Schwa's duration and acoustic position in American English

Uriel Cohen Priva, uriel_cohen_priva@brown.edu Emily Strand emmstrand@ucdavis.edu

Abstract

Is American English schwa's position determined solely by the context in which it appears? Do vowels neutralize to schwa when their duration is shorter? We address these two inter-related questions using the Buckeye corpus to study vowel behavior across multiple contexts of spontaneous speech. We find that all except tense high vowels shift to lower F1 values when their duration is relatively short, including lax high vowels and lexical schwas, rather than toward a mid-vowel position that schwa occupies when its duration is long. However, we also replicate the finding that schwa is more dependent on both context and duration than other vowels. The results are not consistent with the idea that schwa's position is determined exclusively by the context in which it appears. However, schwa's shift to higher F1 values when its duration is longer is not necessarily different from other vowels' shift to higher F1 values when their duration is longer, making it unnecessary to argue that schwa's mid-vowel properties are due to having a target in F1 terms.

1 Introduction

1.1 Overview

Is there such a thing as a neutral vocalic position? From a phonological standpoint, there is. Schwa is regarded as having a neutral, featureless position (e.g. Harris 2005; Oostendorp 1995), and is defined by the properties it lacks: it is not front, high, low, back, or round. In phonetic terms, however, it is not clear what properties define schwa, and whether there are reasons to treat such properties (if any) as neutral. This question is closely related to the study of the production of vowels when their duration is short. One view expects that as vowels shorten, they would become more schwa-like (Miller 1981, among others). This view corresponds to the phonological expectation, namely that if vowels cannot express their distinctive identity, they should phonologically *neutralize* to schwa. A different approach is more biomechanically-motivated (Lindblom 1963; Moon & Lindblom 1994), and argues that reduced-duration vowels should not head toward any position in particular, but should rather exhibit a growing effect of the context in which they occur as their duration becomes shorter, but they should not deviate from that compromise toward any one neutral position. From a phonological perspective, this would mean that vowels *assimilate* to their environment when their duration is shorter, rather than neutralize to schwa.

The contrast between the two views is further complicated by the properties of American English schwa. American English schwa phonologically alternates with full vowels (e.g. the first vowel in *photograph* vs. *photography*). If there is a neutral position, which schwa occupies, then reduced-duration vowels should head toward it. But if there is no neutral position, what is the phonetic manifestation of schwa's phonological neutral quality? If the assimilation-based prediction is correct, and vowels that become schwa lose their target rather than gain a different one, then schwa would be completely determined by the context in which it occurs. However, there is evidence that American English schwa has non-neutral articulatory and acoustic properties, which involve a more open jaw or tongue gesture (Browman & Goldstein 1994), a tongue root gesture distinct from the resting position (Gick 2002), or an acoustic target in F1, though not in F2 (Kondo 1994).

One limitation of existing investigations of the effect of duration on vowels in general and on schwa in particular, is that they overwhelmingly used a discrete number of distinctions in speech rate, e.g. fast vs. slow (see Moon & Lindblom 1994).[1] This limitation makes it statistically impossible to disentangle the possible contribution of different sources contributing to vowels' behavior when their duration is reduced. Another limitation is the contrast between articulatory and perceptual evidence for schwa's position. The bulk of evidence for schwa having a target comes from articulatory data (Browman & Goldstein 1994; Gick 2002), while acoustic evidence for that effect is not as robust (Flemming 2009). We address both limitations by using rich acoustic data from spontaneous speech in the Buckeye Corpus (Pitt et al. 2007), which provides more data in a richer set of contexts than previous studies.

We return to American English schwa and the way it compares to other vowels. One goal is to use rich data from spontaneous speech to find whether schwa has an acoustic target in American English. Another goal is to shed light on the dynamics of vowels when their duration gets shorter, and their relationship to schwa. Specifically, can vowel reduction be attributed to contextual effects alone? Do vowels exhibit any property of neutralizing to schwa? For both goals we focus on changes in F1, because it stands in the center of the apparent conflicting evidence between articulatory and acoustic evidence, and because American English schwa is a central vowel in a system that has few other central vowels,[2] making comparisons in F2 less useful.

Although a fair amount of research has been conducted on schwa and on vowel reduction (e.g. Browman & Goldstein 1994; Flemming 2009; Moon & Lindblom 1994), there is little evidence that draws on 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. Moreover, speech produced in a lab might be hyperarticulated, which is not expected in spontaneous speech. We take advantage of this in the current study through the use of the Buckeye corpus (Pitt et al. 2007), expanded on in the section below.

1.2 Current findings on American English schwa

Articulatory evidence suggests that schwa has an articulatory target in American English. Browman & Goldstein (1994) find that the articulation of American English schwa requires a lower tongue position than

[1]Bates (1995) is a notable exception, but the annotated speech was from a read corpus, rather than spontaneous speech, and therefore has less variable speech. The corpus is also substantially smaller than the one we use, and therefore provides less statistical power.

[2]Davidson (2006) finds that speakers insert a central high vowel to break non-native clusters. Other authors use /ɨ/ or /ɪ/ it to describe some American English epenthetic vowels, such as the second vowel in *roses* (e.g. eSpeak NG development team use /ɪ/; Weide 2008 use /ɪ/ for the same vowel).

coarticulation alone would predict. Gick (2002) finds that schwa is characterized by a tongue root gesture distinct from the resting position. Although Flemming (2009) similarly finds that schwa has higher F1 values than would be otherwise expected, he argues that evidence for schwa having a particular acoustic target is inconclusive. Flemming (2009) argues for two distinct schwa-forming processes. The first is a mid-central schwa (where the IPA chart places [ə]), which is often the counterpart of non-high vowels in unstressed syllables, contrasting only with high vowels. The second is a *variable schwa*, which is argued to "potentially [neutralize] all vowel qualities" and "[occur] where the speaker is not concerned to realize any particular vowel quality." Flemming stresses that schwa results from assimilation to context:

> "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: sec. 3) provides several possible reasons for why schwa would appear to warrant a lower tongue position and higher-than-expected F1 values if its acoustic position were completely determined by its context.[3] First, the data in Browman & Goldstein (1994) may have been interpreted as word-final, where coarticulation pressures are not as strong and underlying vowel qualities may be lower (e.g. the famous *Rosa's roses* minimal-pair contrast). Second, he proposes that assimilation to consonants may not always involve a 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. Third, he proposes that since schwa is a vowel, and vowels require a more open vocal tract position regardless of their identity, schwa's F1 position may follow from being a vowel, rather than from having a concrete acoustic 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 a minimal in-between vowel (excrescent vowels, see Hall 2011). In her study, 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, presumably reflective of some minimal vocalic position. This vowel seemed to have dif-

[3] By *acoustic position* we refer to the area in F1/F2/F3 space in which vowel tokens produced by a particular speaker are expected to be found in normal speech, if we remove the effect of context. This is meant to contrast with *acoustic target*, which also implies intent.

ferent articulatory properties than underlying schwas, as the corresponding F1 values were significantly lower. This difference, or the fact that this position is not where we find schwa, suggests 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 have specific articulatory targets, so these findings could be interpreted as evidence only for schwa as not being the epenthetic vowel in American English. But even so, epenthetic vowels apart from schwa tend to be high (other non-high epenthetic vowels do appear in languages, but they occur substantially less frequently than high epenthetic vowels, Kitto & Lacy 1999; Kim & Kochetov 2011). The particular context in Davidson makes it likely that the vowels are excrescent vowels, phonetically present but not phonologically driven (Hall 2011), and therefore likely to lack a target altogether. We may therefore ask why schwa, labeled as one of the most common epenthetic vowels (Hall 2003; Kim & Kochetov 2011), would not share this same property. Along with the evidence from Davidson's (2006) study, this not only suggests that minimal vowels might actually be higher (or of a lower F1 value) than previously suggested, but also that schwa might not be representative of this targetless position.

In the same vein, Bergem (1994), who looked at the acoustics of Dutch schwa with respect to coarticulatory effects, found that Dutch schwa's F1 values resemble those of high vowels, not mid vowels (Bergem 1994: fig. 3b), as would be expected given that [ə] is used to denote "schwa." Bergem (1994) and Beinum (1994) assert that schwa should be regarded as lacking an articulatory target, such that it is solely determined by coarticulatory pressures, which suggests that the coarticulation alone would dictate a higher position in the acoustic space. In terms of shared ancestry, Dutch and American English share many similarities; thus, we might expect the vowels that maximally reflect coarticulatory pressures in American English to also be higher than where the IPA chart places [ə]. If schwa occupies a position in the middle of the vowel space, it may signify that schwa has an acoustic target.

1.3 Approach

Our approach depends on the assumption that the articulation of vowels depends on two or more of the following three elements:

CONTEXT The context in which a vowel occurs, which restricts the vowel, at the very least, at the beginning and end of its production.

TARGET The specific acoustic properties of the vowel itself, different for each vowel, and present only if the vowel has an acoustic target. These properties are what makes vowels contrast. Schwa may or may not have that property.

VOWEL Some researchers argue that all vowels, including schwa, have properties that are common to all vowels. Bates (1995) relates such properties to the phonological designation [-consonantal, +syllabic]. Such properties could involve greater jaw and tongue displacement relative to consonants (Stevens 2000, ch. 6), and tongue root displacement (Gick 2002). Flemming (2009) attributes the findings in Browman & Goldstein (1994) to schwa having vocalic properties, rather than a target. Even if a vowel's production is largely determined by the surrounding context, being a vowel could still entail a more open vocal tract; otherwise, the resulting sound would not be a vowel. All sounds that surface as vowels are assumed to obey that requirement, including schwa.

We assume that the influence particular contexts have on the production of vowels can be inferred by statistical models in which the specific contexts in which the vowel occurs are known in advance. In Section 3 in the supplementary materials, we use an alternative method in which the initial and final formant values are measured directly, using Praat (Boersma & Weenink 2020). The acoustic properties of specific vowels are their corresponding formant values at the vowels' midpoint, independent of context, as calculated using FAVE (Rosenfelder et al. 2014).

We follow Lindblom (1963) and Moon & Lindblom (1994) and assume that the relative weight of the three properties listed above changes when vowel duration is reduced, such that the effect of context increases for shorter-duration vowels. This principle holds not only for vowels: Sproat & Fujimura (1993) found 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.

