ and /ʊ/ do not show this trend may have to do with reduced duration affording less time for lip-rounding, but we cannot conclude that this is the case.

4 General discussion

We provide two studies, each using a different corpus and a different way to approximate the strength of coarticulatory pressures, while arriving at largely consistent results. Both studies show that as coarticulatory pressures increase in American English, vowels tend to center along F2 and move to lower values in F1, though not to F1 values associated with tense high vowels. This view is largely consistent with the general understanding of the pressures that lead to vowel neutralization in unstressed syllables (e.g. Flemming, 2004), though with two important caveats: We do not observe evidence for neutralization to schwa, and we do not observe evidence for neutralization as a process driven largely by the selective inability to produce low vowels in reduced-duration syllables.

Vowels seem to neutralize to a position closer to that of a central (possibly lax) high vowel rather than to where the International Phonetic Alphabet (IPA) places schwa, which is in the middle of the vowel chart. Study 2 does indeed show that when American English schwa's duration is long, it seems to occupy a position that is largely consistent with the position assumed by the IPA symbol placement. However, all vowels that could show a difference between neutralization to schwa and neutralization to a vowel with a higher position seem to show

it: In Study 1, lax high vowels shift to lower-still F1 values, and in Study 2, both /ɪ/ and schwa shift to lower F1 values. It is therefore difficult to argue that schwa's position is determined solely by coarticulatory pressures given that such pressures do not result in neutralization to schwa. We propose that the difference between the common assumption and our finding is that existing literature seems to assume that two neutral positions are one and the same: the *articulatorily* neutral position and the *perceptually* neutral position. Our studies are only suited to find the former, not the latter: We simply observe what happens to vowels when coarticulatory pressures increase. This leaves room for perceptual pressures to have distinct effects in determining the position of schwa. For instance, following dispersion-theoretic considerations (Flemming, 2004), one could assume that schwa needs to contrast with other vowels, at least in signaling that the syllable in which it occurs is unstressed, or, if this is the case, in signaling that it contrasts with other vowels that also occur in unstressed syllables, e.g. with unstressed /ɪ/. The higher-than-expected F1 values associated with schwa could just be a vehicle for that contrast. This explanation parallels the explanation provided by Kingston (2008) and Katz (2016) for consonant lenition. Kingston and Katz argue that just as there are pressures signaling prosodic prominence, there could be pressures signaling prosodically weak positions as well. For vowels, having a central F1 value alongside a central F2 value could serve this purpose well.

Our findings also bear on the suggestion that vowel neutralization follows from low vowels encroaching on the acoustic space of other vowels when their duration decreases (Flemming, 2004). This explanation is meant to explain a common cross-linguistic pattern e.g. in Central Italian, in which /ɛ/ and /ɔ/ neutralize to /e/ and /o/ in weak positions (collapsing /u,o,ɔ,a,ɛ,e,i/ systems to /u,o,a,e,i/ systems), and in more extreme cases in which all back and front vowels neutralize to /u/ and /i/ respectively (to /u,a,i/ systems). The pattern we observe in studies of stressed syllables is more consistent with a pure articulatory explanation for these two stages of vowel neutralization patterns: All non-high vowels shift to lower F1 values when their duration gets shorter. The rest of the dispersion-theoretic explanation provided by Flemming (2004) seems to hold: The reduced acoustic space in itself is insufficient in explaining why /e,ɛ/ and /o,ɔ/ *cannot* contrast, and neutralization probably follows from an inability to make the contrast clear enough.

Finally, though our studies were not designed to test whether some vowels are more variable than others, the random effect sizes in the different models, as provided in Table 6 do not seem to support the assumption that schwa is more likely than other vowels to be affected by coarticulatory pressures, at least not in absolute values when compared to its two closest competitors in F1, /ɪ/ and /ʊ/.

In the future, we hope to replicate these findings in other languages, with different consonantal contexts and different vowel systems, to make sure the results reflect articulatory pressures overall, rather than those typical in English.

