Phonetic rhythm in American English *-ization*[*]

Juliet Stanton
New York University

December 5, 2021

Abstract

It is commonly assumed that *CLASH and *LAPSE evaluate syllable-sized constituents: a sequence of two adjacent stressed syllables ($\acute{\sigma}\acute{\sigma}$) violates *CLASH, while a sequence of two stressless syllables ($\acute{\sigma}\underline{\sigma\sigma}\acute{\sigma}$) violates *LAPSE (see e.g., Prince 1983, Gordon 2002 for *CLASH; Nespor & Vogel 1989, Green & Kenstowicz 1995, Gordon 2002 for *LAPSE). In this paper I argue, based on patterns of secondary stress in American English *-ization*, that there exist rhythmic constraints that are evaluated with respect to (normalized) duration. I present evidence consistent with this claim from corpus, judgment, and production studies.

1 Introduction

It has been a common assumption since at least Liberman & Prince (1977) that stress is the manifestation of linguistic rhythm. The term *rhythm* implies alternation, or the timed succession of weak and strong beats. In English, for example, rhythmic alternation can be found at the phrase level. In the phrase *twenty-seven Mississippi legislators* (Hayes 1995:28), there is an alternation of weak and strong beats at multiple levels of structure (1): every second syllable carries some level of stress, and every fourth syllable carries a stronger stress.

(1) Rhythmic alternation within a phrase (Hayes 1995:28)

								x			
x				x				x			
x		x		x		x		x		x	
x	x	x	x	x	x	x	x	x	x	x	x
twen	ty	sev	en	Miss	i	sipp	i	leg	is	la	tors

Alternation can also be found at the word level. Words like *Apalachicola*, *Winnepesaukee*, and *reconciliation* demonstrate that there is a general trend towards rhythmic alternation in English words (though exceptions to this tendency, such as *expèllée*, certainly exist).

[*]For comments and helpful feedback I am grateful to Gillian Gallagher, Maria Gouskova, Adam Albright, Donca Steriade, three anonymous reviewers, an anonymous associate editor, and audiences at NYU's PEP Lab, AMP 2020, Berkeley, Cornell, and the Keio x ICU Colloquium Series.

(2) Rhythmic alternation within a word (Hayes 1995:29)

```
                x
x               x
x         x     x
x    x    x  x  x    x
re  con  ci  li  a  tion
```

The term *alternation*, in turn, implies that a sequence of beats is spaced out over time. Weak and strong beats do not happen at once; a strong precedes a weak, which precedes a strong.

This paper addresses the question of how rhythmic spacing should be measured. Current theories of stress propose different answers to this question. In constraint-based approaches to stress that assume feet (e.g. McCarthy & Prince 1993), constraints like PARSESYL require syllables to be parsed into feet, while constraints on foot form (like IAMB, requiring prominence to be foot-final) and constraints on foot alignment (like ALLFTLEFT, requiring each foot to coincide with the word's left edge) regulate the spacing, or *distance*, between the stresses. In constraint-based approaches to stress that eschew feet (e.g. Gordon 2002), constraints like *LAPSE (e.g., Nespor & Vogel 1989, Green & Kenstowicz 1995, Gordon 2002) and *CLASH (e.g., Prince 1983, Gordon 2002) regulate the distance between stressed and stressless syllables directly. *LAPSE (3) prohibits stresses from occurring too far apart, while *CLASH (4) prohibits them from occurring too close together.

(3) *LAPSE: assign one * for each sequence of two adjacent stressless syllables.

(4) *CLASH: assign one * for each sequence of two adjacent stressed syllables.

Despite their differences, these approaches share something fundamental: both calculate distance over units of formal structure (syllables and feet). This paper explores an alternative: rhythm is calculated not over units of formal structure but over (normalized) duration. Evidence for this alternative comes from suffixal stress in American English *-ization* (see Stanton 2019 for a similar case study from *-ative*). In these forms, stress on *-ize* is often variable: one can pronounce *historicization*, for example, as [hɪˌstɔrəsəˈzeɪʃ(ə)n] or [hɪˈstɔrəˌsaɪˈzeɪʃ(ə)n] (transcriptions from the Oxford English Dictionary). I claim that this variation is, at least in part, governed by rhythm: *-ize* stress becomes more likely as the suffix's distance from the rightmost stem stress increases. To give an example: a word like *culturalization* (with a stem-final stress lapse) is more likely to bear *-ize* stress than a word like *acidization* (without one). I show that the metric of distance speakers use when deciding whether or not to stress *-ize* must reference duration in a more direct way than is generally assumed by theories of stress.

1.1 Roadmap

Section 2 of this paper introduces stress in *-ization*, provides a preliminary analysis of suffixal stress, and documents the existence of rhythmic effects in data from the Oxford English Dictionary (OED). Section 3 advances the hypothesis that *-ize* stress is sensitive to duration. Section 3 provides evidence from a forced-choice task that supports this hypothesis, and Section 5 develops the proposal. Section 6 presents evidence from a production task that is consistent wiht the proposal. Section 7 discusses further evidence consistent with the idea that (normalized) duration plays a role in stress assignment, both in English and elsewhere, and Section 8 briefly concludes.

1.2 Departures from Stanton (2019), on *-ization*

Patterns in the stress of *-ization* words have been previously addressed by Stanton (2019), who reports on the results of a corpus study using the OED. In addition to a more detailed corpus study that takes into account frequency information, this study expands on her findings in a few ways.

The primary difference between the work reported in Stanton (2019) and here is the inclusion of three experimental tasks that probe speaker intuitions regarding the stress of *-ization* words. These tasks, in addition to providing us with evidence that speaker preferences match the dictionary data, allow us to probe larger questions unanswered in Stanton's (2019) work on *-ative*. For example, Stanton leaves open the question of whether the constraints that regulate rhythm ought to reference raw or normalized duration; the results in Section 4 of this paper strongly suggest that they ought to reference normalized duration. This paper thus makes an independent contribution to both our empirical knowledge of English stress as well as our understanding of its theoretical underpinnings.

2 Stress in *-ization*

Words ending in *-ization* vary in whether or not *-ize* bears stress. A look at the OED makes this clear: some words are transcribed with stress on *-ize* (e.g. *solarization*), others are transcribed without stress on *-ize* (e.g. *fascization*), while still others are transcribed with both variants (e.g. *serialization*). Before focusing on this variation, it is necessary to review more general properties of stress in *-ization* to understand what conditions stressing and destressing of *-ize*.

2.1 Background

For analysis, it is useful to separate words that end in *-ization* into two domains: the stem domain (all of the material that precedes *-ization*) and the suffixal domain (composed of *-ize* and *-ation*). I illustrate with *solarization* in (5).

(5) Division of *-ization* forms into stem and suffixal domains

solar – ize – ation

stem *suffixal*

We need a few assumptions to illustrate why stress on *-ization* varies. I assume, for the purposes of this paper, that stress falls on *-ize* and the first syllable of *-ation* due to the presence of underlying diphthongs. Since English stress is weight-sensitive (Chomsky & Halle 1968 and many others), the constraint WEIGHT-TO-STRESS ((6); Prince 1990) compels stress on these syllables.[1]

(6) WEIGHT-TO-STRESS (WSP): assign one * for each heavy syllable (containing a long vowel or coda consonant) that does not bear a stress.

Note, however, that if both *-ize* and the first syllable of *-ation* were stressed, a stress clash would result. While clashes with a heavy syllable in first position are generally acceptable in English (cf.

[1] To be clear, I assume that the choice between [aɪz] and [əz] is made in the phonological grammar. More generally, while I assume that the choice between *-ize* and another morpheme (e.g. *ify*, see Raffelsiefen 2004) is made by the morphological grammar, I assume that stress and its segmental effects are computed by the phonological grammar.

bàndána, *Nàntúcket*; examples from Pater 2000:244), *-ate* is what Liberman & Prince (1977) refer to as a *strong retracter*: the syllable immediately preceding it is generally unstressed. I use this observation to motivate a suffix-specific markedness constraint, *CLASH$_{\textit{-ate}}$ (7), which forbids a sequence of two stressed syllables in the case that one of those syllables is the verbalizing *-ate*.

(7) *CLASH$_{\textit{-ate}}$: assign one * for each sequence of two stressed syllables in which one of the syllables is the verbalizing *-ate*.

I attribute the preference for *-izátion* (vs. *-ízation*) to a preference for penultimate stress, implemented here as *LAPSER ((8); for positional lapse constraints, see Gordon 2002).

(8) *LAPSER: assign one * if neither of the final two syllables bears stress.

There are no arguments for rankings among any of the three constraints under consideration. The variation between *ìzátion* and *izátion* is illustrated in (9) for *serialization*. I do not take a position here on why the first stress of *serial* bears stress; it could be due to either cyclic preservation (see Stanton & Steriade in prep for cyclic preservation in Level 1 derivatives) or to a preference for word-initial stress.

(9) Tableau for *serialization*

		sérial-ize-ation	*LAPSER	*CLASH$_{\textit{-ate}}$	WSP
☞	a.	sèrialìzátion		*	
☞	b.	sèrializátion			*
	c.	sèrialízation	*		*!

Here, (9c) is harmonically bounded by (9b) – it incurs a violation of *LAPSER in addition to the shared violation of WSP – and is eliminated. Candidates (9a) and (9b) tie because they each violate one constraint: (9a) (with stress on *-ize* and *-ate*) violates *CLASH$_{\textit{-ate}}$, while (9b) (with stress only on *-ate*) violates WSP. Which candidate surfaces as optimal depends on the ranking of these two constraints. If *CLASH$_{\textit{-ate}}$ dominates WSP, (9b) surfaces as the winner. If WSP dominates *CLASH$_{\textit{-ate}}$, (9a) wins.

The question of interest, then, is whether or not we can predict when *-ize* will be stressed: is the variation between (9a) and (9b) conditioned by anything, or is it random?

2.2 Rhythmic effects in *-ization* stress: the available evidence

Expanding on Stanton (2019:7.2), I conducted a corpus study to verify that rhythmic factors are implicated in *-ization* stress. The corpus consisted of all relevant *-ization* forms in the OED as of February 2019 (n=858[2]); for analysis, only those forms with an American English transcription were considered (n=681). The inner suffix was counted as "stressed" if *-ize* was transcribed with the diphthong [aɪ] and as "stressless" if *-ize* was transcribed with a reduced vowel, [ə] or [ɪ]. Cases like *relativization* and *serialization*, where both variants are recorded, represent the majority of tokens (544/681) and were excluded because it was not clear whether these variable tokens should

[2]The number of forms in the OED was actually 773. I report a higher number here because some stems have two stress patterns, e.g. *multimerization* can be 202-x10 or 020-x10. In such cases, variants are counted as separate stems.

be counted as "stressed" or "stressless".[3]

The results (10) demonstrate a clear rhythmic effect. Stress on *-ize* is least frequent when it would result in a clash with the stem (as in *àlcohòlìzátion*), more frequent when it would resolve a *Lapse violation (as in *chànnelìzátion*), and most frequent when it would resolve an *ExtLapse violation (as in *pèrsonalìzátion*). (*ExtLapse, or *ExtendedLapse, is a rhythmic constraint that disprefers sequences of three adjacent stressed syllables; see Gordon 2002.)