These assumptions lead to the following expectations for all vowels. Initially, a vowel's formants are most strongly affected by the preceding context. This effect diminishes during the articulation of the vowel, given that the articulators have time to move away from their previous position; they may even cease to affect the vowel after a while (Hertz 1991). Conversely, the vowel's formants become progressively more affected by the following context toward the end of the vowel (this influence too does not necessarily affect the vowel throughout its articulation). Both effects fall under the category CONTEXT above. The path from a position largely dictated by the previous context to a position largely dictated by the following context should also satisfy VOWEL, or the resulting sound would not be a vowel. This influence is expected to result in vowel-like displacement. Crucially, the path, dictated by the influence of the previous and following contexts, should also approach the vowel's acoustic target (TARGET), if such a target exists (Klatt 1973). For vowels that have a target, we expect that there would be room to satisfy TARGET, such that long-duration vowels would actually reach their respective acoustic targets, if they have one, and short-duration vowels would approach their targets. For target-less vowels, only CONTEXT and VOWEL determine production. This means that if changes in duration result in a vowel having different acoustic properties, then either the vowel has a target, or that the manifestation of being a vowel is affected by changes in duration.

Given this outline, there are at least three types of evidence that can be taken into account in order to determine whether schwa has a particular acoustic target, or alternatively, that the property of being a vowel is affected by vowel duration. One is specific to schwa itself, and two depend on either direct or indirect comparison to other vowels.

First, if schwa's acoustic position is determined exclusively by CONTEXT, then schwa's acoustic position should not be affected by its duration. That is, schwa should not move in the acoustic space when its duration is shorter than usual. Furthermore, the effect of context on schwa should apply independently from its duration. The generalization that shorter duration should result in stronger coarticulatory pressures should simply not hold for a vowel that lacks an acoustic target, as a targetless schwa should be affected by context to the same extent, even when its duration is long. This expectation hinges on the assumption that all vowels that surface as vowels satisfy the requirement in VOWEL, regardless of their duration or specific acoustic correlates. The expectation outlined in this paragraph is one-sided: If schwa does have a target, and that target is maximally compatible with the context in which it occurs, it may not move further. Study 1 aims to establish whether schwa's acoustic properties shift when its duration is shorter.

Second, Moon & Lindblom (1994) predict that shorter-duration vowels (that do have a target) should head toward articulatory positions that facilitate articulation but compromise the pressure to reach their respective acoustic targets, given a greater expected influence of CONTEXT relative to their TARGET. If schwa lacks an acoustic target, this means that other vowels should become more schwa-like when their duration is shorter (and that schwa itself should not be affected by duration, as explained above). One expectation is that increased coarticulatory pressures would lead to reduced jaw-lowering (Lindblom 1963: 1776–1777; Flemming 2004: sec. 3.2.2; Keating 1985). Lindblom (1963) found that low vowels shift to lower F1 values when their duration is short. This tendency stops at about 375 Hz (higher than lax high vowels, but not as high as /i/), but it is crucially higher than the mid-central position associated with schwa. If other vowels shift past IPA schwa' position, as Lindblom (1963) found, then the association of schwa with a mid-center position (e.g. in Recasens 2021: 500 Hz F1, 1500 Hz F2) has to follow from some influence other than CONTEXT. Finding that all vowels progress towards a higher-than-schwa position could also undermine our assumption that satisfying TARGET satisfies VOWEL by default. This set of questions is addressed in Study 2.

Finally, several authors assume that schwa lacks an acoustic target because it is affected by context to a greater extent than other vowels (Recasens 1991; Bates 1995; Flege 1988). The comparison between schwa and other vowels can be tricky because American English schwa is unstressed, has shorter duration than most vowels, and is therefore subject to greater coarticulatory pressures than longer-duration vowels, everything else being equal. Another complication is that other vowels are closer to the periphery of the vowel space and perhaps cannot move as much in response to coarticulatory pressures. Thus, if schwa is indeed more variable, its articulation should be more dependent on context than the articulation of similar (or more central) vowels, even when its duration is factored out. Study 3 replicates previous findings after controlling for duration.

We focus on casual spontaneous speech, as it provides many instances for the analysis of reduced-duration vowels (Turnbull 2019; Bergem & Beinum 1989). We use the Buckeye corpus (Pitt et al. 2007), which provides segment-level annotations and higher-quality recordings of 40 speakers interviewed in Columbus, Ohio. Each vowel appears in a variety of contexts, produced in several instances by many individuals, and the very nature of casual spontaneous speech suggests that we may deal with greater variability than we would otherwise get with careful articulation. This is particularly useful for the direct comparison between schwa and other vowels.

2 Methods and materials

2.1 Materials

We use the Buckeye Corpus of Conversational Speech (Pitt et al. 2007), which provides data collected at 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 (2020). 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 [bɛz], the procedure would align /b/ with [b], /æ/ with [ɛ], /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.[4] 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 remove schwas that Flemming (2009) treats as possibly different vowels. We also excluded second-to-last schwas in words that ended with *-ed* or *-es* to remove epenthetic vowels (e.g. in *roses* and *wanted*).[5] Finally, we removed all function words, using R's `tm` package's function `stopwords()` (Feinerer & Hornik 2015). The resulting number of tokens for each vowel, as well as the three most frequent words are found in Table 1.

We used FAVE (Rosenfelder et al. 2014) to get mid-vowel values. The procedure entailed providing the program information about the gender of the speakers in the corpus, and making it process schwa as if it were another stressed vowel.

We attempted to decide whether to transform the formant values in our models, so we compared (i) no transformation, (ii) log transformation, and (iii) mel-transformation of the formant values. The best method was defined as the method that would yield the highest consistency when using the vowel iden-

[4]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.

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

tity as a predictor. We therefore tried to find which method would yield the highest R^2 for a model that predicts formant values (transformed or not) using vowel identity, F0 mean, and log F0 mean as the only three predictors, and applying no other grouping.

For F1 values, log-transforming formant values led to higher consistency (R^2=0.55) compared to other methods. For F2 values, log-transforming formant values led to slightly worse consistency compared to not transforming formant values (R^2=0.525 compared to R^2=0.533). The best methods remained the same even when we added further controls, such as the identity of the previous and following segments.

Using similar principles, we determined which summary function and transformation of F0 to use, with speaker identity as the sole predictor, and applying no other grouping, given that speakers' F0 values are extremely consistent (Cohen Priva & Sanker 2018; Cohen Priva & Sanker 2020). We considered the following summary statistics: mean, geometric mean, median, and mode, as well as the following transformations: identity (no change), log, and mel. Log mean F0 yielded the highest consistency (R^2=0.61).

Table 1: The number of tokens and the most frequent words for all the monophthongs used in the Buckeye corpus. Word frequencies are listed by the number of times the word was used without the deletion of the vowel in question.

Vowel	Count	Top words	IPA
/ɑ/	5595	lot, got, probably	lɑt, gɑt, pɹɑbəbli
/æ/	4311	back, actually, bad	bæk, æktʃuəli, bæd
/ə/	3669	people, columbus, away	pipəl, kəlʌmbəs, əweɪ
/ʌ/	4512	just, much, stuff	dʒʌst, mʌtʃ, stʌf
/ɔ/	3810	always, thought, talk	ɔlweɪz, θɔt, tɔk
/ɛ/	10110	well, get, said	wɛl, gɛt, sɛd
/ɪ/	6663	really, kids, years	ɹɪli, kɪdz, jɪɹz
/i/	5030	people, even, need	pipəl, ivɪn, nid
/ʊ/	1226	good, look, put	gʊd, lʊk, pʊt
/u/	2194	school, used, schools	skul, juzd, skulz
Total	47120		

2.2 Statistical methods

2.2.1 Modeling considerations

The acoustic properties of vowels are affected by preceding and following consonants and vowels (Stevens & House 1963; Lindblom 1963; Bates 1995), as well as by the physiology and individual tendencies of the speakers that produce them (Fabricius, Watt & Johnson 2009). We use control variables to remove variance that can be associated with such factors, and provide some input regarding the association between such variance and vowel duration.

There are challenges and advantages for corpus studies that our models need to tackle. The environments in which the vowels occur are greatly varied, and require a host of controls, as well as their interaction with the main variable of interest. Such situations are less common in lab settings, in which it is possible to keep the number of contrasting conditions small. Whatever modeling we choose, we will never be able to match with controls the small number of contrasts an experimental setting would have included. Why use corpora then? Because spontaneous speech production gives rise to more relaxed and less anticipated combinations than an experiment could devise, and because spontaneous speech naturally gives rise to speech tokens in varied speech rates, which this set of studies leverages. While corpus studies may suffer from more noisy data, they can compensate for that with substantially more data.

Several different approaches are applicable to modeling context in corpus studies. We use one approach in the main text, and complement it with a different approach in Section 3 in the supplementary materials. We describe the three approaches below.

- **Model context as random effects.** This approach makes it possible to study pooled effects (e.g. the common influence of duration on all variables), but makes the interpretation of non-pooled effects difficult. This approach is less practical when comparing the interaction of context with the fixed effects, as we do in Study 3, and we therefore do not use it.

- **Estimate model contexts using a model.** This approach creates point estimates for different contexts using their effect on some observed variable. This can be done by averaging observed values (e.g. the average first formant measurement in a vowel, when the preceding sound is /k/) or by using a regression model. The advantage of this approach is that the estimation stage yields a single continuous variable (“`context`”), which can then be used in subsequent regression models. This is the approach we use in the paper, and it is described below in section 2.2.2.

- **Model context using actual values.** This approach uses the actual observed variables of the context. For instance, rather than specify that the preceding vowel is /i/, provide the actual formant values of the preceding vowel as a control. The advantages of this approach is that the models can use actual performance to control for what the speaker did in the utterance. For instance, perhaps the preceding vowel was lexically specified as /i/, but the speaker produced an /ɪ/ instead. The disadvantage is that this approach is causally opaque, because the speakers' performance (e.g. their speech rate) is likely to affect both the target variable and the measured context variables, and the relationship between the underlying variable and the target variable can therefore falsely seem to follow from the actual measurements of context. Another downside is that there is often no usable observable variable. For instance, if the formant values of the preceding vowels are included in the regression, such values are missing when there is no preceding vowel, and the missing values need to be imputed. We use this approach in Section 3 in the supplementary materials.

2.2.2 Models outline

We used several contextual variables to estimate the expected effect of context on formant values. These include the following variables of interest and controls:

(1) The identity of the vowel (when contrasting vowels), as a fixed effect. We label this variable `vowel.id`.