5   Conclusion

There seems to be an open question in our understanding of schwa and vowel neutralization: Does schwa occupy an articulatorily neutral position, based on acoustic data, as Flemming (2009) argues, or does it have an

articulatory-specified position based on articulatory data, as Browman & Goldstein (1994) suggest. We used a novel approach, which relies on an observation by Lindblom (1963), that as vowels get shorter, they increasingly reflect coarticulatory pressures. We benefited from using large corpora of carefully annotated speech (the Switchboard and Buckeye corpora), and although we used a different measure of coarticulation for each corpus (speech rate and vowel duration), we arrived at the same conclusion as Browman & Goldstein (1994), namely that the articulatorily neutral position is similar to that of a lax high central vowel, and that schwa does not occupy this position. Our results bear on our understanding of schwa possibly as a perceptually rather than an articulatorily neutral vowel, and on the mechanics of vowel neutralization as a whole.

Table 8: Study 1 vowel F1 / F2 mean locations, by quintile (lower quintile is faster speech rate), in the Switchboard corpus. F1 and F2 ranges are two standard errors above and below the mean. F1 and F2 are calculated as the mode of F1 and F2 in the word that contains them. With the exception of /i/ and /u/, all other vowels including /ɪ/ and /ʊ/ seem to shift to lower F1 values when speech rate increases.

Phoneme	Quintile	F1	F1 range	F2	F2 range
æ	1	695	676–715	1806	1778–1834
æ	2	706	686–726	1813	1785–1841
æ	3	716	697–736	1818	1791–1846
æ	4	722	702–742	1819	1792–1847
æ	5	736	715–756	1821	1793–1849
ɑ	1	758	732–785	1434	1387–1482
ɑ	2	781	754–809	1411	1365–1458
ɑ	3	795	768–824	1393	1348–1440
ɑ	4	814	785–843	1374	1330–1420
ɑ	5	826	798–856	1348	1305–1394
ɛ	1	567	535–601	1856	1816–1897
ɛ	2	577	545–611	1846	1807–1887
ɛ	3	588	555–623	1844	1804–1885
ɛ	4	596	563–632	1837	1798–1878
ɛ	5	621	586–658	1825	1786–1866
i	1	394	384–404	2188	2160–2216
i	2	394	384–403	2200	2173–2227
i	3	392	383–402	2208	2182–2235
i	4	389	380–399	2226	2200–2253
i	5	389	379–398	2237	2210–2265
ɪ	1	484	465–504	1891	1832–1953
ɪ	2	490	471–510	1906	1847–1968
ɪ	3	493	474–513	1911	1851–1973
ɪ	4	498	479–519	1913	1853–1975
ɪ	5	513	493–534	1916	1856–1978
ɔ	1	752	732–773	1359	1306–1414
ɔ	2	757	737–778	1339	1287–1394
ɔ	3	769	748–790	1316	1265–1369
ɔ	4	778	757–799	1289	1239–1341
ɔ	5	791	770–813	1261	1212–1312
u	1	452	429–477	1651	1512–1804
u	2	456	433–480	1636	1498–1786
u	3	454	431–478	1642	1503–1793
u	4	453	431–477	1634	1496–1784
u	5	461	438–485	1611	1475–1758
ʊ	1	523	497–549	1631	1513–1758
ʊ	2	524	498–550	1618	1501–1744
ʊ	3	524	499–551	1606	1490–1731
ʊ	4	538	512–565	1594	1479–1718
ʊ	5	555	528–583	1573	1459–1696
ʌ	1	647	619–676	1569	1511–1630
ʌ	2	655	627–684	1567	1508–1627
ʌ	3	660	632–690	1560	1502–1621
ʌ	4	664	635–694	1553	1495–1614
ʌ	5	683	654–714	1545	1488–1605

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. http://doi.org/10.18637/jss.v067.i01

Bergem, D. R. van, & Beinum, F. J. K.-v. (1989). Vowel reduction in natural speech. In *Eurospeech-1989* (pp. 2285–2288).

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (Version 6.0.37) [Computer program].

Browman, C. P., & Goldstein, L. (1994). "Targetless" schwa: An articulatory analysis. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II gesture, segment, prosody* (pp. 26–67). Cambridge: Cambridge University Press.

Carr, P. (2019). *English phonetics and phonology*. John Wiley & Sons.

Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics*.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, *6*(2), 243–278. http://doi.org/10.1515/lp-2015-0008

Cohen Priva, U. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, *160*, 27–34. http://doi.org/10.1016/j.cognition.2016.12.002

Cohen Priva, U., & Gleason, E. (2018). The role of fast speech in sound change. In *Proceedings of the 40th annual conference of the cognitive science society* (pp. 1512–1517). Austin, TX: Cognitive Science Society. Retrieved from http://mindmodeling.org/cogsci2018/papers/0293/index.html

Cohen Priva, U., & Gleason, E. (n.d.). The causal structure of lenition. *Language*. http://doi.org/10.31234/osf.io/awvzd

Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics*.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, *34*(1), 104–137. http://doi.org/10.1016/j.wocn.2005.03.004

Davidson, L., & Stone, M. (2004). Epenthesis versus gestural mistiming in consonant cluster production. In *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL) 22*. University of California, San Diego.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-Based Phonology* (pp. 232–276). Cambridge University Press.

Flemming, E. (2009). The phonetics of schwa vowels. In *Phonological weakness in English* (pp. 78–95). England: Palgrave Macmillan Houndsmill.

Flemming, E., & Johnson, S. (2007). Rosa's roses: Reduced vowels in American English. *Journal of the International Phonetic Association*, *37*(01), 83. http://doi.org/10.1017/S0025100306002817

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *The Journal of Acoustical Society of America*, *90*(4), 1816–1827. http://doi.org/10.1121/1.401662

Gick, B., & Wilson, I. (2006). Papers in laboratory phonology VIII: Varieties of phonological competence. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), (pp. 635–660). Mouton de Gruyter.

Giegerich, H. J. (1992). *English phonology: An introduction*. Cambridge University Press.

Godfrey, J. J., & Holliman, E. (1997). Switchboard-1 release 2.

Hall, N. E. (2003). *Gestures and segments: Vowel intrusion as overlap* (PhD thesis). University of Massachusetts.

Harkins, D., Feinstein, D., Lindsey, T., Martin, S., & Winter, G. (2003). Switchboard MS State manually corrected word alignments. Retrieved from https://www.isip.piconepress.com/projects/switchboard/

Harris, J. (2005). Vowel reduction as information loss. In *Current issues in linguistic theory* (pp. 119–132). John Benjamins Publishing Company. http://doi.org/10.1075/cilt.259.10har

Katz, J. (2016). Lenition, perception and neutralisation. *Phonology*, *33*(1), 43–85. http://doi.org/10.1017/S0952675716000038

Katz, J., & Fricke, M. (2018). Auditory disruption improves word segmentation: A functional basis for lenition phenomena. *Glossa: A Journal of General Linguistics*, *3*(1), 38. http://doi.org/10.5334/gjgl.443

Kaźmierski, K. (2015). *Vowel-shifting in the English language*. de Gruyter Mouton.

Kingston, J. (2008). Lenition. In *Selected proceedings of the 3rd conference on laboratory approaches to Spanish phonology* (pp. 1–31). Cascadilla Press Somerville, MA.

Kondo, Y. (1994). Targetless schwa: Is that how we get the impression of stress-timing in english? In *Proceedings of the edinburgh linguistics department conference '94*.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. http://doi.org/10.18637/jss.v082.i13

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, *35*(11), 1773–1781. http://doi.org/10.1121/1.1918816

McMahon, A. M. S. (2001). *An introduction to English phonology*. Edinburgh University Press.

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America*, *96*(1), 40–55. http://doi.org/10.1121/1.410492

Mooshammer, C., & Geng, C. (2008). Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association*, *38*(2).

Oostendorp, M. van. (2002). The second glot international state-of-the-article book. In L. Cheng & R. Sybesma (Eds.), (pp. 431–461). De Gruyter Mouton.

Oostendorp, M. van. (1995). *Vowel quality and phonological projection* (PhD thesis). Tilburg University.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye corpus of conversational speech (2nd release). Department of Psychology, Ohio State University.

Proctor, M., & Walker, R. (2012). Articulatory bases of sonority in English liquids. In *The sonority controversy*. De Gruyter.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in english /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*, 291–311.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech & Hearing Research*, *6*, 111–128.

Turnbull, R. (2019). Listener-oriented phonetic reduction and theory of mind. *Language, Cognition and Neuroscience*, *34*(6), 747–768. http://doi.org/10.1080/23273798.2019.1579349

Weide, R. (2008). The CMU pronunciation dictionary, release 0.7a.