(10) Results from OED corpus study

Effect of *-ize* stress	Stressed *-ize*	Stressless *-ize*	% stressed
*Clash violation	*àlcohòlìzátion* (n=18)	*párallèlizátion* (n=26)	40.9% (18/44)
*Lapse satisfaction	*chànnelìzátion* (n=35)	*dichòtimizátion* (n=31)	53.0% (35/66)
*ExtLapse satisfaction	*pèrsonalìzátion* (n=24)	*cùlturalizátion* (n=3)	88.9% (24/27)

As shown in (11), a logistic regression shows that the difference between the *Lapse and the *ExtLapse contexts is statistically significant. The dependent variable is whether or not *-ize* bears stress; the independent variable is rhythmic context, with *Lapse as the baseline. Comparisons are therefore between *Clash and *Lapse as well as *Lapse and *ExtLapse.

(11) Results of statistical analysis for OED data

Predictor	Coefficient	z value	Significant?
Intercept	0.12	–	–
Context: *Clash	-0.49	-1.24	No (p = .21)
Context: *ExtLapse	1.96	3.00	Yes ($p < .01$)

What other factors might be relevant here? Perhaps frequency-related factors play a role. For example, if the *-ize* base of the *-ization* word is frequent, it could be the case that speakers are accustomed to hearing *-ize* stress associated with that lexical item, leading them to be more likely to stress *-ize* in the *-ization* form. It could also be the case that frequency of the *-ization* form impacts its likelihood to bear stress: perhaps more frequent *-ization* forms, for example, are less likely to bear *-ize* stress. (For more on frequency-related factors that impact the pronunciation of derived words, see Hay 2004; see also Collie 2008 for applications to English stress.) To assess these possibilities, I first fit a logistic regression to the data whose predictors included only the frequency of the *-ize* base and the frequency of the *-ization* derivative.[4] The frequency measures are from the OED's frequency bands, which assigns words a frequency score of 1 through 8.[5] This model did not find

[3]I have also considered what the data would look like were the variable cases counted as .5 in both categories. Under this interpretation, 48.9% of the items in the *Clash category carry *-ize* stress, 62.7% of the items in the *Lapse category carry *-ize* stress, and 61.3% of the items in the *ExtLapse category carry *-ize* stress. The comparisons between the *Clash and *Lapse contexts (χ^2 = 4.41) and the *Clash and *ExtLapse contexts (χ^2 = 4.55, $p < .05$) are significant assuming an alpha of $p < .05$; the comparison between *Lapse and *ExtLapse is not. I report the results with smaller numbers below because I do not know how to fit and compare regression models with multiple factors on data where some forms have been double-counted.

[4]The results discussed here are qualitatively identical if one frequency-related measure is used, namely the frequency of the *-ize* base minus the frequency of the *-ization* derivative.

[5]Frequencies are log-scaled and controlled for homonyms and other complexities; see the OED's discussion of their frequency measures (https://public.oed.com/how-to-use-the-oed/key-to-frequency/) for more.

a significant effect of either derivative frequency (*p* = .14) or base frequency (*p* = .10). Adding the rhythmic factor in (11) results in a significant improvement to the model (χ^2 (1) = 14.22, $p < .001$), indicating that rhythm makes a contribution independent of frequency. Additionally, I compared a model that incorporates rhythmic and frequency-related predictors to a model that includes only the rhythmic predictor, and the larger model does not result in a better fit to the data (χ^2 (2) = 2.52, *p* = .28); it is unlikely that frequency plays a role in predicting *-ize* stress. This finding should be treated with caution, however, as the numbers in (10) are small (and see Section 5 for evidence that frequency effects surface in production).

Another factor that might play a role has to do with the identity of the stem-final consonant. Some forms ending in *-ize*, like *exorcize* and *Catholicize*, have a stem-final [s]. As the *-ization* derivatives of these forms violate a constraint on the co-occurrence of identical segments, *[s]. . . [s] (see Suzuki 1998 on the instantiation of the Obligatory Contour Principle as **x. . . x* constraints), speakers might prefer to lengthen the vowel in between them to [aɪ] to increase the time between the [s]s. This, however, is not the case in the dictionary data (though see Sections 4 and 5 for evidence that the nature of the stem-final segment is important in both judgment and perception tasks). Adding a predictor to the model encoding whether or not the stem-final segment is [s] does not result in an improved fit to the data relative to the model in (11) (χ^2 (1) = 1.10, *p* = .29).

While there have been no studies on the pronunciation of *-ization* words in natural speech, available evidence suggests that speaker productions match the OED trends. The data in (12) are from Forvo, an online pronunciation dictionary. The dictionary was searched in February 2019 for each *-ization* word in the OED. Native speaker status and *-ize* stress were determined by ear,.

As is clear from (12), the rates of *-ization* stress are overall lower in the Forvo data, suggesting that the rate of *-ize* stress is inflated in the OED data (see Sections 4 and 6 for further evidence that speakers of American English generally disprefer stress on *-ize*). The trend, however, is the same: *-ize* stress is least frequent in the *CLASH context, more frequent in the *LAPSE context, and most frequent in the *EXTLAPSE context.

(12) Results from the Forvo corpus study

Effect of *-ize* stress	Stressed *-ize*	Stressless *-ize*	% stressed
*CLASH violation	*rèalìzátion* (n=4)	*tàblòidizátion* (n=21)	19% (4/25)
*LAPSE satisfaction	*fòssilìzátion* (n=49)	*demòbilizátion* (n=151)	24.5% (49/200)
*EXTLAPSE satisfaction	*àctualìzátion* (n=21)	*làbializátion* (n=51)	29.2% (21/72)

A statistical model of (12) does not find a significant difference between the rates of *-ize* stress in the *CLASH and *LAPSE contexts (*p* = .56), nor does it find a significant difference between the rates of *-ize* stress in the *LAPSE and *EXTLAPSE contexts (*p* = .06). This is likely due to the relatively small number of forms in the *CLASH and *EXTLAPSE contexts. As was true for the OED data, adding frequency-related factors, namely the frequency of the *-ization* derivative and that of its *-ize* base, does not result in an improvement in model fit (χ^2 (2) = 0.61, *p* = .74), nor does adding a predictor differentiating [s]-final stems from all others (χ^2 (2) = 0.00, *p* = .98).

Rhythm, then, appears to affect how likely *-ize* is to bear a secondary stress in *-ization* forms. Other factors, such as frequency and the identity of the stem-final segment, do not have an effect in these data (though these results should be treated with caution, and cf. Sections 4-6).

3 Hypothesis

In what follows, I investigate the hypothesis that *-ize* stress in *-ization* is sensitive to normalized duration. The longer the stressless string between the rightmost stem stress and the stem boundary is, in relation to the entire word, the more likely *-ize* is to carry stress. To give an example: *-ize* stress is more common in *personalization*-type words because the post-stress material in *pérsonal* takes up a large percentage of the overall duration of *personalization*. By comparison, the post-stress material in *álcohòl* takes up a smaller percentage of the overall duration of *alcoholization*. Analytically, this amounts to a hypothesis that stress in *-ization* is governed by rhythmic constraints that reference normalized duration, a more fine-grained metric than is typically assumed.

If this hypothesis is correct, trends in the dictionary data should mirror trends in the phonetics. As the number of syllables between the rightmost stem stress and *-ize* increases, so should the normalized duration of the material between the rightmost stress and the stem boundary. This is expected given the findings in (10) and (12): the more syllables that follow the rightmost stem stress, the more likely *-ize* stress is. This prediction is diagrammed in (13): words like *àlcohòlizátion* should have a shorter post-stress duration than words like *chànnelizátion*, which should have a shorter post-stress duration than words like *pèrsonalizátion*.

(13) Different post-stress durations (in black) in *-ization* forms

a. $\grave{V}$ C_0 -izátion (*àlcohò**l**izátion*): shortest
$\grave{V}$ $\boxed{C_0}$ *izátion*

b. $\grave{V}$ C_0VC_0 -izátion (*chà**nnel**izátion*): longer
$\grave{V}$ $\boxed{C_0V\ C_0}$ *izátion*

c. $\grave{V}$ $C_0VC_0VC_0$ -izátion (*pè**rsonal**izátion*): longest
$\grave{V}$ $\boxed{C_0V\ C_0V\ C_0}$ *izátion*

While this prediction might seem self-evident, there have been results of lapse compression in the literature. Nespor & Vogel (1989:102), for example, report that speakers increase their speech rate in order to lessen the duration of a stress lapse. If compression is used as a way to ameliorate *LAPSE violations in American English, it could be the case that the duration of the post-stress material is shortened in (13c) such that it becomes close or equivalent to the duration of the post-stress material in (13b). This possibility means that the prediction that more material is associated with more duration should be verified rather than assumed.

To know if this hypothesis is plausible, we need to know whether or not the trends observed in the dictionary data correlate with the predicted trends in duration. We also need to know whether or not the speakers' preferences and productions match these trends.

4 Experimental support: forced-choice

To test the hypothesis, I ran a set of experimental tasks. The forced-choice task discussed in this section asked participants to choose between *ìzátion* and *-izátion* variants of the same form. The results of this task are best-modeled under the assumption that speakers are sensitive to normalized duration and use this knowledge when judging an *-ization* form.