(2) A continuous control for log mean F0 during vowel articulation. We label this variable `F0`. We include F0 because it has been found to affect vowel perception (Syrdal & Gopal 1986).

(3) The combined identity of the immediately preceding (`preceding1`) and immediately following (`following1`) segments (replaced by phrase-initial or phrase-final symbols if unavailable). We label this variable `preceding1_following1`. Given the number of possible contexts, we opted to model this effect as a random intercept. The random intercept models absolute changes in the formant value, without regard to vowel duration. This variable and the next four variables are meant to control for the effect of CONTEXT on vowel production (see Bates 1995), as explained above.

(4) The identity of the non-immediate preceding segment, if available, or otherwise a phrase-initial symbol. We label this variable `preceding2`. As with `preceding1_following1`, and for similar reasons, we modeled this as a random intercept. The random intercept models absolute changes in the formant value as affected by the non-immediate preceding segment.

(5) The identity of the non-immediate following segment, if available, or otherwise a phrase-final symbol. We label this variable `following2`. As with `preceding2`, and for similar reasons, we modeled

this as a random intercept. The random intercept models absolute changes in the formant value as affected by the non-immediate following segment.

(6) The identity of the preceding vowel, if available, or otherwise a “missing” symbol. This variable is modeled using a random intercept, which we label `preceding.vowel`.

(7) The identity of the following vowel, if available, or otherwise a “missing” symbol. This variable is modeled using a random intercept, which we label `following.vowel`.

(8) The identity of the word, as a random intercept. The random intercept models absolute changes that may follow from different words having particular acoustic targets (cf. Pierrehumbert 2001). It may also account for cases in which some combinations of preceding and following contexts are not captured by the cumulative effect of the preceding and following variables. We label this variable `word`.

(9) The identity of the speaker as a random intercept, and a random slope for vowel identity (when contrasting vowels). Both are meant to control for speakers having different performance for different vowels (Fabricius, Watt & Johnson 2009). We label this variable `speaker`.

All F0 and F1 measurements were log-transformed because this transformation yielded the highest consistency, as discussed above. Also aforementioned, all F2 measurements were not log-transformed. To ease model convergence and interpretation, all continuous variables were z-transformed as well.

We trained two mixed effects linear models, one for each formant, using vowel duration, vowel identity, their interaction, and F0 as the fixed predictors, and `preceding1_following1`, `preceding2`, `following2`, `preceding.vowel`, `following.vowel`, `word`, and `speaker` as random intercepts. Vowel duration was also added as a random slope to `speaker`. The models were fitted using `lme4` (Bates et al. 2015) to produce point estimates for the random effects. All the continuous variables were z-transformed. Crucially, the models did not include context-dependent random slopes for duration, which means that context was modeled as constant deviations along F1 and F2. This yields the formula in (1).

(1)
```
formant ~ 1
+ duration * vowel.id
+ F0
+ (1 | preceding1_following1)
```

```
+ (1 | preceding2)
+ (1 | following2)
+ (1 | preceding.vowel)
+ (1 | following.vowel)
+ (1 + vowel.id || speaker)
+ (1 | word)
```

Subsequently, we used the cumulative effect of `preceding1_following1`, `preceding2`, `following2`, `preceding.vowel`, and `following.vowel` to predict for each vowel token a single continuous variable that combines all the individual contextual variables, which we labeled `context`. This variable is supposed to predict how the vowel is expected to shift along the formant in question when found in the specific combination of contextual variables. The variable `context` therefore captures all the variance that the linear mixed effects models can explain, with the exception of any variance explained by the vowel itself and vowel duration.

To study the relationship between a given vowel's duration and its position in the acoustic space, we used a bayesian mixed effects linear regressions using R's brms package (Bürkner 2018), with the formant value (F1 or F2) as the predicted value. All continuous variables were z-transformed, and the priors we used were uninformative (`normal(0, 3)` for fixed effects, `normal(0, 2)` for the SD of random effects). We label the dependent variable `formant`.

Our main variable of interest is vowel duration, which we label `duration`. A non-zero coefficient (whose credible interval does not include zero) for vowel duration would suggest that on average the vowel's position changes in particular ways and in particular directions when coarticulatory pressures change (following Moon & Lindblom 1994). The interaction terms between vowel duration and the context variable is also of interest, as it indicates whether the effect of other variables changes when vowel duration changes, as explained in detail in the individual studies.

The basic model in Studies 1–2 is provided in R' lme4 syntax (R Core Team 2021; Bates et al. 2015) in (2).

(2)
```
formant ~ 1
+ duration * context
+ (1 + duration | speaker)
+ (1 | word)
+ F0
```

In Study 3 we compare schwa to other vowels directly. This leads to including `vowel.id` as a predictor, with schwa as the base level. We also include the two-way interactions between `vowel.id` and `duration`, between `vowel.id` and `context`, and the three-way interaction among all three, leading to the formula in (3). The predicted directions of the interaction terms are explained in detail in Study 3.

(3)
```
formant ~ 1
+ vowel.id * duration * context
+ (1 + vowel.id + duration | speaker)
+ (1 | word)
+ F0
```

3 Studies

3.1 Study 1: The correlation between schwa's formants and its duration

3.1.1 Introduction

The goal of Study 1 is to explore the first type of evidence detailed in Section 1.3, and repeated here. If schwa does not have an acoustic target and the property of being a vowel is satisfied by any surface vowel, then its position would be affected by the context in which it appears, and by virtue of being a vowel, rather than by its duration. After factoring out the contributions of particular contexts and the possible variable effects that impact the location of schwa in the acoustic space, we try to identify whether durational changes in schwa are accompanied by movement within this space. We applied the single-vowel model described in 2.2 to schwa's F1 and F2 values.

If we detect that schwa's formants correlate with its duration, it would suggest that speakers' performance follows not only from contextual effects. If schwa would not move in either formant it would mean that its acoustic target is not different from the one dictated by context. As mentioned above, it would not entail that schwa does not have a target, but if it does, it would be indistinguishable from having no target.

Context is expected to be strongly correlated with schwa's formants, given that schwa has been found to be greatly affected by the context in which it occurs (Recasens 1991; Bates 1995; Flege 1988). If schwa does not have a target, we would expect the interaction term between context and duration not to differ from zero: The effect of context on schwa's formants should not differ when its duration is shorter. A negative interaction term would mean that longer-duration schwa is less likely to be affected by its context, a finding which would directly contradict the idea that schwa's position is dictated exclusively by its context. A positive interaction term is not expected, but if found it would suggest that the effect of context *increases* when schwa's duration is longer. This would be unexpected given our simplifying assumptions, but is possible e.g. if schwa's longer duration follows from slower speech, in which schwa's context may reach more extreme articulation.

3.1.2 Methods and materials

For all tokens that had underlying schwa in the Buckeye corpus (see Section 2.1 above), we used the formula in (2) to determine positions in the acoustic space with respect to duration. We fit the model twice, once with log-transformed F1 as the dependent variable, and once with F2 as the dependent variable.

3.1.3 Results

The full results of the revised Study 1 are in Table 2 for F1, and in Table 3 for F2.

Longer-duration schwas were associated with greater F1 values (Estimate=0.17, 95% CrI=[0.12–0.23]. This is consistent with schwa being influenced by having a target (TARGET), or if the effect of being a vowel (VOWEL) is modulated by vowel duration, and unexpected otherwise. The contextual effects explain much of the variance in F1 (Estimate=0.36, 95% CrI=[0.33–0.39]). Since all continuous variables were z-transformed, the size of the coefficients for duration and context can be compared directly, and it shows that the effect of context on schwa's position is greater than the effect of duration in determining schwa's position in F1. The interaction between duration and context was small and negative (Estimate=-0.036, 95% CrI=[-0.062–-0.01]). A negative interaction term is not consistent with the targetless schwa hypothesis, in which we expect the effect of context to remain the same regardless of schwa's duration. We see instead that the effect of context diminishes when vowel duration is longer.

For F2, longer-duration schwas were not associated with higher or lower F2 values (Estimate=0.021, 95% CrI=[-0.0033–0.046]. The contextual effects were very high (Estimate=0.75, 95% CrI=[0.72–0.78]), which suggests that schwa's F2 values are highly influenced by the context in which schwa appears. The interaction term between duration and context was very small but *positive* (Estimate=0.03, 95% CrI=[0.011–0.049]), which means that longer-duration schwas were more susceptible to the effect of context than shorter duration schwas. This is not predicted by our simplifying assumptions, but could be explained if schwa's duration indicates to what extent its context is hyperarticulated, e.g. in fast or slow speech. For instance, if a long-duration schwa occurs mostly in slow speech, then perhaps the context in which it appears is hyperarticulated, which could result in the small but positive interaction.

Table 2: Study 1 schwa F1 model, using a single context variable

	Estimate	2.5% CrI	97.5% CrI
Intercept	-0.02	-0.20	0.16
Duration	0.17	0.12	0.23
Context	0.36	0.33	0.39
F0	0.02	-0.01	0.06
Duration : context	-0.04	-0.06	-0.01

Table 3: Study 1 Schwa F2 model, using a single context variable

	Estimate	2.5% CrI	97.5% CrI
Intercept	0.02	-0.089	0.14
Duration	0.02	-0.003	0.05
Context	0.75	0.724	0.78
F0	0.05	0.022	0.08
Duration : context	0.03	0.011	0.05

A reviewer on a previous version of this manuscript was concerned that the results for F1 may follow from the over-representation of F1-lowering contexts in the data. That is, the observed effect of `context` is actually due to the majority of contexts having F1-lowering effects, rather than a true effect of `context`. We therefore wanted to see whether the coefficient value for duration, when combined with the context-specific random slope for preceding and following context, was still expected to be positive. Consequently, we retrained the F1 model using the `brms` package (Bürkner 2018), but replaced the single context variable with random intercepts and slopes for the immediately preceding and following contexts, and the preceding and following vowels. We also omitted the slopes for the non-immediate context to facilitate interpretation. This yielded the formula in (4). We then extracted the posterior samples for the coefficient and the random slopes.