Table 1: *-ization* items, by rhythmic profile and interstress C(s)

*CLASH (n=10) Interstress C(s)	*LAPSE (n=10) Interstress C(s)	*EXTLAPSE (n=10) Interstress C(s)
Pràgueizátion [g]	*Ègyptizátion* [dʒ], [pt]	*Pròvidencizátion* [v], [d], [ns]
Quebècizátion [k]	*Wyòmingizátion* [m], [ŋ]	*Sènegalizátion* [n], [g], [l]
Chàdizátion [d]	*Cùbanizátion* [b], [n]	*Ìndianàpolisizátion* [n], [p], [l]
Ròmeizátion [m]	*Bròoklynizátion* [kl], [n]	*Antàrcticanizátion* [ɹ(k)t], [k], [n]
Japànizátion [n]	*Àustinizátion* [st], [n]	*Blòomingtonizátion* [m], [ŋt], [n]
Brònxizátion [ŋks]	*Tèxasizátion* [ks], [s]	*Mèxicanizátion* [ks], [k], [n]
Vermòntizátion [nt]	*Phòenixizátion* [n], [ks]	*Mìchiganizátion* [ʃ], [g], [n]
Frànceizátion [ns]	*Alàskanizátion* [sk], [n]	*Òberlinizátion* [b], [ɹl], [n]
Bàsqueizátion [sk]	*Rùssianizátion* [ʃ], [n]	*Màdisonizátion* [d], [s], [n]
Mìnskizátion [nsk]	*Ìcelandizátion* [sl], [nd]	*Ròcehsterizátion* [tʃ], [st], [ɹ]

4.1 Items and acoustic analysis

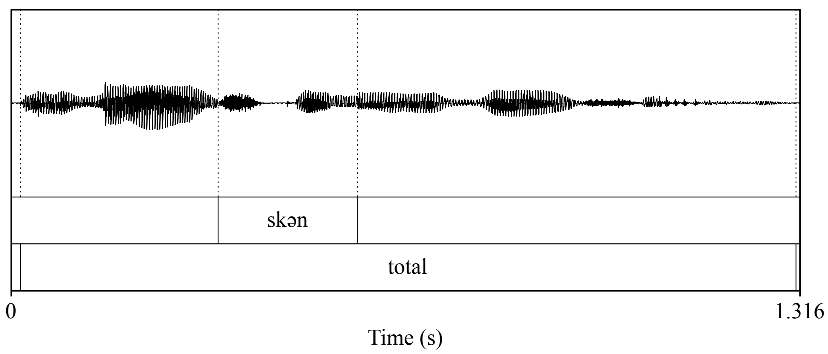
For the experiment, I recorded a single native speaker of American English producing *-ìzátion* and *-izátion* variants of forms that ended in *-ization*, all placenames or demonyms. The forms were recorded in a list, with fillers at the beginning and the end of the list to minimize intonational differences. The placenames themselves were recorded along with the *-ization* derivatives.

As shown in Table 1, these items included ten where *-ize* stress would introduce a violation of *CLASH with the stem (*Chadization*), ten where *-ize* stress would satisfy *LAPSE (*Brooklynization*), and ten where *-ize* stress would satisfy *EXTLAPSE (*Madisonization*). Within these categories, the consonants following the rightmost stem stress were varied by item; this variation was introduced to ensure that there was some durational variation among the items.

To measure the normalized duration of each stress-to-boundary string, the distance between the rightmost stem stress and the end of the stem was divided by the total duration of the form. For each item, measurements were taken off of the *-ìzátion* productions. This means that in *Alàskanizátion*, for example, the interval between the offset of the stressed stem vowel and the onset of stressed [aɪ] (0.233 s), [skən], was divided by the overall duration of the form (1.294 s), yielding a normalized durational measurement of 0.18. Figure 1 illustrates how these measurements were taken in Praat (Boersma & Weenink 2021).[6]

[6] An anonymous reviewer notes that the presence or absence of *-ize* stress affects the normalized durations of the

Figure 1: Calculating normalized stem-to-boundary duration in *Alàskanìzátion*



Normalized duration here is a relational measure: it is the proportion of the entire word's duration that is taken up by the stem-to-boundary duration. In this, I follow earlier authors (notably Lunden 2013) in assuming that the right representation of duration in phonology is a normalized one, as this allows us to abstract away from issues like speech rate (see Section 5.1 for evidence that speech rate does not impact *-ize* judgments) and idiosyncratic variation in the durational properties of segments.[7] Note, however, that this conception of normalization makes predictions that are yet to be explored. For example, the present approach predicts that a lapse in a word of the shape *σ́σσσ́σ* is treated as 'longer' than a lapse in a word of the shape *σσσ́σσσ́σ*, and should therefore be penalized more heavily. Further work is necessary to verify that this prediction is correct.

Investigation of the durational properties of these forms reveals that the prediction regarding the association between normalized duration and rhythmic context is verified (Figure 2): the normalized duration of the stem-to-boundary material is shortest in the *CLASH context (where there are no interstress stressless syllables), longer in the *LAPSE context (where there is one stressless syllable), and longest in the *EXTLAPSE context (where there are two).

This finding renders the first part of the hypothesis, outlined in Section 3, plausible: the broad trends discovered in the dictionary study correlate with properties of the productions.

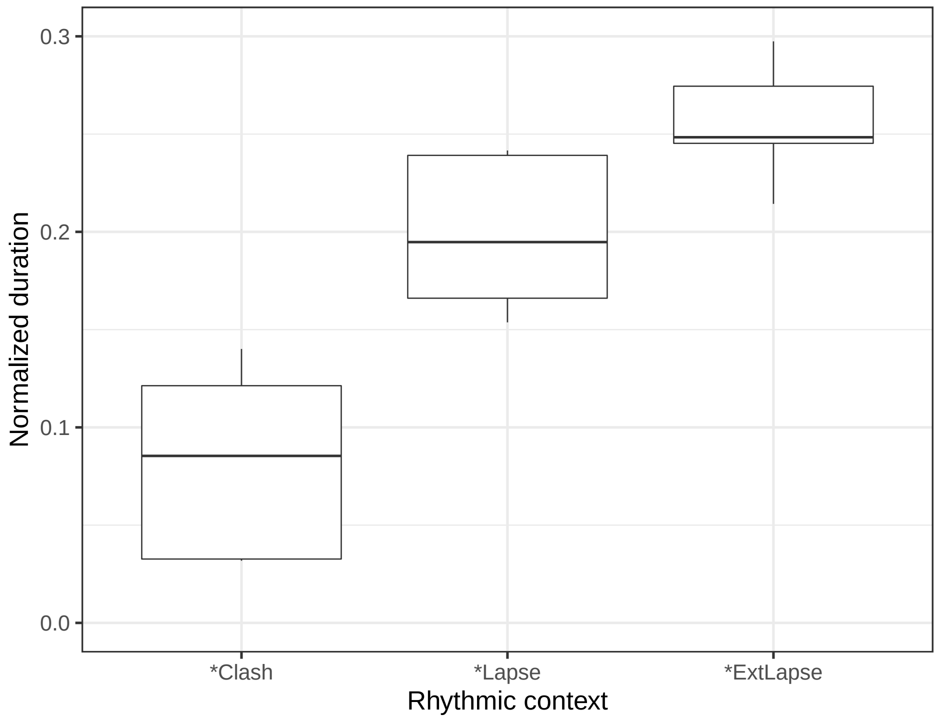
4.2 Design

Thirty stimuli were created from the forms in Table 1, differing only in suffixal stress. For example, the item for *Alaskanization* included two forms, *Alàskanizátion* and *Àlaskanìzátion*. The words were presented in the frame *Prepare for the ___ of your vacation!* (e.g. *Prepare for the* ***Alaskanization*** *of your vacation!*), and participants were told that they were helping a travel company pronounce words in a new slogan. An illustration of what occurred at each trial is provided in Figure 3. First, participants clicked on the radio button associated with the placename (e.g. ***Alaska***) and listened to

stem-to-boundary string. While true, the normalized durations scale linearly; my choice to base the modeling off of the normalized duration of the *-ize*-stressed form is essentially arbitrary and the results do not qualitatively differ if the modeling is based off the normalized duration of the form where *-ize* does not bear a stress.

[7] An open question here is if, in running speech, the normalizing duration ought to be the word or the phrase. The stimuli employed in this experiment are single words, so it is not possible to distinguish between these options at present.

Figure 2: Normalized duration by rhythmic context



a recording of that word. After this, the slogan was revealed, along with the two radio buttons that corresponded to the two pronunciations of the (bolded and italicized) test item. After both options had been played, participants were instructed to choose which option they preferred by clicking on either Option 1 or Option 2. The experiment then advanced to the next item.

The order of *-izátion* and *-ìzátion* was randomized by item and participant; item order was also randomized by participant. Experiments were constructed with Experigen (Becker & Levine 2013).

4.3 Participants

Fifty participants were recruited using Amazon's Mechanical Turk. Prerequisites for inclusion included a U.S. IP address and a HIT approval rate of greater than 95%. All participants indicated in a demographic questionnaire that they were native speakers of English, and no speakers were excluded from the analysis. Participants were compensated $1.50 for their time.

4.4 Results

Patterns in the data suggest that the hypothesis is correct. First, the distinctions among rhythmic categories are consistent with the expectations, given the data and acoustic results (though these distinctions are not statistically significant; see below). These are laid out in (14). Rates of *-ize* stress are lowest in the *CLASH context (forms like *Quebecization*), higher in the *LAPSE context (forms like *Alaskanization*), and highest in the *EXTLAPSE context (forms like *Mexicanization*).

Figure 3: Example of an experimental item

This is a photo of a trail in [Alaska].

Prepare for the ***Alaskanization*** of your vacation!

[Option 1]

[Option 2]

Which of these options do you prefer?

[Option 1] [Option 2]

Notice, however, that *-ize* stress is chosen at fairly low rates overall; this is consistent with the Forvo data (12), where *-ize* stress appeared to be the dispreferred option.

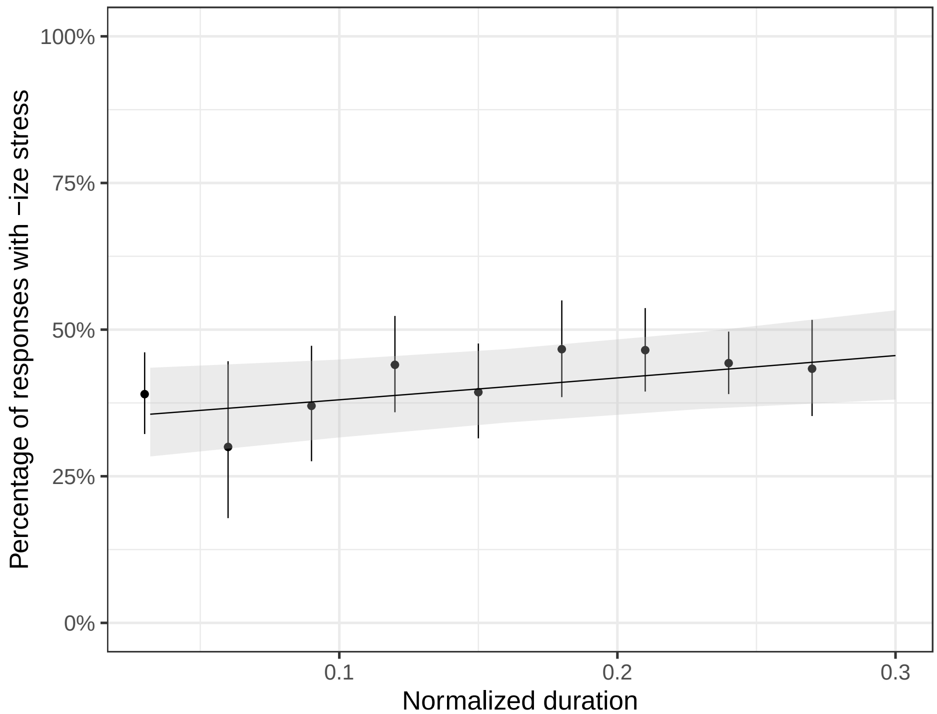
(14) Results by rhythmic context

Context (Example)	*-ize* stressed	*-ize* stressless
*CLASH (*Quebecization*)	34.9%	65.1%
*LAPSE (*Texasization*)	39.4%	60.6%
*EXTLAPSE (*Mexicanization*)	40.2%	59.8%

Turning our attention to the role of normalized duration in conditioning *-ize* stress, we find a similarly predictable result: there is a positive correlation between *-ize* stress and normalized duration. The longer the normalized duration of the stress-to-boundary material, the more likely participants were to prefer the form with *-ize* stress. Figure 4's x axis represents normalized duration. The raw results are represented as point ranges, where the point represents the percentage of responses preferring *-ize* stress and the line represents the 95% binomial proportion confidence interval. To aid reability, data were binned into arbitrary intervals separated by a consistent measure. The best-fit line in Figure 4 as drawn using R's effects package (Fox 2003, Fox & Weisberg 2018, Fox & Weisberg 2019) and is based on the best-fit model in (15).

Interestingly, the statistics indicate that duration provides a better explanation than does rhythmic category as to why the participant responses patterned as they did. The best-fit mixed-effects logistic regression model finds a significant effect of normalized duration as well as the identity

Figure 4: relationship between normalized duration and *-ize* stress



of the final segment (/s/ vs. others). All models reported from this point forward, including (15), were fitted using the glmer function of R's lme4 package (Bates et al. 2015) and include a random intercept for participant. (A random effect for item was not included for convergence reasons.) Significance values are from lmerTest (Kuznetsova et al. 2017).

(15) Model with duration as a fixed effect

Factor	Coefficient	z value	Significant?
Intercept	-0.70	–	
Normalized duration	1.55	2.19	Yes ($p < .05$)
Final /s/	0.34	2.31	Yes ($p < .05$)

As is true for the corpus data, the frequency of the *-ization* derivative and its *-ize* base do not play a role in speaker responses ($\chi^2 = 4.12$, $p = .12$). The crucial finding here is that (15) has a lower AIC (1934.59 vs. 1936.53) and BIC (1955.83 vs. 1963.09) than does an otherwise equivalent model that substitutes rhythmic context for normalized duration. (See Section 5.1 for evidence that this best-fit comparison replicates with even higher AIC/BIC differences.) In this model, the factor referencing rhythmic context is treatment-coded, with *LAPSE as the reference level. A look at the results (16) shows that the comparisons between *CLASH and *LAPSE on one hand, and *LAPSE and *EXTLAPSE on the other, are not significant: the model does not find an effect of rhythmic context.

(16) Results for model with rhythmic category

Factor	Coefficient	*z* value	Significant?
Intercept	-0.38	–	
*LAPSE vs. *CLASH	-0.22	-1.58	No (p = .11)
*LAPSE vs. *EXTLAPSE	0.08	0.56	No (p = .58)
Final /s/	0.37	2.52	Yes ($p < .05$)

A further suggestion that normalized duration is the right way to model the data comes from within-context effects of normalized duration. To investigate this, I fit separate models to the three relevant subsets of the data: one subset with the *CLASH items, a second subset with the *LAPSE items, and the third subset with the *EXTLAPSE items. These models, like the model in (14), include predictors for normalized duration and the presence of a stem-final /s/; also like the model in (14), they include a random effect for participant. As is clear from (17–19), while the effects of normalized duration are not significant within each context, they consistently trend in the expected duration. Within each context, longer stress-to-boundary durations are associated with higher rates of *-ize* stress.

(17) Effects of normalized duration in the *CLASH context

Factor	Coefficient	*z* value	Significant?
Intercept	-0.68	–	
Normalized duration	0.83	0.28	No (p = .78)
Final /s/	0.38	1.26	No (p = .21)

(18) Effects of normalized duration in the *LAPSE context

Factor	Coefficient	*z* value	Significant?
Intercept	-0.76	–	
Normalized duration	1.13	0.38	No (p = .70)
Final /s/	0.32	1.31	No (p = .19)

(19) Effects of normalized duration in the *EXTLAPSE context

Factor	Coefficient	*z* value	Significant?
Intercept	-0.89	–	
Normalized duration	0.56	0.13	No (p = .90)
Final /s/	0.40	1.13	No (p = .26)

We can take the following away from these results. First, phonetic rhythmic information plays a role in speakers' judgments about whether or not to stress *-ize* in *-ization* forms. Second, it's (normalized) duration that matters. The rhythmic category the form belongs to (*CLASH, *LAPSE, *EXTLAPSE) only matters insofar as these categories are shorthand for duration.

5 Analysis

I have suggested above that CON needs to contain rhythmic constraints that refer to normalized duration. This is not, however, the only possible interpretation of the results in Section 4. An equally probable hypothesis, given what we have seen so far, is that speakers are instead reacting to raw duration: the more milliseconds intervene between the rightmost stem stress and the stem's right edge, the more likely the listener is to prefer *-ize* stress (see Stanton 2019 for an instantiation of this proposal for *-ative*).

The raw and normalized duration hypotheses make different predictions regarding what might happen under different experimental conditions. Let us imagine a task, for example, in which half of the experimental items are played at the normal speech rate and half of the experimental items are slowed down. The raw duration hypothesis would predict a difference between listeners' reactions to the normal and the slowed items, as the slowed items would have longer stem-to-boundary durations. The normalized duration hypothesis does not predict a difference: as the stem-to-boundary duration increases, so does the overall duration of the form. Varying the speech rate of the item changes the raw duration, but not the normalized duration, of the stem-to-boundary material.

Section 5.1 reports the results of such an experiment and shows that they support the normalized duration hypothesis. Section 5.2 analyzes the results of the forced-choice task in Section 4 using a Maxent grammar (e.g. Hayes & Wilson 2008) and shows that the proposed analysis results in a relatively close fit to the data.

5.1 Distinguishing between hypotheses: a second forced-choice task

To distinguish between the raw duration and normalized duration hypotheses, I conducted a second forced-choice task. This forced-choice task used half of the *-ization* items from the original task in Section 4; these items are provided in Table 2 for reference. In all other ways (design, recruitment, payment), this second forced-choice task was identical to the first.

Table 2: Half of *-ization* items, by rhythmic profile and interstress C(s)

*CLASH (n=5) Interstress C(s)	*LAPSE (n=5) Interstress C(s)	*EXTLAPSE (n=5) Interstress C(s)
Quebècizátion [k]	*Ègyptizátion* [dʒ], [pt]	*Sènegalizátion* [n], [g], [l]
Chàdizátion [d]	*Cùbanizátion* [b], [n]	*Ìndianàpolisizátion* [n], [p], [l]
Ròmeizátion [m]	*Àustinizátion* [st], [n]	*Antàrcticanizátion* [ɹ(k)t], [k], [n]
Brònxizátion [ŋks]	*Tèxasizátion* [ks], [s]	*Mèxicanizátion* [ks], [k], [n]
Bàsqueizátion [k]	*Phòenixizátion* [n], [ks]	*Ròchesterizátion* [tʃ], [st], [ɹ]

For this experiment, two versions of each item, or thirty items total, were used. In one, both forms were presented at the normal speech rate. In the second, both forms were artificially slowed by 20% (using Praat Vocal Toolkit, Corretge 2012). 20% is enough slowing for the difference between the normal and the slowed forms to be audible, but not enough for the slowed forms to sound unnatural.[8]

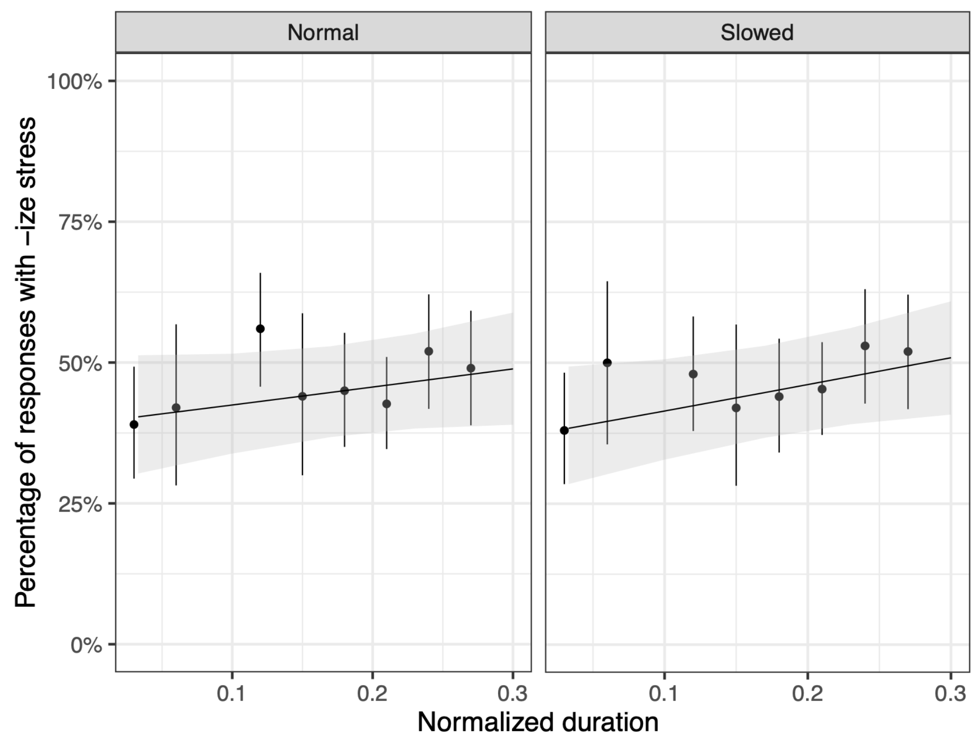
[8]By implementing different speech rates in this way, I make the implicit assumption that different speaking rates involve proportional rescaling of corresponding temporal patterns. An anonymous reviewer points out that this is not in line with what we know about differing speaking rates; see Turk & Shattuck-Hufnagel (2020).

The prediction: if the raw duration hypothesis is correct, we should find a stronger preference for *-ize* stress in the slowed items. This is because the stem-to-boundary duration is longer. If the normalized duration hypothesis is correct, however, we do not expect to find such a difference.

The results are clear and support the prediction of the normalized duration hypothesis.