(4)
```
formant ~ 1
+ duration
+ (1 | preceding1) + (0 + duration | preceding1)
+ (1 | following1) + (0 + duration | following1)
+ (1 | preceding.vowel) + (0 + duration | preceding.vowel)
+ (1 | following.vowel) + (0 + duration | following.vowel)
+ (1 + duration | speaker)
+ (1 | word)
+ F0
```

Figure 1 illustrates the density plots of the combined samples for the coefficient and slope for the ten most common following contexts and the ten most common preceding contexts. Though some of the probability mass falls below zero for a few contexts, the bulk of the density for most contexts lies above zero, which means that schwa did not shift in the opposite direction when its duration was short in any of the frequent contexts.

3.1.4 Discussion

The results are consistent with schwa having a target (TARGET), or that vowel duration modulates the property being a vowel (VOWEL), or both. The reason is that if schwa is targetless and vowel duration does not modulate being a vowel, we would expect all schwas in all contexts to simply represent the constraints imposed by the context in which they appear (CONTEXT), and we would also not predict that the effect of context would diminish in long durations, as we find for F1.

In contrast, the results for F2 do not seem to provide much evidence in favor of schwa having a target or that being a vowel is related to a property that affects F2. This means that perhaps schwa does not have a target in F2, being a vowel has no effect on F2, or schwa just happens to have a target that is consistent with what context alone would dictate. However, the effect of context on F2 is very large, which is less consistent with schwa having a target in F2.

The results are consistent with Browman & Goldstein (1994), who found that the articulation of schwa involves a component that is not predicted by context alone, and with Kondo (1994), who argued for a difference between schwa's F1 and F2, such that the former seems to deviate from positions dictated solely by context, and the latter does not. Schwa is associated with higher F1 values when its duration is longer, as we would expect of a mid-vowel (Ladefoged & Johnson 2010). Since higher F1 positions are found where

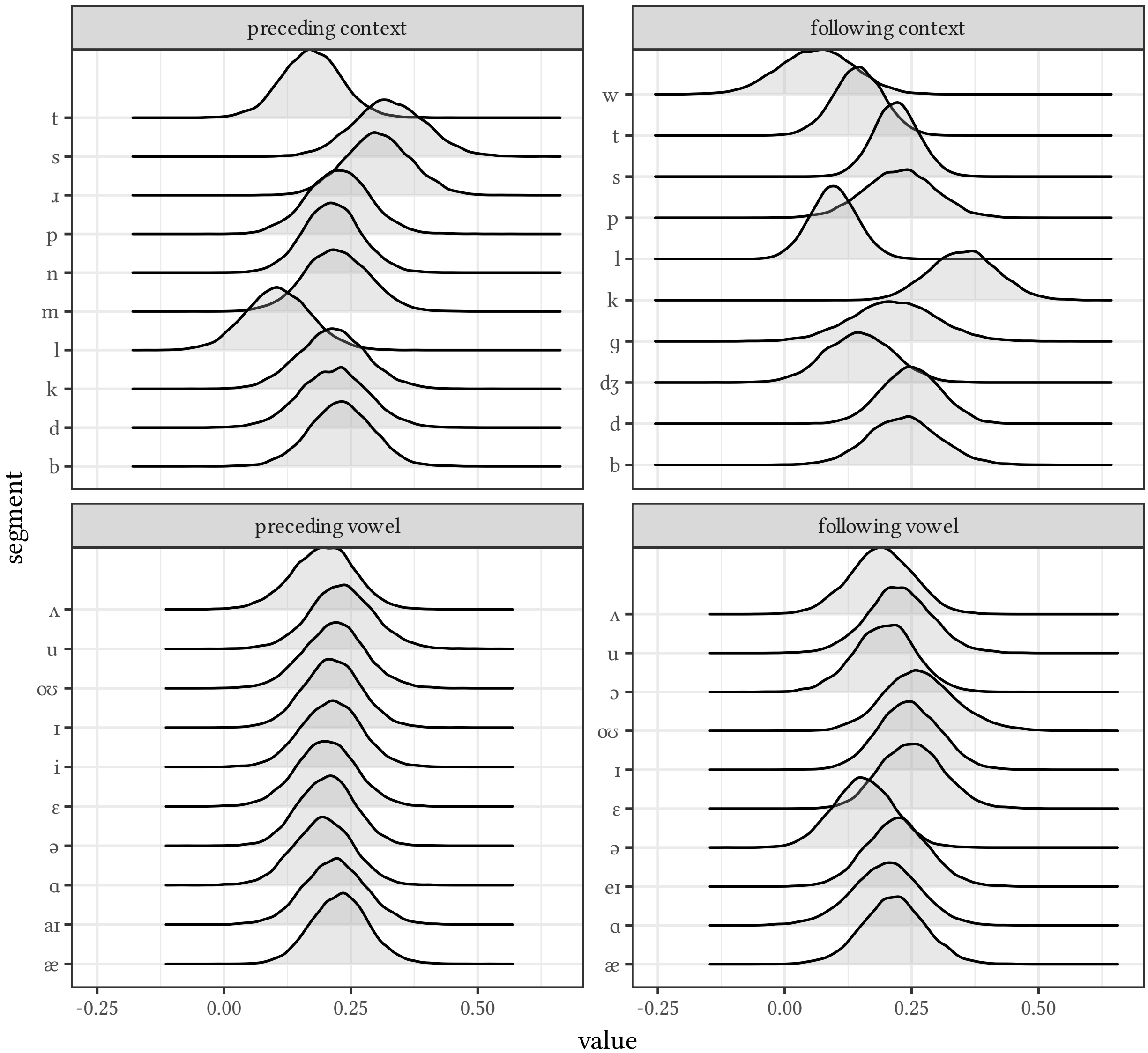

Figure 1: Density plots of the combined samples for the vowel duration coefficient and the context-specific random slope for the ten most common following and preceding contexts. Similarly, the most frequent preceding and following vowels (either schwas or vowels with primary stress).

contextual coarticulatory pressures are assumed to not be as strong, it would suggest that schwa's position in the F1 space is not dictated solely by contextual coarticulatory pressures, but by having a target in F1, or by having a greater jaw or tongue displacement when its duration is long than when it is short for some other reason.

What could that reason be? Previous accounts group the property of being a vowel with contextual effects (Lindblom 1963; Moon & Lindblom 1994), but other possibilities exist. For instance, if "being a vowel" means a more open vocal tract regardless of vowel identity per se (Flemming 2009: 84), then the results presented above could be interpreted as schwa having more prototypical vowel production when its duration is longer than when its duration is short, rather than having a target in F1. The results do not provide evidence that could disentangle schwa having a target in F1 and the manifestation of abstract vocalic property.

Though schwa is not an epenthetic vowel in the contexts we studied, a higher-than-mid position resembles the inserted vowels that American English speakers insert in non-native sequences (Davidson 2006), and the cross-linguistic tendency for epenthetic vowels to be high (Kitto & Lacy 1999). However, these observations suggest that the minimal requirement to satisfy the property of being a vowel (VOWEL) does not necessitate the more open vocal tract observed for longer-duration schwas.

3.2 Study 2: F1 values for reduced-duration vowels

3.2.1 Introduction

In Study 1, we observed that shorter duration was correlated with lower F1 values for schwa. It is clear that there is a pressure for low vowels to shift to lower F1 values when their duration gets shorter, but is there a clear reason to assume that the same would hold true for (other) mid and high vowels? Lindblom (1963) found that this tendency stops at about 375 Hz (higher than lax high vowels, but not as high as /i/). This suggests that vowel neutralization due to reduced duration should not head toward the mid-central position schwa occupies when its duration is longer, but rather toward the position schwa occupies when its duration is short. The goal of this study is to extend Study 1 to other vowels, in order to test this prediction.

3.2.2 Methods and materials

We repeated the method used in Study 1 for all non-diphthongs in American English. We did not collapse /ɑ/ and /ɔ/, as Labov, Ash & Boberg (2005: 64) show that they are at least partially distinct in Columbus, even though they greatly overlap in the Buckeye corpus.[6] Using the same formula used in Study 1, we then measured which vowels had a correlation between shorter duration and lower F1 values.

Since all the vowels except Schwa have an acoustic target, we are not interested in the degree to which the coefficient for vowel duration deviates from zero. Rather, we are interested in seeing the overall change in vowel space as vowel duration is reduced.

For `context`, we anticipate a positive coefficient for all vowels, though likely smaller than schwa's, because schwa is expected to be more susceptible to the effects of context, as discussed above. Following the results of Study 1, the expectations for the interaction term between duration and context are that it would be negative for all vowels, signifying that the effect of context diminishes for longer-duration vowels. However, Study 1 has already provided an exception to that expectation.

3.2.3 Results and discussion

The results for the relationship between vowel duration and F1 values for all vowels can be found in Table 4, and for the relationship between vowel duration and F2 in Table 5.

For F1, all vowels except high tense ones shifted to lower F1 values when their durations were shorter, just as schwa did. Lax high vowels also had lower F1 values when their duration was shorter, consistent with Lindblom (1963). For /u/ there was no correlation between duration and F1, which suggests that the acoustic position of /u/ is close to the position which coarticulation would predict. This too is consistent with Lindblom (1963). For /i/ the relationship reversed, with longer duration being associated with lower F1 values. We are not sure why this relationship holds. Perhaps coarticulatory pressures do not lead vowels to have a position similar higher than the one associated with /u/.

All vowels' F1 values were correlated with context, as anticipated, and for most vowels there was a negative

[6] It seems that /ɑ/ and /ɔ/ are not completely merged in the Buckeye corpus, even though they have substantial overlap, as verified using a mixed effects linear regression in which we include data from both vowels, and used vowel identity, log vowel duration, and their interaction to predict F1 values (using log(F1) ~ vowel.id * log(duration) + (1 + vowel.id | speaker) + (1 | word)). The model yielded small but highly significant differences for all three variables. /ɔ/ had lower F1 values (β=-0.34, SE=0.02, df=1268.9, t=-15.94, p<.001).

coefficient for the interaction term between duration and context, which means that the effect of context was more pronounced when vowel duration was shorter. There were two exceptions for /ɔ/ and /ʊ/, for which the interaction was small (relative to the main effect of context) but *positive*, as was the case with schwa and F2, which suggests that for these two vowels the effect of context is less pronounced when their duration is shorter. This is unexpected given our assumptions, but could follow from the smaller acoustic space associated with shorter vowels, as discussed above.

For F2, back vowels (including /ʌ/) had higher F2 when their duration was shorter), and front vowels (except /æ/, whose credible interval included zero) had lower F2 values when their duration was shorter. All vowels were highly influenced by context, and the interaction term between duration and context was mostly negative, with the exception of schwa, which was positive (as mentioned above), and /u/ and /æ/, whose credible interval includes zero. This means that for those two vowels there is no evidence that longer duration would result in a smaller effect for context.

Table 4: Study 2 duration estimates for F1

Vowel	Estimate	2.5% CrI	97.5% CrI
ə	0.173	0.122	0.226
ʌ	0.196	0.151	0.241
ɑ	0.189	0.157	0.221
ɔ	0.061	0.032	0.090
ʊ	0.239	0.166	0.312
u	0.026	-0.013	0.064
i	-0.089	-0.118	-0.061
ɪ	0.112	0.087	0.135
ɛ	0.153	0.121	0.185
æ	0.159	0.121	0.199

Table 5: Study 2 duration estimates for F2

Vowel	Estimate	2.5% CrI	97.5% CrI
ə	0.021	-0.003	0.046
ʌ	-0.074	-0.106	-0.043
ɑ	-0.123	-0.149	-0.096
ɔ	-0.142	-0.181	-0.103
ʊ	-0.130	-0.182	-0.077
u	-0.061	-0.093	-0.031
i	0.168	0.139	0.196
ɪ	0.082	0.051	0.113
ɛ	0.037	0.013	0.062
æ	0.048	-0.002	0.095

Table 6: Study 2 point estimate coefficients for F1, by vowel. Values whose middle credible interval 95% range has the same sign are marked in bold.

	ə	ʌ	ɑ	ɔ	ʊ	u	i	ɪ	ɛ	æ
Intercept	-0.02	**0.33**	0.06	-0.01	0.02	-0.04	-0.03	0.07	0.13	0.06
Duration	**0.17**	**0.20**	**0.19**	**0.06**	**0.24**	0.03	**-0.09**	**0.11**	**0.15**	**0.16**
Context	**0.36**	**0.20**	**0.25**	**0.63**	**0.20**	**0.33**	**0.31**	**0.43**	**0.35**	**0.23**
F0	0.02	**0.06**	**0.07**	**0.05**	**0.12**	0.04	**0.08**	**0.03**	**0.07**	**0.11**
Context : duration	**-0.04**	**-0.02**	-0.01	**0.08**	**0.08**	-0.02	**-0.02**	**-0.02**	0.00	**-0.06**

Table 7: Study 2 point estimate coefficients for F2, by vowel. Values whose middle credible interval 95% range has the same sign are marked in bold.

	ə	ʌ	ɑ	ɔ	ʊ	u	i	ɪ	ɛ	æ
Intercept	0.02	**-0.25**	-0.15	**-0.25**	-0.15	**0.12**	0.07	0.08	-0.04	0.04
Duration	0.02	**-0.07**	**-0.12**	**-0.14**	**-0.13**	**-0.06**	**0.17**	**0.08**	**0.04**	0.05
Context	**0.75**	**0.38**	**0.37**	**0.35**	**0.72**	**0.67**	**0.31**	**0.57**	**0.55**	**0.26**
F0	**0.05**	**0.05**	**0.03**	0.01	0.00	**0.04**	**0.10**	**0.11**	**0.08**	**0.13**
Context : duration	**0.03**	**-0.04**	**-0.07**	**-0.07**	**-0.06**	-0.02	**-0.05**	**-0.07**	**-0.03**	-0.00

The results help put Study 1 in context. Schwa's association between shorter duration and lower F1 values can be seen as the usual trend for all vowels except /i/ and /u/. Shorter duration seems to involve a reduction in F1 values, with the exception of tense high vowels. We proposed above that the results for schwa could follow from either having a target or from a modulation of the property of being a vowel. Under this interpretation, it is possible that the lower F1 values for the other vowels could follow from

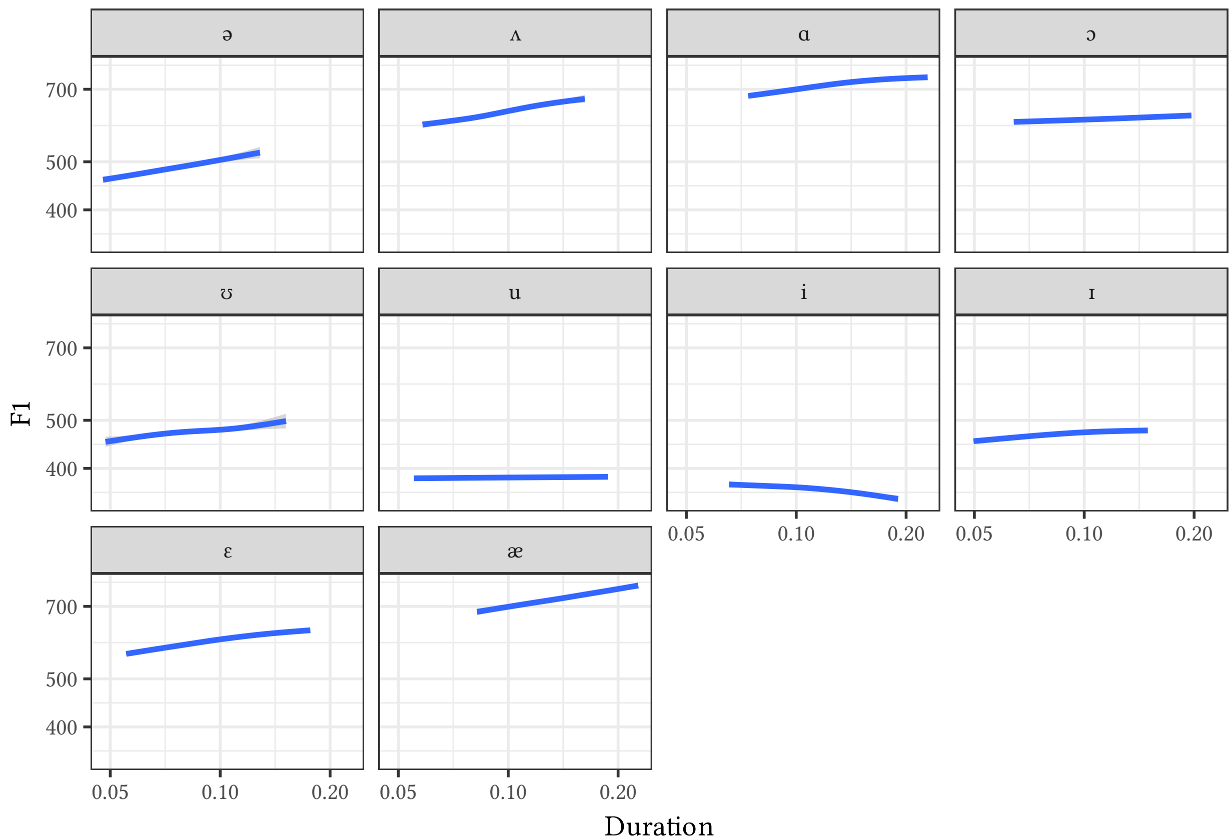

Figure 2: A visualization of the relationship between vowel duration and F1 values. Both duration and F1 are the residual values after the effects of speaker, word, and phonological context have been controlled for using random intercepts. Both axes are log-scaled, and only the middle 95% of duration values are shown. The line is drawn using a smoothing spline. The gray zones (visible only for /ʊ/) are the 95% confidence intervals of the smoothing model.

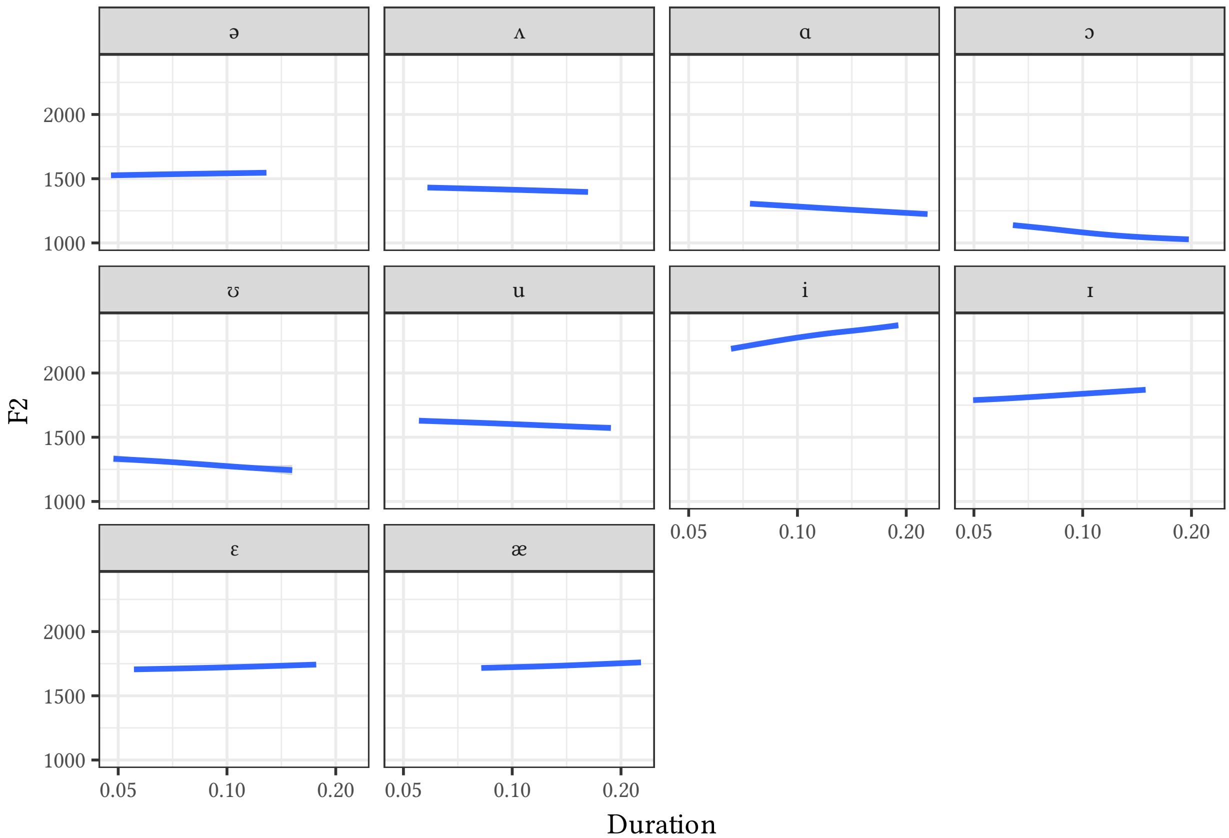

Figure 3: A visualization of the relationship between vowel duration and F2 values. Both duration and F1 are the residual values after the effects of speaker, word, and phonological context have been controlled for using random intercepts. The x-axis is log-scaled, and only the middle 95% of duration values are shown. The line is drawn using a smoothing spline. The gray zones (hardly visible) are the 95% confidence intervals of the smoothing model.