First, the results from the experiment show that the effect of normalized duration found in Section 4's forced-choice task replicates: normalized duration is a significant predictor of *-ize* stress, with a higher normalized duration leading to higher rates of *-ize* stress (z = 2.23, $p < .05$). The plot in Figure 5.1 breaks up the items by speech rate – I refer to the two categories as *normal* and *slowed* – and shows that the positive trend is evident in both subsets of the data. (Note that the absence of point ranges at intermediate positions in these plots means that there are no items included with a normalized duration falling in that range.)

Figure 5: relationship between normalized duration and *-ize* stress, by item type



Also evident in Figure 5.1 is the lack of a difference between the rates of *-ize* stress in the normal and slowed items: the trend lines are close to identical. (If the slowed items had *-ize* stressed at higher rates, we would have expected to see the trend line for the slowed items have a higher intercept than the trend line for the normal items.) This visual finding is borne out in the numbers: participants selected *-ize* stress for the normal items 53.6% of the time, and *-ize* stress for the slowed items 53.5% of the time. Given this, it is unsurprising that adding a predictor to the model for speech rate does not result in an improvement in fit (χ^2 (1) = 0.13, p = .71), nor does adding an interaction between normalized duration and speech rate (χ^2 (2) = 0.35, p = .84).[9]

[9]Other factors that do not result in an improvement in the fit of the model include base frequency, derivative frequency,

As before, the model that appeals to normalized duration does better on measures of best fit than does an otherwise equivalent model that appeals to rhythmic context (1868.00 vs. 1940.85 for AIC; 1883.94 vs. 1962.10 for BIC). Also as before, within each rhythmic category, the rate of *-ize* stress goes up as the stress-to-boundary distance increases; this effect is significant in the *CLASH context but not in the *LAPSE or the *EXTLAPSE contexts. These results reinforce the suggestion that participants pay attention to normalized duration and not to rhythmic context. Rhythmic context matters only insofar as it it is a stand-in for normalized duration.

5.2 Proposal

For a formal proposal, I assume the relevant phonetic rhythmic constraint takes the form of a single U-shaped function, which assigns increasingly harsh penalties to long lapses and harsh clashes. (See Flemming 2001 and Flemming & Cho 2017 on U-shaped functions in phonology.) To implement this constraint, we need two assumptions. The first is a reference point, or an ideal distance between the stresses. It is not immediately obvious what such a reference point should be: no prior work that I am aware of has determined the ideal normalized duration that must separate two stresses.

For the purposes of this paper, then, I use a measure that comes from the experimental items. To determine the idealized normalized duration, or the RATIOTARGET, I averaged the normalized durations separating the stem-final stress from the suffixal stress in two types of item, both of which exhibit perfect rhythmic alternation. The first type of item is the *Chàdizátion*-type item: items from the *CLASH subset that do not bear *-ize* stress. The second type of item is the *Tèxasìzátion*-type item: items from the *LAPSE subset that do bear *-ize* stress. I collected the normalized durations of the stressless strings for all 20 forms and averaged them; the result, in (20), was 0.21. (The normalized durations ranged from .16 to .32; compare to the overall range of .03 to .41.)

(20) RATIOTARGET = 0.21

With the RATIOTARGET established, we need a constraint whose definition can map normalized duration to violations. Following Flemming (2001) and Flemming & Cho's (2017) lead, I assume the constraint in (21), which takes the difference between a normalized duration and the RATIOTARGET and squares it. Squaring the difference is necessary because it ensures that all violations are positive numbers, which is necessary for the constraint to be evaluated. (The name of the constraint, PERFECTGRIDPHON, is inspired by Prince 1983 and Stanton & Steriade in prep, the latter of which uses a syllable-based, categorical version of the constraint to penalize clashes or lapses.)

(21) PERFECTGRIDPHON: For any stressless string with the normalized duration x, assign a violation equal to $(x - \text{RT})^2$, where RT is the RATIOTARGET defined above.

For an example, let us consider *Alàskanìzátion*, whose stressless string has a normalized duration of 0.18. (Here, and in the analysis more generally, I abstract away from all stressless strings aside from the one between the stem-final stress and the suffixal stress.) The difference between 0.18 and 0.21 is -0.03; squared, the violation of PERFECTGRIDPHON comes out to 0.0009. Now consider *Chàdìzátion*, whose stressless string has a normalized duration of 0.03. The difference between 0.03 and 0.21 is -0.18; squared, the violation of PERFECTGRIDPHON comes out to 0.0324. Finally,

---

and quality of the final segment (/s/ vs. others). In this way, the results of this second forced-choice task differ from the results of the first task; recall that the first task found a significant effect of the quality of the final segment.

consider *Mèxicanìzátion*, whose stressless string has a normalized duration of 0.30. The difference between 0.30 and 0.21 is 0.09; squared, the violation of PERFECTGRIDPHON comes out to 0.0081. Notice, now, that the violation for the form exhibiting perfect alternation (*Alàskanìzátion*, 0.0009) is smaller than the violation of the form that exhibits a *CLASH violation (*Chàdìzátion*, 0.0324) and that of the form that exhibits a *LAPSE violation (*Mèxicanìzátion*, 0.0081). This confirms that the constraint is working as intended.

To be clear, by adopting a constraint like PERFECTGRIDPHON, I assume that phonological constraints have access to phonetic detail. This is not an uncommon assumption; see e.g. Steriade (1997) for explicit discussion of the architecture of the grammar necessary for such constraints to exist, as well as Flemming (2008) for a model of the grammar that allows phonetically detailed output candidates.

5.3 Application to experimental data

To determine if the constraints introduced throughout this paper are sufficient to model the result of Section 4's forced-choice task, I fit a Maxent grammar to the results using the Maxent Grammar Tool (Hayes et al. 2009). The structure of the input files is illustrated in Table 3. The first column contains the forms, the second column contains the two possible choices speakers were presented with, and the third column contains the frequency at which those forms were chosen in the experiment. For the purposes of this model, I included four constraints: *CLASH-$_{ate}$, WSP, PERFECTGRIDPHON, and *səz, the latter of which penalizes two coronal fricatives that are separated by a reduced vowel. This constraint is included due to the significant of Final /s/ in the model summarized in (15) and formalizes the intuition that a co-occurrence constraint between two coronal fricatives may hold more strongly when the fricatives are separated by a short vowel (e.g. [ə]) than it does when they are separated by a long vowel (e.g. [aɪ]). (See also Suzuki 1998:82, where identical elements separated by a single mora are penalized more harshly than identical elements separated by two.)

Table 3: structure of input file for the Maxent Grammar Tool

			*CLASH-$_{ate}$	WSP	PERFECTGRIDPHON	*səz
Alaskanization	*Alàskanìzátion*	.48	1	0	.0011	0
	Alàskanizátion	.52	0	1	.0088	0
Chadization	*Chàdìzátion*	.32	1	0	.0328	0
	Chàdizátion	.68	0	1	.0012	0
Texasization	*Tèxasìzátion*	.48	1	0	.0001	0
	Tèxasizátion	.52	0	1	.0186	1

The Maxent Grammar Tool discovers the weights in (22) for the four constraints.

(22) Constraint weights found by the Maxent Grammar Tool

Constraint	Weight
CLASH-$_{ate}$	0.401
WSP	0.000
PERFECTGRIDPHON	3.613
*səz	0.369

Table 4: comparison between expected (E) and observed (O) frequencies of *-ize* stress

Form	E	O	**Diff**		E	O	**Diff**
Alaskanization	0.41	0.48	**0.07**	*Mexicanization*	0.43	0.40	**0.03**
Antarcticanization	0.42	0.50	**0.08**	*Michiganization*	0.43	0.40	**0.03**
Austinization	0.42	0.40	**0.02**	*Minskization*	0.40	0.42	**0.02**
Basqueization	0.39	0.38	**0.01**	*Oberlinization*	0.43	0.40	**0.03**
Bloomingtonization	0.43	0.34	**0.09**	*Phoenixization*	0.50	0.52	**0.02**
Bronxization	0.49	0.52	**0.03**	*Pragueization*	0.37	0.36	**0.01**
Brooklynization	0.41	0.48	**0.07**	*Providenceization*	0.51	0.56	**0.05**
Chadization	0.37	0.32	**0.05**	*Quebecization*	0.38	0.30	**0.08**
Cubanization	0.40	0.34	**0.06**	*Rochesterization*	0.43	0.36	**0.07**
Egyptization	0.42	0.50	**0.08**	*Romeization*	0.42	0.42	**0.00**
Franceization	0.48	0.40	**0.08**	*Russianization*	0.40	0.44	**0.04**
Icelandization	0.42	0.34	**0.08**	*Senegalization*	0.42	0.48	**0.06**
Indianapolisization	0.50	0.52	**0.02**	*Texasization*	0.51	0.48	**0.03**
Japanization	0.38	0.46	**0.08**	*Vermontization*	0.39	0.34	**0.05**
Madisonization	0.41	0.50	**0.09**	*Wyomingization*	0.40	0.40	**0.00**

Given these weights, the model achieves a decent fit to the experimental data. Table 4 shows a direct comparison between the expected and observed frequencies for each item, along with the difference between these two frequencies (in bold). The average deviation between observed and expected rates of *-ize* stress is .049, or 4.9%.

5.4 Discussion

One interesting result of the modeling exercise above is that while three of the constraints were assigned weights, WSP was not. This indicates that, with the inclusion of PERFECTGRIDPHON in the constraint set, it may be possible to rethink the analysis of stress in *-ization* that was presented in Section 2. Effectively, it may be possible to assume that *-ize* stress is compelled not by WSP, but rather entirely by rhythmic considerations. (The fact that stress always appears on *-ation* could then be due to a constraint requiring stress to fall on one of the word's final two syllables, e.g. *LAPSER from Gordon 2002, and a prohibition against stress falling on the final suffix, *-ion*.)

A comment is necessary here regarding the generality of PERFECTGRIDPHON and its role in the broader English grammar. This study is not the first to find evidence that phonetically-defined rhythmic constraints are active in English: Stanton (2019) documents similar effects in forms that end in *-ative*. Briefly, Stanton finds that the greater the distance between the rightmost stem stress and *-ative*, the more likely speakers are to prefer secondary stress on *-at-*. I find it unlikely that the patterns observed in *-ization* and *-ative* are peculiar to these suffixes, because it is unlikely that children would hear enough words ending in *-ization* and *-ative* to learn the pattern based on these forms alone. In the OED, the average frequency bin for a word ending in *-ization* or *-ative* is 2, meaning that they occur less than .01 times per million words. Unsurprisingly, they are uncommon in the CHILDES parental corpus (Li & Shirai 2000, MacWhinney 2000), where we find three *-ization* (*pluralization*, *specialization*, *civilization*) and zero *-ative* tokens, from a total of

2,579,966. Given the rarity of these forms, it is unlikely that children would acquire the secondary stress patterns documented here and in Stanton (2019) with reference to *-ization* and *-ative* forms alone. It more likely the case that evidence for phonetic rhythm can be found throughout English (see Section 7.1 for a few places where further evidence could be found), and that speakers apply to forms in *-ization* and *-ative* their general knowledge of the English grammar.