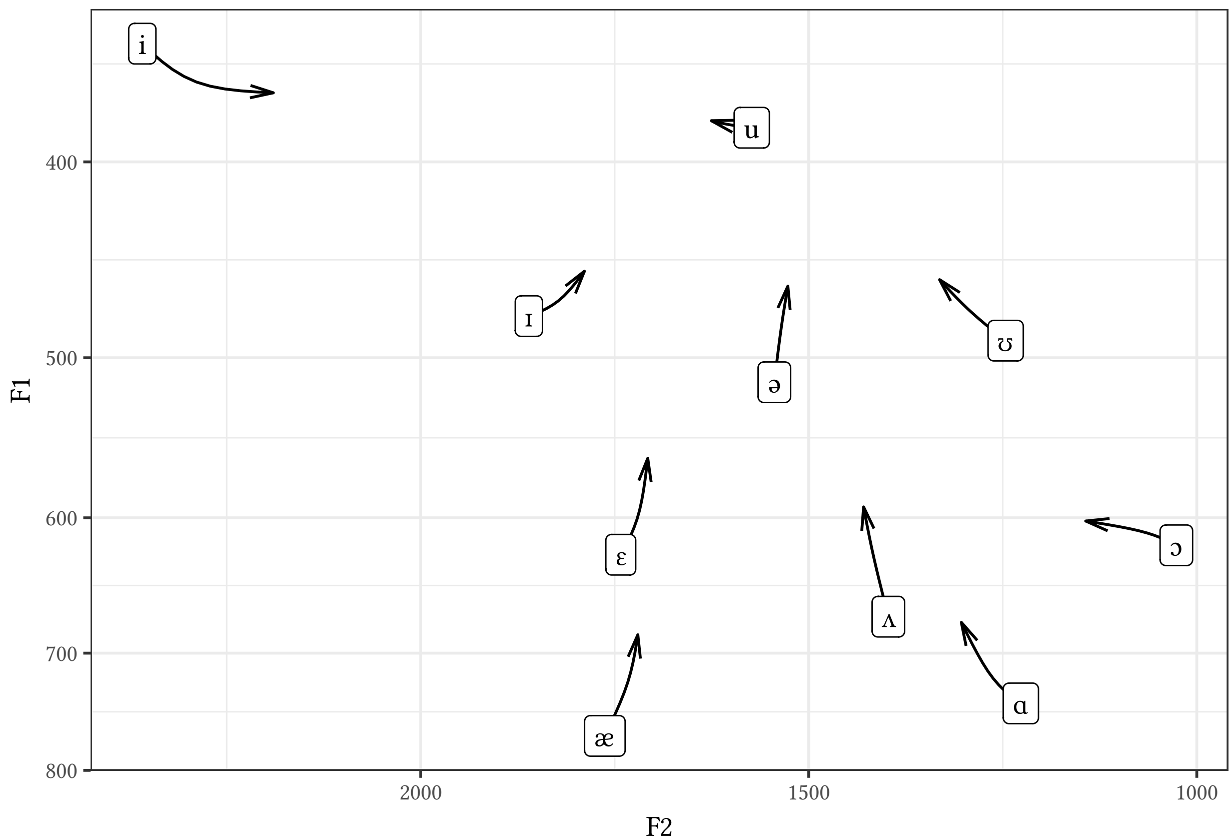

Figure 4: A visualization of the movement of individual vowels in the F1/F2 space as correlated with duration, combining Figures 2 and 3. Duration, F1, and F2 are the residual values after the effects of speaker, word, and phonological context have been controlled for using random intercepts. Only the middle 95% of duration values are shown.

having less time to reach the vowels' intended target, or from having less time to express the vowel's more abstract vocalic property when their duration is short.

It is interesting to note that the interaction term between vowel duration and context was not negative in a number of cases, as it was for schwa's F2 in Study 1. This means that for those vowels too, context was not modulated by duration, even though no account argues that other vowels lack an acoustic target.

With respect to the phonological question regarding whether vowels neutralize to schwa or assimilate to their context when their duration is shorter, the results of Study 2 seem to offer a more nuanced but different answer. We observe that vowels overall do assimilate more when their duration is shorter, though there are a few exceptions to this rule. However, they *also* neutralize toward a high central position, because all except the tense high vowels shift to lower F1 values independently from context when their duration is short, and all except /u/ and /æ/ centralize (/u/ has fairly central F2 values even when its duration is long). Crucially though, the movement to lower F1 values does not make vowels shift toward the center of the vowel space, because lax high vowels shift toward lower F1 values too. However, the supposed neutral position is where we do find long duration schwa vowels after we remove the effect of individual contexts, and where cross-linguistically most schwa vowels are found (Recasens 2021: fig. 1).

3.3 Study 3: Pairwise comparison between schwa and other vowels

3.3.1 Introduction

In Study 1 and Study 2 we found that schwa behaves much like other vowels, in the sense that it shifts to lower F1 values when its duration is shorter. We argue that this property is consistent with either having a target, or with duration modulating a stipulated abstract vocalic property, which is associated with more open production. The results were not consistent with the argument that schwa lacks a target if the abstract vocalic property (VOWEL) is satisfied by any vowel. For F2 the results were more ambiguous, because schwa did not shift in F2 when its duration was shorter. This effect could follow from schwa lacking a target in F2, but also if schwa does have a target, but that target is already compatible with extreme coarticulation pressures. There was no correlation between duration and F2 for two other vowels too, /u/ and /æ/.

In Study 3 we examine a different kind of evidence for schwa having or lacking a target. Several researchers proposed that schwa does not have a target because it is affected by the context in which it appears to a greater extent (Recasens 1991; Bates 1995; Flege 1988). As outlined in Section 1.3, this comparison is difficult because schwa has a shorter duration and can move more in the F1/F2 space relative to other vowels. The goal of this study is to see whether schwa is more variable than other vowels in the same context, even when duration is controlled for.

3.3.2 Methods and materials

We compared each American English non-rhotic monophthong in the Buckeye corpus to schwa. We determined vowel quality by the underlying (intended) representation rather than by the annotation used for that token, because we wanted to predict speakers' performance relative to an intended baseline. Since we are not interested in comparing the other vowels to one another, we modeled each vowel in comparison to schwa directly, using the formula listed above in (3) and repeated here in (5).

(5)
```
formant ~ 1
+ vowel.id * duration * context
+ (1 + vowel.id + duration | speaker)
+ (1 | word)
+ F0
```

This study focuses on the difference between schwa and other vowels. Therefore, the main variables of interest are the interaction terms between vowel identity and other variables. These include the interaction with `context` and the interaction with `duration`. The interaction term with context should be significantly negative if the contrasting vowel is not affected by context to the same extent as schwa. The interaction term with duration is only interesting in the context of F1 (schwa was not affected by duration in F2). The interaction term should be negative if duration has a greater effect on schwa's F1 relative to other vowels.

The three-way interaction between vowel identity, duration, and context could also be illuminating. The two-way interactions between duration and each of the contextual variables model the moderation of the contextual effect by duration for schwa. In Study 1 that coefficient was negative in the F1 model, suggesting that schwa is affected by context to a lesser extent when its duration is longer. A positive coefficient would suggest that the contrasting vowels' dependence on context is not modulated by duration to the same extent as schwa's, while a negative coefficient would suggest that the contrasting vowel's dependence on context is modulated to a greater extent by duration than schwa's (in the expected direction). The two-way interaction was positive for F2 in Study 2, which was unexpected, and we would expect the contrasting vowels to have negative three-way interaction terms if their behavior is more typical.

All the variables that do not depend on vowel identity model the same variance in all the models (because schwa is the default level), and are expected to be similar to one another and to those in the Study 1 model. However, they are not expected to be identical to one another because the random effects fit different data in the different models due to the inclusion of an additional vowel, and because they were z-transformed in each data set separately. We therefore provide these variables in the model summaries, but we will not discuss them further.

3.3.3 Results

The full models are listed in Section 2 in the supplementary materials. Table 8 provides a summary of the coefficients and their estimates for F1, and Table 9 provides a summary of the coefficients and their estimates for F2.

Table 8: Study 3 point estimate coefficients for F1, by vowel. Values whose middle CrI 95% range has the same sign are marked in bold. The row labels of variables of interest are also marked in bold. The full models can be found in Section 2 in the supplementary materials.

	ʌ	ɑ	ɔ	ʊ	u	i	ɪ	ɛ	æ
Intercept	**-0.39**	**-0.82**	**-0.39**	-0.01	**0.38**	**0.74**	0.05	**-0.63**	**-0.58**
F0	**0.04**	**0.04**	**0.03**	**0.04**	0.02	**0.04**	**0.02**	**0.05**	**0.05**
Duration	**0.19**	**0.22**	**0.23**	**0.21**	**0.21**	**0.19**	**0.23**	**0.22**	**0.20**
Vowel id	**0.99**	**1.37**	**0.81**	-0.01	**-0.96**	**-1.23**	-0.03	**0.96**	**1.13**
Context	**0.33**	**0.38**	**0.50**	**0.37**	**0.34**	**0.29**	**0.50**	**0.39**	**0.23**
Duration : vowel id	-0.02	**-0.09**	**-0.15**	**-0.07**	**-0.20**	**-0.26**	**-0.14**	**-0.08**	-0.05
context : duration	**-0.04**	**-0.06**	**-0.07**	**-0.04**	**-0.04**	**-0.04**	**-0.05**	**-0.05**	**-0.04**
context : vowel id	**-0.12**	**-0.22**	-0.01	**-0.22**	**-0.10**	-0.03	**-0.14**	**-0.08**	-0.02
context : duration : vowel id	0.01	**0.05**	**0.13**	**0.09**	0.03	0.02	0.03	**0.05**	-0.02

Table 9: Study 3 point estimate coefficients for F2, by vowel. Values whose middle 95% CrI range has the same sign are marked in bold. The row labels of variables of interest are also marked in bold. The full models can be found in Section 2 in the supplementary materials.