There are a number of questions that I leave unanswered here. One major question concerns the interaction between PERFECTGRIDPHON and syllabic rhythmic constraints like *LAPSE and *CLASH. For the time being, I assume that both are essential components of the phonological grammar. Further work focusing on more areas of English stress is necessary to determine if PERFECTGRIDPHON acts as a supplement to, or in fact replaces, the syllabically-defined constraints.

6 Experimental support: production

Results from the forced-choice tasks are consistent with the hypothesis that *-ize* stress is conditioned by the normalized duration of the string between the rightmost stem stress and the edge of the stem. A forced-choice task may, however, not be representative of how people speak, so it is important to verify that the results in Sections 4 and 5 correlate with English speakers' productions. In this section I present the results of a production task, where participants were presented with a collection of *-ization* words and asked to say them aloud. The results of the production task are consistent with those of the forced-choice task: there is a positive correlation between rhythmic category and the rate at which speakers produce *-ize* stress. An additional finding from the study is that not all speakers exhibit variation in *-ize* stress; most speakers consistently leave it stressless.

6.1 Design

For the production study, the stimuli were the same as those included in the forced-choice task (see Table 1). As was the case with the forced-choice task, participants were told that they were helping a travel company pronounce words with new slogans. First, participants heard a recording of the placename (e.g. *Quebec*). Then, participants were asked to read the slogan aloud (e.g. *Prepare for the* ***Quebecization*** *of your vacation!*).

The experiment took place over the Zoom web conferencing platform, with the experimenter advancing from slide to slide once the participants produced the slogans. Participants' audio was recorded using their computer's microphone and they were compensated $10.00 for their time.

6.2 Participants

Fifty-seven participants were recruited for the study from Craigslist (a classified advertisement site), Facebook, and Twitter.[10] All but one indicated that they were native speakers of English from the United States. In addition to the one speaker who was not a native speaker of American English, six other participants were excluded due to the author's judgment that they were not being truthful about their language background. With these exclusions, fifty participants remained.

[10] Thanks to Lisa Davidson for posting my advertisement to her Twitter account.

6.3 Results

The experiment yielded 1495 analyzable tokens. (The five unanalyzable tokens were from a single speaker whose Internet connection was unstable.)

Most productions (992/1495) did not involve rhythmic modification of the item. To understand what I mean by rhythmic modification, consider the form *Quebecization*. In the majority of cases (35/50), this form was pronounced as *Quebècizátion*, with stem stress in the intended place and no stress on *-ize*. Because productions of this type were so frequent, I assume that the default option for American English speakers is to leave *-ize* stressless (this is consistent with the Forvo data discussed in Section 2.2 and the results from the forced-choice task in Section 4).

The remaining 15 productions constitute a mix of rhythmic and non-rhythmic modification. One type of non-rhythmic modification is a segmental change that does not affect the rhythmic profile of the form: *Quebècsizátion*, for example, instead of *Quebècizátion*. Another type of non-rhythmic modification includes some forms modified in multiple ways at once. One speaker, for example, produced the form as *Quèbecìzátion*. The combination of *-ize* stress and stem stress shift (*Quèbec* instead of *Quebèc*) means that, like *Quebècizátion*, the form exhibits perfect alternation.

The category of rhythmic modification includes forms like *Quebècnátion* (which has undergone segmental deletion, creating a clash) and *Quèbecizátion* (which has undergone stress shift, creating a lapse). There were, in total, four types of changes that participants employed which led to rhythmic modification. These were segmental addition (Section 6.3.1), stem stress shift (Section 6.3.2), segmental deletion (Section 6.3.3), and *-ize* stress (Section 6.3.4). Our interest will ultimately be in deletion and *-ize* stress, but the other two repairs are discussed here to provide a complete picture.

6.3.1 Segmental addition

In cases that demonstrate addition, material is added to either a stem or a suffixal domain. Two types of addition that were common in the data included suffix doubling (e.g. *France**iz**ization*) and demonym formation (e.g. *Rom**an**ization*, instead of the expected *Romeization*). The 46 tokens that exhibited addition were not evenly distributed by rhythmic context: the majority of additions happened in the *CLASH context. In other words, additions in forms like *Bronxization* (yielding, for example, *Bronx**in**ization*) was more frequent than addition in forms like *Egyptization* (yielding, for example, *Egypt**iz**ization*) (23).

(23) Addition by rhythmic context

Context (Example)	No addition	Addition	% addition
*CLASH (*Quebecization*)	*Bronxization* (n=456)	*Bronx**in**ization* (n=40)	10.3% (40.496)
*LAPSE (*Texasization*)	*Egyptization* (n=495)	*Egypt**iz**ization* (n=5)	1.8% (5/500)
*EXTLAPSE (*Mexicanization*)	*Michiganization* (n=498)	*Michigan**if**ication* (n=2)	.002% (1/499)

The skewed distribution of addition is independently interesting. Addition doesn't make sense from a rhythmic perspective: with *CLASH context forms like *Bronxization*, addition creates a lapse out of what could have been perfect alternation. But the limitation of addition to the *CLASH context tells us something about how long lapses can be. Speakers are willing, if infrequently, to take a form that could have had perfect alternation and change that to a form that includes a lapse. The lower rows of (23), however, show us that speakers are unwilling to insert material that would lengthen a pre-existing lapse.

Within the *CLASH context, errors are distributed fairly evenly across items. The most common type of addition (in 21 or 46 forms) was creation of a demonym from a placename (e.g. *Romanization*, *Vermontenization*, *Praguenization*), a strategy that could be attributed to the large number of demonyms included among the stimuli. Suffix doubling was also common (occurring in 9 or 46 forms) and occurred randomly across items. The remaining cases of addition were idiosyncratic, including forms like *Romyization* (perhaps a misreading of the orthography, *Romeization*) and *Vermonitization* (insertion of a vowel into the stem-final consonant cluster).

6.3.2 Stress shift

In forms that exhibited stress shift, the stress of the stem was pronounced differently than the stem is pronounced in isolation. A common example of this type of error comes from the form *Egyptization*. In isolation, the stem is produced as a trochee (*Égypt*); in the *-ization* derivative, however, many speakers pronounced it as an iamb (*Egýpt*). The 98 tokens that exhibit stress shift are concentrated in four items (24): *Egyptization* (produced as *Egỳptizátion*), *Japanization* (*Jàpanizátion*), *Icelandization* (*Ìcelàndizátion*), and *Rochesterization* (*Rochèsterizátion*).

(24) Rates of stress shift by item

Item	No stress shift	Stress shift	% stress shift
Egypt	17	32	66%
Japan	27	21	46%
Iceland	35	15	30%
Rochester	38	12	24%
Senegal	46	4	8%
Quebec	46	3	6%
Providence, *Vermont*	48	2	4%
Antarctica, *Austin*, *Cuba*, *Michigan*, *Oberlin*, *Phoenix*	49	1	2%
Mexico	48	1	2%

Why should stress shift occur mostly in these items? For *Egyptization* and *Icelandization*, the answer is straightforward. In both cases, there exists a related derivative, *Egýptian* and *Ìcelándic*, that bears the same stress as the stress-shifted form. Using the stress of *Egỳptian* and *Ìcelándic*, rather than the stress of *Égypt* and *Íceland*, achieves lapse satisfaction in *Egỳptizátion* and *Ìcelàndizátion*. These cases can thus be thought of pseudo-cyclic effects (see Steriade 1999 *et seq.*, Stanton & Steriade 2014, in prep; Steriade & Stanton 2020): using the stress of a lexically related form in these cases allows for the satisfaction of *LAPSE.

For *Japanization* and *Rochesterization*, the answers are less straightforward. For the case of *Japanization*, there exist derivatives with initial stress: *Jàpanése* and the less-frequent *Jápanìze*. It is unclear, however, why speakers should adopt this stress, because it is not accentually improving.[11] One possibility is that some speakers know the form *Japanize* and use this to determine the

[11]The associate editor notes that the *Jàpanizátion* production is potentially accentually improving: English has been

stress of the *-ization* form; see Section 6.3.5 for discussion of the existence of these kinds of effects in deletion and *-ize* stress. Finally, in the case of *Rochesterization*, I hypothesize that some speakers have a secondary stress on the second syllable of *Rochester* (thus the word is *Róchèster*, not *Róchester*, as produced by the recorded speaker). For the speakers with a secondary stress, the choice to promote the secondary stress to primary, and to demote the primary to zero stress, can be seen as a move to satisfy both *CLASH (**Róchèsterizátion*) and *EXTLAPSE (**Róchesterizàtion*).

Stress shift is interesting but orthogonal to the questions of interest addressed here, which concern the role of rhythm in the production of American English *-ization*. This is because the factors that govern stress shift are not entirely rhythmic in nature but have more to do with lexical idiosyncrasies and the properties of lexical families.

6.3.3 Deletion

In the 159 tokens that exhibited deletion, segments from either the stem or the suffix were deleted. Deletion was not localized to particular places in the word: to give an example, pronunciations of *Madisonization* involving deletion included *Madonization*, *Madisization*, and *Madisonation*. There are two factors involved in determining whether or not a form undergoes deletion. The first of these is rhythm: deletion is more common in the *EXTLAPSE context than in the *LAPSE context, and deletion is more common in the *LAPSE context than in the *CLASH context (25).

(25) Rhythmic distribution of deletion

Context (Example)	No deletion	Deletion	% deletion
*CLASH (*Quebecization*)	*Bronxization* (n=482)	*Bronxation* (n=14)	2.8% (14/496)
*LAPSE (*Texasization*)	*Texasization* (n=454)	*Texization* (n=46)	9.2% (46/500)
*EXTLAPSE (*Mexicanization*)	*Bloomingtonization* (n=400)	*Bloomingtization* (n=99)	19.8% (99/499)

Given the effects observed in the forced-choice task, the results in (25) are unsurprising. As the stem-to-boundary duration increases, speakers are more likely to delete material, thereby reducing the normalized duration.

The other factor that appears to determine whether or not a form undergoes deletion is the quality of the final segment: /s/ versus others. As is clear from (26), forms with a final /s/ (like *France*, *Texas*, and *Indianapolis*) are significantly more likely to undergo deletion than are forms that end with a different segment ($p < .001$, Fisher's Exact Test). Deletion can thus be seen as a strategy to repair a violation of *səz: instead of lengthening the vowel between the two sibilants, speakers get rid of one of the sibilants. This is independently interesting, as segmental co-occurrence constraints are typically satisfied through modification rather than deletion.

claimed to exhibit an initial dactyl effect, per Liberman & Prince 1977 and others. However, this can't be the whole story: if the initial dactyl effect were decisive, we would expect *Ègyptizátion* to be the most frequent production, contrary to fact.

(26) Distribution of deletion according to word-final segment: /s/ versus others

Stem-final segment	No deletion	Deletion	% deletion
/s/	*Texasization* (n=232)	*Texasation* (n=67)	22.4% (67/299)
Another segment	*Mexicanization* (n=1104)	*Mexicazation* (n=92)	7.7% (92/1196)

It is possible, of course, that these two factors are confounded: perhaps the majority of /s/-final stems occur in the *EXTLAPSE context. This point is addressed in Section 6.3.5.

6.3.4 *-ize* stress

Of the four possible types of rhythmic modification, the most common was *-ize* stress, occurring in 226 of 1496 items. It is worth noting however that *-ize* stress, while common, was not evenly distributed among speakers. Two speakers consistently stressed *-ize* while thirty-three speakers never did, meaning that variation in *-ize* stress was exhibited by only fifteen of the fifty speakers who participated in the experiment.

Nevertheless, there were patterns in the data. First, *-ize* stress was distributed unevenly across rhythmic contexts: *-ize* stress in the *CLASH context was less common than *-ize* stress in the *LAPSE context, which was less common than *-ize* stress in the *EXTLAPSE context (27). This pattern, where the rate of *-ize* stress increases from the *CLASH to *LAPSE to *EXTLAPSE contexts, mirrors those found in the OED data, the Forvo data, and the forced-choice task.

(27) Rhythmic distribution of *-ize* stress

Context (Example)	No *-ize* stress	*-ize* stress	% *-ize* stress
*CLASH (*Quebecization*)	*Quebècizátion* (n=447)	*Quebècìzátion* (n=49)	9.9% (49/496)
*LAPSE (*Texasization*)	*Tèxasizátion* (n=416)	*Tèxasìzátion* (n=84)	16.8% (84/500)
*EXTLAPSE (*Mexicanization*)	*Mèxicanizátion* (n=406)	*Mèxicanìzátion* (n=93)	18.6% (93/499)

As was the case with deletion, *-ize* stress is also more likely to happen in forms with an /s/-final stem (28), though this difference is not quite significant (p = .053, Fisher's Exact Test). The trend is however consistent with what was observed in the forced-choice task: a stem-final /s/ causes speakers to be more likely than usual to prefer a form with *-ize* stress.

(28) Rhythmic distribution of *-ize* stress

Stem-final segment	No *-ize* stress	*-ize* stress	% *-ize* stress
/s/	*Tèxasizátion* (n=250)	*Tèxasìzátion* (n=49)	16.3% (49/299)
Another segment	*Mèxicanizátion* (n=1019)	*Mèxicanìzátion* (n=177)	14.8% (188/1196)

Overall, the trends that exist in *-ize* stress production mirror exactly those trends that were observed in dictionary data and in the forced-choice tasks. This finding supports an assumption, implicit through the first part of this paper, that dictionary data and forced-choice tasks are valid ways of accessing speaker intuition.

6.3.5 Modeling rhythmic repair

Deletion (Section 6.3.3) and *-ize* stress (Section 6.3.4) have something in common: both act to reduce the distance between the rightmost stem stress and the suffixal stress. They accomplish this differently – deletion brings the rightmost stem stress closer to the suffixal stress, while *-ize* stress brings the suffixal stress closer to the rightmost stem stress – but both repairs can be seen as part of a conspiracy to reduce the normalized duration of the stressless span. Because of this similarity, I refer to them together as rhythmic repair.

To make this conspiracy explicit, consider (29). Let us assume that a speaker is presented with the form *Madisonization* and high-ranked *EXTLAPSE forbids them from producing this form as the default *Màdisonizátion* (29c). (I use syllable-based rhythmic constraints here for ease of exposition.) I consider two possible repairs here: stressing *-ize* (which introduces a stress clash with *-ation*, leading to a violation of *CLASH) (29a), and deleting part of the material between the two stresses, violating MAX (McCarthy & Prince 1995, (29b)).

(29) Conspiracy of repairs for *EXTLAPSE

				*EXTLAPSE	*CLASH	MAX
☞	a.	Màdisonìzátion	200210		*	
☞	b.	Màdonizátion	20010			*
	c.	Màdisonizátion	200010	*!		

The choice between these two repairs depends on the ranking of *CLASH and MAX. If *CLASH dominates MAX, the speaker deletes segments so as to move the two stresses closer together. If MAX dominates *CLASH, then the speaker stresses *-ize*, adding a new stress to reduce the distance between the rightmost stem stress and the suffixal stress.

Given that these two repairs perform the same function, we can model them together. To do this, I created a subset of data that excludes all forms exhibiting segmental addition or stress shift. I fit a mixed effects logistic regression model to the remainder of the data. The model asks the following question: which factors influence whether or not a form exhibits rhythmic repair? I included the following predictors in the model: rhythmic context (with *LAPSE as the baseline), identity of the final segment (/s/ vs. others), frequency of the *-ization* derivative, and frequency of the *-ize* base. (For consistency with the previous models, the only random effect included in the model is a random intercept for participant.) We see in (30) that all four factors have significant effects.

(30) Results of statistical model for rhythmic repair

	Factor	Coefficient	z value	Significant?
a.	Intercept	-2.67	–	
b.	Rhythmic context: *CLASH	-1.28	-4.74	$p < .001$
c.	Rhythmic context: *EXTLAPSE	1.18	5.18	$p < .001$
d.	Final /s/	1.92	8.07	$p < .001$
e.	Derivative frequency	-0.62	-2.46	$p < .05$
f.	Base frequency	0.46	2.10	$p < .05$

The negative coefficient for (30b) shows that the overall rate of rhythmic repair is lower in the

*CLASH context than it is in the *LAPSE context. The positive coefficient for (30) shows us that the overall rate of rhythmic repair is lower in the *LAPSE context than it is in the *EXTLAPSE context. The positive coefficient for (30d) shows that if the stem ends with an /s/, rhythmic repair is more likely to happen than if the stem ends with any other consonant. The effects of base frequency and derivative frequency, while both significant, go in opposite directions. The negative coefficient for (30e) shows that, as the derivative becomes more frequent, it becomes less likely that it will exhibit rhythmic repair. The positive coefficient for (30f), in turn, shows us that as the *-ize* base becomes more frequent, the *-ization* derivative is more likely to exhibit rhythmic repair.

An important question to ask, given the questions at issue in this paper, is whether or not substituting a measure of normalized duration for rhythmic context results in a better fit to the data than the model in (30) achieves. But such a measure is difficult to obtain: participants were not exposed to recordings of the *-ization* forms like the participants in the forced-choice task were, and we cannot assume that the speaker who contributed their recordings to the forced-choice task is representative of the population as a whole. Thus while it is not possible to answer this question at present, it is clear that the results from this production experiment are, at the very least, consistent with the hypothesis advanced in this paper.

7 Broader evidence for phonetic rhythm

This paper has argued for a particular conception of phonetic *LAPSE and *CLASH based on data from American English *-ization*. In this section I discuss further potential evidence for phonetically defined rhythmic constraints. Section 7.21 focuses on potential evidence within English, discussing possible phonetic effects in the English rhythm rule (Hayes 1986, Beames 2020) as well as post-tonic syncope (Hooper 1978, Polgárdi 2015). Section 7.2 focuses on potential evidence in other language, including the acceptability of secondary stress in Russian compounds (Gouskova & Roon 2013) and the distribution of secondary stress in longer Finnish words (Karvonen 2008).

7.1 Potential sources of evidence in English

The English rhythm rule (Liberman & Prince 1977, Prince 1983, Hayes 1984, *a.o.*) is a phrasal process in which primary stress retracts to avoid a stress clash. The rhythm rule applies most frequently given a stress-final first word and a stress-initial second word (e.g. *bàmbóo trúmpets* → *bámbòo trúmpets*) and less frequently given a stress-medial or stress-final second word (e.g. *bàmbóo viólas*, *bàmbóo trombónes*). For stress to retract, there must be a stress already residing in the target position of retraction: *complete trumpets*, for example, cannot be *cómplete trúmpets* because the initial syllable of *compléte* does not bear any degree of stress. The rhythm rule is subject to syntactic limitations, discussed in more detail by Liberman & Prince (1977), among others.

While it is typically assumed that the rhythm rule operates on a syllabic representation, Hayes (Hayes 1984:70-3) posits that what conditions the rhythm rule is not the number of syllables between the stresses but rather the duration between them. In support of this hypothesis, Hayes notes a difference in intuition regarding retraction in the phrases *Korbel whiskey*, *Korbel tequila*, and *Korbel champagne*. Hayes claims that retraction is most acceptable in *Kòrbél whískey*; this is expected, given that *whískey* is stress-initial. He claims that, in addition, retraction in *Kòrbél tequíla* seems more likely than retraction in *Kòrbél champágne*. This is not expected under a syllabic formulation of the rhythm rule, because in both cases the primary stress in the second word is peninitial.

Hayes proposes that the greater likelihood of retraction in *Korbel champagne* is due to the fact that there is more material before the second-word stress in *champagne* ([ʃæmp]) than there is before the second-word stress in *tequila* ([tək]).

Hayes, however, does not back this intuition up with data. To fill this gap, Beames (2020) conducted an experiment with the aim of determining if varying the segmental material before the second stress affects the rate at which the rhythm rule applies. Beames's stimuli consisted of all possible combinations of a list of first and second words, provided in (31). In each case, the first word was cast as an adjective and the second was cast as a noun. Beames also included fillers: phrases beginning with *complete* as well as phrases ending with iambic forms such as *mistake*. These fillers were included to ensure that participants adhered to the basic principles of the rhythm rule and are not discussed here. Each combination of words was embedded within a larger carrier phrase (e.g. *when you least expect it, bamboo strikes*), which participants were asked to read aloud.

(31) Stimuli for Beames's rhythm rule experiment (adapted from Beames 2020:5)

Word 1 (Orthography / IPA)	Word 2 (Orthography / IPA)
bamboo / bæ̀mbú	*likes* / laɪks
concrete / kɑ̀nkrít	*bikes* / baɪks
fifteen / fɪ̀ftín	*strikes* / stɹaɪks
	aches / eɪks
	flakes / fleɪks
	wakes / weɪks
	cakes / keɪks
	rakes / ɹeɪks
	fakes / feɪks

Beames found a modest effect of segmental material, summarized in (32).[12] The rate of retraction is lowest when the second word begins with a cluster: this is the case for phrases like *bamboo strikes*. Retraction is more frequent when the second word begins with an obstruent (as in *bamboo cakes*), and more frequent still when the second word begins with a vowel (as in *fifteen aches*) or a sonorant (as in *bamboo rakes*). The effect thus goes in the direction we'd expect, given the discussion in this paper: assuming that a cluster onset is longer than an obstruent, which is longer than a sonorant (see Stanton 2019 for some evidence to support this assumption), the pattern here is one in which the rhythm rule applies more frequently when there is less time between the two stresses.