	ʌ	ɑ	ɔ	ʊ	u	i	ɪ	ɛ	æ
Intercept	**0.15**	**0.47**	**0.59**	**0.18**	0.06	**-0.74**	**-0.40**	**-0.41**	**-0.26**
F0	**0.04**	**0.03**	**0.02**	**0.03**	**0.04**	**0.05**	**0.08**	**0.07**	**0.07**
Duration	**0.03**	**0.04**	0.03	0.03	0.02	0.02	**0.03**	**0.03**	0.04
Vowel id	**-0.42**	**-0.82**	**-1.25**	**-0.67**	0.04	**1.31**	**0.72**	**0.54**	**0.54**
Context	**0.70**	**0.85**	**0.65**	**0.80**	**0.73**	**0.48**	**0.72**	**0.81**	**0.73**
Duration : vowel id	**-0.09**	**-0.12**	**-0.11**	**-0.11**	**-0.09**	**0.10**	**0.04**	0.00	0.00
context : duration	**0.03**	**0.05**	**0.03**	**0.04**	**0.04**	**0.02**	**0.03**	**0.04**	**0.04**
context : vowel id	**-0.23**	**-0.54**	**-0.38**	**-0.12**	-0.05	**-0.26**	**-0.19**	**-0.29**	**-0.46**
context : duration : vowel id	**-0.07**	**-0.10**	**-0.08**	**-0.08**	**-0.05**	**-0.06**	**-0.10**	**-0.06**	**-0.04**

The interaction terms between duration and vowel identity in the F1 models was negative for all vowels except /ʌ/ and /æ/. Though we are not interested in the difference between schwa and /i/ and /u/, which seem to respond to duration differently, it is striking that schwa differs from other mid and lax high vowels. Whatever causes schwa to shift in F1 when its duration changes makes it behave less like vowels which we assume have a target in F1.

For F1, some of the interaction terms between vowel identity and context included zero, but most were negative. This suggests that previous authors' observations that schwa is affected by context more than other vowels are replicated in the corpus data, even when vowel duration is controlled for. The three exceptions were the low vowels /ɔ/, /ɑ/, and /æ/. For F2, the results were more uniform such that almost all the coefficients (except for /u/) were negative. This means that schwa's F2 values are determined by context more than other vowels.

The picture for the three-way interactions is less informative. The two-way interactions between duration and each of the contextual variables model the moderation of the contextual effect by duration for schwa. As explained in Section 3.3.1, For F1 a negative coefficient for would suggest that the contrasting vowels' dependence on context is modulated to a greater extent than schwa's, while a positive coefficient would mean that duration does not have as strong of an effect on their dependence on context. We find that in the F1 models there were four positive interactions (/ɑ/, /ɔ/, /ʊ/, and /ɛ/), and the rest included zero in the credible interval. We conclude that the effect of context on the contrasting vowels is not modulated to the same extent by duration as it is for schwa, but the evidence for it is less conclusive than for the two-way interactions.

As discussed in Section 3.3.1, for F2 the expectation is that the contrasting vowels would have negative three-way interaction terms, and this is indeed the case, which means that schwa is unique in expressing stronger contextual effects when its duration is longer.

3.3.4 Discussion

The interaction terms between duration and vowel identity and between context and vowel identity suggest that whatever causes schwa to have greater F1 values when its duration is longer is less resistant to the effects of context and short duration relative to vowels that are assumed to have a target.

The results for the interaction between vowel identity and context replicate previous findings that argued that schwa does not have a target based on its greater dependence on context. Unlike the majority of previous work, we can assert that this difference is not due to schwa's shorter duration. Schwa is more susceptible to context even when duration is controlled for. Schwa is indeed not stressed, while all the contrasting vowels were stressed, leaving this aspect as one competing alternative explanation, namely that schwa is more sensitive to context because it is not stressed.

Finally, it is not possible to argue that schwa moves more in F1 relative to other vowels because its position is not high or low. This is because we find that several non-high non-low vowels shift less than schwa when their duration is short.

In sum, the results of Study 3 show that schwa is more dependent on both context and its duration compared to other vowels.

4 General discussion

4.1 Schwa's shift to higher F1 values

Study 1 showed that reduced duration is associated with lower F1 values for schwa. Though the effect was not as strong in every context, it was present in the majority of the most frequent contexts, both preceding and following. The effect of context diminished when schwa's duration was longer, and both findings point to some factor other than context that is involved in dictating schwa's acoustic position, and which has a greater effect when schwa's duration is longer. That other factor could be having an acoustic target, but our results support other possibilities as well. The position in which we find schwa when its duration is long is similar to the cross-linguistic schwa position at the center of the acoustic space (Recasens 2021), while the position in which we find schwa when its duration is shorter is similar to where speakers of American English produce minimal epenthetic vowels (Davidson 2006). We did not observe a correlation between duration and F2.

These findings are consistent with two possibilities. The first is that schwa does have a target in F1 (but possibly not in F2, Kondo 1994), which is compromised when its duration is shorter. If schwa does have a target in F1, that target must be weaker than other vowels' targets, because speakers give up on reaching it more readily than they do with other vowels, as shown in Study 3. The need for variable target strength is necessary on other grounds as well. For instance, Cohen Priva (2015) shows that different stops in American English have different propensity to delete. Another possibility is that schwa does not have a target, but the property of being a vowel is responsible for the effect of having higher F1 values when schwa's duration is longer (Flemming 2009). In other words, schwa's draw to higher F1 values when its duration is longer is not unique to schwa, but stems from a property common to all vowels. The main challenge to this possibility is that the exclusion of tense high vowels from this generalization needs to be justified, as shown in Study 2. One obvious justification is that having a close vocal tract is one of the key defining features of tense high vowels.

Both accounts therefore require adjusting the weight of one of the attractors for some of the vowels. Phonological theories with a hierarchy or weighting of preferences (e.g. Prince & Smolensky 1993; Flemming 2001) can make schwa's target less important to reach than other vowels', or exclude tense high vowels from the effect observed for all other vowels. Neither account can be rejected on these grounds. However, we believe that attributing schwa's position to properties common to all vowels except tense high ones requires fewer auxiliary assumptions, as would become evident in the following sections.

Regardless of the underlying cause, our findings provide further acoustic support to the idea that schwa's position is not dictated exclusively by the context in which it occurs. Our findings do not mean that schwa would have any one target in acoustic space, but they do suggest that there are attractors that draw schwa to higher F1 values when its duration is longer.

4.2 Reduced vowel neutralization vs. assimilation

Do vowels assimilate or neutralize when their duration is shorter? Study 2 shows that most vowels assimilate to their context more when their duration is shorter than when it is longer, consistent with an assimilation approach. However, like schwa, all vowels except tense high ones shift toward lower F1 values when their duration is shorter, and all vowels except /u/ and /æ/ shift toward more central F2 values when their duration is shorter, consistent with a neutralization approach (Crystal 2008; Kondo 1994; Ladefoged & Johnson 2010: 97). However, the neutralization-like behavior is not neutralization to the position schwa vowels occupy in many different languages (Recasens 2021), as even lax high vowels shift to lower F1 values when their duration is short, not to the middle of the vowel space.

This would suggest that phonological and diachronic neutralization to mid-central positions involves an additional component that pushes the reduced vowel toward higher F1 values compared to the variable reduction patterns observed here. One possibility is that schwa repeatedly obtains specific targets in different languages, which happen to approximate the assumed neutral position, e.g. in response to dispersion pressures that push it away from other vowels (Flemming 2004). This gains support by the observation that the production of schwa in different languages seems to depend on the presence of other mid-vowels (Recasens 2021). Another possibility is that the property of being a vowel translates to a draw toward the neutral position, and then schwa need not have a target of its own, providing a more parsimonious explanation. These findings parallel our findings for schwa's synchronic draw to higher F1 values when its duration is longer.

That lax high vowels also shift to lower F1 values when their duration is short also serves as counterevidence to the possibility raised by Bergem (1994), that the mid central position is simply the average of the many different contexts in which schwa is articulated.

4.3 Schwa's sensitivity to context and duration

One of the main arguments for schwa's targetlessness is that it is more susceptible to contextual variability than other vowels (Bates 1995; Recasens 1991). We raised the possibility that schwa is simply shorter than other vowels and therefore more susceptible to the influence of its context, but Study 3 rules out this possibility. We compared schwa to other vowels, and found that schwa is never less susceptible to the effect of context relative to other vowels, and it is also more susceptible to the correlation between duration and

F1.

Such results could be taken as evidence that schwa lacks a target altogether, but as we discussed in 4.1, the shift in F1 is not consistent with schwa having no attractor to higher F1 values when its duration is longer. We proposed that the acoustic position of vowels that have a target is dictated by three factors: the target (TARGET), the context in which they occur (CONTEXT), and the property of being a vowel (VOWEL), which are weighted differently when coarticulatory pressures increase (a version of Moon & Lindblom 1994). The results are consistent with two possibilities. First, it is possible that schwa does have some TARGET which attracts it to higher F1 values when its duration is longer but speakers learn that the importance or weight of TARGET for schwa is lower than for other vowels, and more readily compromise it (Keating 1990; Flemming 2001). The other alternative, consistent with Flemming (2009: 88), is that schwa is not affected by TARGET at all, and the draw to higher F1 values is due to VOWEL. In this case, speakers do not try to reach higher F1 values to approximate any target in particular, they do the same thing they do for all other vowels except tense high ones, and since schwa does not have an actual target, the effect of the two remaining attractors is more pronounced than it is for other vowels. On these grounds too, we find the latter possibility as being more parsimonious.

4.4 Theoretical motivation for schwa having or lacking a target

Short duration vowels increasingly assimilate to their context, but also neutralize, such that they occupy a smaller acoustic space, and generally shift toward a higher and more central position (Flemming 2004, as well as Study 2 in this paper). This applies to lax high vowels too, which suggests that the cross-linguistic mode for schwa vowels is not motivated by vowel reduction directly. Some other factor, a repeatedly acquired target or some other factor must be responsible for finding schwa at these mid-central positions. The mid-central position is also where the average production of American English schwa is centered around when schwa's duration is longer.

Why would schwa acquire similar targets in different languages, and perhaps in American English too? Perhaps the more mid-vowel quality serves a perceptual role of distinguishing schwa from the lax high vowels it approximates when its duration is short, and provides evidence for its unstressed quality. If so, lower F1 articulations could highlight schwa's contrast relative to other vowels (Dispersion Theoretic motivation, Flemming 2004), which is consistent with the idea that so-called weak prosodic positions need to signal their contrast relative to so-called strong prosodic positions, attested similarly in consonant lenition (Kingston 2008; Katz 2016). One possible issue with that goal is that Erickson & Kawahara (2016) argue that more open vocal tract is related to higher prominence, which would imply that the more prototypical mid-central schwa is more prominent.

If schwa does not have a target, and its shift to higher F1 values follows from VOWEL rather than TARGET, then we may wonder why VOWEL would involve a more open vocal tract. Perhaps longer duration in the corpus follows from greater prominence, which translates (along the lines of Erickson & Kawahara 2016) to a more open vocal tract. In other words vowel reduction patterns alone may head toward a central high position, but if speakers produce such vowels in prosodically prominent contexts, they would not keep the vowel high (because it lacks a target), and produce a more mid vowel instead.

We ran post-hoc tests to see if the relationship between prominence and greater F1 values holds in our data, using the maximal intensity of a vowel token as a correlate of its prominence. We used a procedure similar to the one described above for Figure 4, and added log duration as a predictor, with intensity as the dependent variable. Duration and intensity were z-transformed by vowel and speaker. In the first set of models log duration was used to predict maximal intensity, and the two were positively correlated for all vowels. In the second set of studies duration and maximal intensity were used jointly (along the other controls) to predict changes in F1. The direction of the association between duration and F1 did not change for any vowel. Greater intensity was correlated with higher F1 values for all vowels except /u/ (but including /i/), even after controlling for duration. The post-hoc test does not reveal the causal structure of the positive correlation between intensity and F1 for almost all vowels, but it does mean that making vowels louder and increasing their F1 are causally linked, and that the effects of increased intensity and increased duration on F1 cannot be reduced to one another, even if the increase in F1 is related to an increase in prominence.

5 Conclusion

We addressed two inter-related questions, whether phonological vowel reduction is a neutralization or assimilation type process and whether American English schwa lacks an acoustic target. We used a novel approach, which relies on an observation by Lindblom (1963) and Moon & Lindblom (1994) that as vowels get shorter, they increasingly reflect coarticulatory pressures. We benefited from using the Buckeye corpus, a large collection of carefully annotated speech, to test schwa's response to variable duration, and contrast its variability with other vowels. We find that American English schwa seems to involve a higher-than-expected F1 value when its duration is long, paralleling previous articulatory studies. This effect is limited to F1, not F2, consistent with Kondo (1994). We also replicated existing accounts that show schwa's increased sensitivity to contextual effects relative to other vowels, which form the basis for the argument that schwa is targetless (Bates 1995).

As for vowel reduction, we find that American English vowels do exhibit greater assimilation to context when their duration is shorter, but they also seem to neutralize toward a central high position, not the position that schwa occupies cross-linguistically, and not the position toward which it shifts when its duration is long. Together, these results suggest that the diachronic and phonological reduction toward schwa necessitates an additional attractor, and we speculate that this attractor could follow from dispersion theoretic pressures, or from a pressure to increase vowel prominence in specific contexts.

One limitation of this set of studies is that they involve just one language, American English. In the future, we hope to replicate these findings in other languages, with different consonantal contexts and different vowel systems. We would like to make sure the results reflect the broad effect reduced duration has on vowel formants, and not only reflect what is typical for American English.

6 Acknowledgments

This paper was partially supported by NSF grant BCS-1829290 granted to the first author.

Chelsea Sanker provided extensive feedback on previous versions of this manuscript, and contributed significantly to ideas that appear here. Elinor Amit, Stefon Flego, and others provided substantial useful feedback. We thank the Associate Editor and the three reviewers for their feedback. All mistakes remain ours.

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Bates, Sally Alexandra Rosemary. 1995. *Towards a definition of schwa : An acoustic investigation of vowel reduction in English*. University of Edinburgh PhD thesis.

Beinum, Florien J. Koopmans-van. 1994. What's in a schwa? *Phonetica* 51(1-3). 68–79.

Bergem, Dick R. van. 1994. A model of coarticulatory effects on the schwa. *Speech Communication* 14(2). 143–162.

Bergem, Dick R. van & Florien J. Koopmans-van Beinum. 1989. Vowel reduction in natural speech. In *Eurospeech-1989*, 2285–2288.

Boersma, Paul & David Weenink. 2020. Praat: Doing phonetics by computer. Version 6.1.09. Computer program, available at http://www.praat.org/.

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

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10(1). 395–411.

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

Cohen Priva, Uriel & Emily Gleason. 2020. The causal structure of lenition: A case for the causal precedence of durational shortening. *Language* 96(2). 413–448. https://doi.org/10.1353/lan.2020.0025.

Cohen Priva, Uriel & Chelsea Sanker. 2018. Distinct behaviors in convergence across measures. In *Proceedings of the 40th annual conference of the cognitive science society*, 1518–1523. Austin, TX: Cognitive Science Society.

Cohen Priva, Uriel & Chelsea Sanker. 2020. Natural leaders: Some interlocutors elicit greater convergence across conversations and across characteristics. *Cognitive Science*. Wiley 44(10). https://doi.org/10.1111/cogs.12897.

Crystal, David. 2008. *A dictionary of linguistics and phonetics, sixth edition*. Blackwell Publishing Ltd.

Davidson, Lisa. 2005. Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics*.

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

Davidson, Lisa & Maureen Stone. 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.

Erickson, Donna & Shigeto Kawahara. 2016. Articulatory correlates of metrical structure: Studying jaw displacement patterns. *Linguistics Vanguard*. Walter de Gruyter GmbH 2(1). https://doi.org/10.1515/lingvan-2015-0025.

eSpeak NG development team. eSpeak NG text-to-speech. https://github.com/espeak-ng/espeak-ng.

Fabricius, Anne H., Dominic Watt & Daniel Ezra Johnson. 2009. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change*. Cambridge University Press (CUP) 21(3). 413–435. https://doi.org/10.1017/s0954394509990160. http://dx.doi.org/10.1017/S0954394509990160.

Feinerer, Ingo & Kurt Hornik. 2015. *tm: Text mining package*. http://CRAN.R-project.org/package=tm.

Flege, James Emil. 1988. Effects of speaking rate on tongue position and velocity of movement in vowel production. *The Journal of the Acoustical Society of America*. Acoustical Society of America (ASA) 84(3). 901–916. https://doi.org/10.1121/1.396659.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*. Cambridge University Press 18(1). 7–44.

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

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

Gick, Bryan. 2002. An x-ray investigation of pharyngeal constriction in American English schwa. *Phonetica*. Walter de Gruyter GmbH 59(1). 38–48. https://doi.org/10.1159/000056204.

Hall, Nancy. 2011. Vowel epenthesis. In *The blackwell companion to phonology*, 1576–1596. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444335262.wbctp0067. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781444335262.wbctp0067.

Hall, Nancy Elizabeth. 2003. *Gestures and segments: Vowel intrusion as overlap*. University of Massachusetts PhD thesis.

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

Hertz, Susan R. 1991. Streams, phones and transitions: Toward a new phonological and phonetic model of formant timing. *Journal of Phonetics*. Elsevier BV 19(1). 91–109. https://doi.org/10.1016/s0095-4470(19)30319-5.

Katz, Jonah. 2016. Lenition, perception and neutralisation. *Phonology*. Cambridge, UK: Cambridge University Press 33(1). 43–85. https://doi.org/10.1017/S0952675716000038.

Keating, Patricia A. 1985. Universal phonetics and the organization of grammars. In V. Fromkin (ed.), *Phonetic linguistics*, 115–132. Academic Press.

Keating, Patricia A. 1990. Phonetic representations in a generative grammar. *Journal of Phonetics*. Elsevier BV 18(3). 321–334. https://doi.org/10.1016/s0095-4470(19)30377-8. http://dx.doi.org/10.1016/S0095-4470(19)30377-8.

Kim, Kyumin & Alexei Kochetov. 2011. Phonology and phonetics of epenthetic vowels in loanwords: Experimental evidence from korean. *Lingua* 121(3). 511–532.

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

Kitto, Catherine & Paul de Lacy. 1999. Correspondence and epenthetic quality. In Catherine Kitto & Carolyn Smallwood (eds.), *Proceedings of AFLA VI: The sixth meeting of the austronesian formal linguistics association*, 181–200. Toronto: Department of Linguistics, University of Toronto.

Klatt, Dennis H. 1973. Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America*. Acoustical Society of America (ASA) 54(4). 1102–1104. https://doi.org/10.1121/1.1914322.

Kondo, Yuko. 1994. Targetless schwa: Is that how we get the impression of stress-timing in English? In *Proceedings of the Edinburgh linguistics department conference '94*.

Labov, William, Sharon Ash & Charles Boberg. 2005. *The Atlas of North American English*. De Gruyter Mouton. https://doi.org/10.1515/9783110167467.

Ladefoged, Peter & Keith Johnson. 2010. *A Course in Phonetics*. Cengage Learning.

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

Miller, Joanne L. 1981. Effects of speaking rate on segmental distinctions. In Peter D Eimas & Joanne L

Miller (eds.), *Perspectives on the study of speech*, 39–74. Taylor & Francis.

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

Oostendorp, Marc van. 1995. *Vowel quality and phonological projection*. Tilburg University PhD thesis.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–157. John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.08pie.

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

Prince, Alan S. & Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Recasens, Daniel. 1991. An electropalatographic and acoustic study of consonant-to-vowel coarticulation. *Journal of Phonetics* 19(2). 177–192.

Recasens, Daniel. 2021. Acoustic characteristics and placement within vowel space of full schwa in the world's languages: A survey. *Journal of the International Phonetic Association*. Cambridge University Press (CUP) 52(1). 59–94. https://doi.org/10.1017/s0025100320000122.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard & Jiahong Yuan. 2014. FAVE (forced alignment and vowel extraction) program suite v1.2.2. https://github.com/JoFrhwld/FAVE. https://doi.org/10.5281/zenodo.22281.

Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291–311.

Stevens, Kenneth N. 2000. *Acoustic Phonetics*. The MIT Press. https://doi.org/10.7551/mitpress/1072.003.0007.

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

Syrdal, Ann K. & H. S. Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America*. Acoustical Society of America (ASA) 79(4). 1086–1100. https://doi.org/10.1121/1.393381. http://dx.doi.org/10.1121/1.393381.

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

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