(32) Effect of segmental content on retraction

Segmental context	Retraction	No retraction	% retraction
V	*bámbòo áches* (n=71)	*bàmbóo áches* (n=10)	87.7% (71/81)
R	*bámbòo rákes* (n=220)	*bàmbóo rákes* (n=32)	87.3% (220/252)
O	*bámbòo cákes* (n=215)	*bàmbóo cákes* (n=34)	86.3% (215/249)
CC	*bámbòo stríkes* (n=137)	*bàmbóo stríkes* (n=30)	82.0% (137/167)

[12]Beames does not provide this table in her paper; I have constructed it from her raw results file.

The comparisons among categories are not statistically significant (though a comparison between the V&R and O&CC effects is: $p < .001$, mixed effects logistic regression), meaning that further work is necessary to confirm the existence of phonetic rhythmic effects in the rhythm rule. But the effect is consistent with the patterns that have been observed in *-ative* (Stanton 2019) and *-ization*.

More possible evidence for phonetic rhythm in English comes from post-tonic syncope. It has been known since at least Hooper (1978) that post-tonic syncope is rhythmically conditioned: syncope can only occur when it does not result in a stress clash. The adjective *séparate* can syncopate to *séprate*, but the verb *séparàte* cannot syncopate to *sépràte*. Polgárdi (2015) casts English post-tonic syncope as a form of lapse resolution: it acts to reduce the number of syllables between two stresses or between a stress and the word's right edge. Given this characterization, along with the possibility of observing syncope in a multitude of different contexts (e.g. *séparable* → *séprable*, *sèparabílity* → *sèprabílity*, *séparableness* → *séprableness*), English post-tonic syncope seems ripe for re-investigation along the lines detailed in this paper.

7.2 Potential sources of evidence outside English

Secondary stress does not usually occur in Russian but can be variably realized in compounds or in certain loan prefixes (Avasenov 1964, Yoo 1992, Gouskova & Roon 2013). Thus the compound [bʲitòn-ə-mʲi̯sálkə] 'concrete mixer' (Gouskova & Roon 2013:393) can be realized with or without the secondary stress on the left half of the compound.

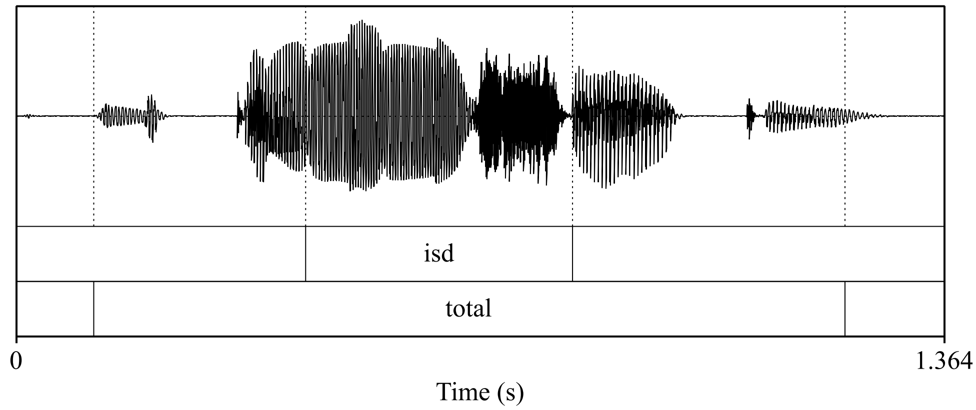
Gouskova & Roon (2013) conduct an experiment to investigate the factors that favor and disfavor secondary stress in subordinating compounds (compounds that contain a linker vowel, such as the [ə] in [bʲitòn-ə-mʲi̯sálkə]). Relevant here is a potential factor involving rhythm: based on prior work by Avasenov (1964), Gouskova & Roon hypothesize that compounds should be more acceptable as the distance between the secondary and primary stresses grows. Thus, all else being equal, [bʲètən-ə-mʲi̯sálkə] (with three syllables between the stresses) should be better-rated than [bʲitòn-ə-mʲi̯sálkə] (with two), and [gòləv-ʌ-lómkə] 'puzzle' (with two syllables between the stresses) should be better-rated than [gʌlòv-ʌ-lómkə] (with only one).

In their experiment, participants were asked to rate the acceptability of each item, with a 1 corresponding to the highest rating (for the most natural items) and a 7 corresponding to the lowest. Gouskova & Roon show that as the distance between the stresses increases, the average rating goes down, meaning that speakers find compounds in which stresses are further apart more natural than compounds in which stresses are closer to gether. They show that a gradient definition of *CLASH fares better with respect to the data than does a more categorical definition (due to Yoo 1992), in which compounds whose stresses are separated by zero or one syllables are less acceptable than compounds whose stresses are separated by more. Concretely, they find that a statistical model assuming a gradient definition of *CLASH does better on measures of best fit than does a statistical model assuming a categorical definition.

Gouskova & Roon measure the distance between stresses by counting syllables and do not address the question of whether the appropriate measure could be a durationally-based measure that does not appeal to formal structure. It is thus an open question whether or not their results could be better modeled by appealing to a measure like normalized duration instead of the number of syllables between the stresses.

To address this question, I calculated a normalized duration for each of Gouskova & Roon's multi-stress compounds by dividing the interstress string by the total duration of the compound.

Figure 6: Calculating normalized duration in [bʲitòn-ə-mʲis̩álkə] (isd = interstress duration)



This calculation follows the calculation done for the experiment reported in Section 4; an example showing the measured distances for [bʲitòn-ə-mʲis̩álkə] is in Figure 6.

I then replaced Gouskova & Roon's factor for "interstress distance" (number of syllables between the stresses) with "normalized duration", leaving everything else in the statistical model the same. In addition to the factor for duration, the model includes a four-level factor for stress type (whether the first member of the compound when in isolation has fixed stress mobile stress, or no vowel) and an interaction between normalized duration and stress type. These additional factors are included in gray, as our interest here is only on the comparison between number of syllables and normalized duration. The results of this new model are provided in (33); see Gouskova & Roon's Table 4 (their p. 402) for the original. The two models are qualitatively similar, differing notably only in Stress type = vowelless; this effect was not significant in Gouskova & Roon's model.

(33) Model for stress clash in Russian compounds, using normalized duration

Factor	Coefficient	*t* value	Significant?
Intercept	3.95	–	–
Normalized duration	-12.43	-6.29	$p < .001$
Stress type = fixed	-1.35	-2.97	$p < .01$
Stress type = mobile	-0.58	-1.14	$p = .27$
Stress type = vowelless	-1.16	-2.31	$p < .05$
Normalized duration * fixed	6.77	1.92	$p = .06$
Normalized duration * mobile	2.48	0.91	$p = .38$
Normalized duration * vowelless	9.00	2.73	$p < .05$

To compare Gouskova & Roon's model (referencing syllables) with the model in (33) (referencing normalized duration), I obtained the AIC and BIC for both models. Both measures indicate that (33) is a better fit to the data than is Gouskova & Roon's model (AIC: 6782.45 vs. 6826.20; BIC: 6947.38 vs. 6991.13). The suggestion, then, is that speakers were more likely to be paying attention to the normalized duration between the primary and secondary stresses, rather than the number of syllables between them, when they were judging Russian compounds.

Another effect potentially providing evidence for phonetic rhythm comes from Finnish, where secondary stress is variable in five-syllable words (Karvonen 2008). In the general case, five-syllable words have a primary stress on the initial syllable and a secondary stres son the penult (e.g. kó.les.te.rò.li ‘cholesterol’, bí.ha.ra.mù.lo ‘city in Tanzania’). If the word-final syllable lacks an onset, however, secondary stress falls on the antepenultimate syllable instead (e.g. ér.go.nò.mi.a ‘ergonomics’, mé.lan.kò.li.a ‘melancholy’). A possible interpretation of this effect is that stress cannot fall on the first vowel in a hiatus sequence (Karvonen 2008:312). However, it could also be possible to attribute this effect to a phonetically-defined version of NONFINALITY, which prohibits stress from coinciding with the word’s right edge. Let us assume there is a constraint, NONFINALITYPHON, which requires a certain amount of normalized duration between the rightmost stress and the word’s right edge. It is then possible to understand the Finnish effect in a different way: stress retracts in forms like ér.go.nò.mi.a because placing a second stress on the penult, as in ér.go.no.mì.a, would be too close to the right edge of the stem. Penultimate stress on a form like bí.ha.ra.mù.lo is acceptable, however, because more material intervenes between the rightmost stress and the word’s right edge.

These two effects – clash in Russian compounds and secondary stress in Finnish five-syllable words – are likely only two of a large number of effects that could be explained by reference to phonetic rhythmic constraints. Further work is necessary to locate and explicate them fully.

8 Conclusion

In this paper, I have argued form patterns of secondary stress in *-ization* that a phonetic rhythmic constraint needs to be included in CON. As detailed in Section 7, these results add to a growing base of evidence that rhythmic constraints pay greater attention to duration than is commonly assumed. All of the work discussed here is thus consistent with a broader view in which stress placement can be directly informed by phonetics (see also e.g. Lunden 2013, Ryan 2016).

The results discussed here also have broader implications for theories of stress. In order to accommodate the finding that phonetic versions of rhythmic constraints are part of CON, the theory must recognize the existence of rhythmic constraints (see also Mołczanow & Łukaszewicz 2021 for an argument that rhythmic constraints are part of CON). This is not true of all theories of stress: Martínez-Paricio & Kager (2015), for example, propose a foot-based theory of stress that does not include rhythmic constraints as part of CON. It is not clear how the effects documented here could be captured using only the constraints that they propose. This case study thus presents a strong argument for theories of stress that include rhythmic constraints and against those that don’t.

References

Avasenov, R. I. 1964. *Udarenije v sovremennon russkom jazyke [modern russian stress]*. New York: MacMillan.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67. 1–48.

Beames, Samantha. 2020. Experimental Evidence for the Rhythm Rule in English. B.A. thesis, New York University.

Becker, Michael & Jonathan Levine. 2013. Experigen – an online experiment platform. Available at http://becker.phonologist.org/experigen.

Boersma, Paul & David Weenink. 2021. Praat: doing phonetics by computer [computer program]. version 6.0.31. http://www.praat.org.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, MA: MIT Press.

Collie, Sarah. 2008. *English stress preservation and Stratal Optimality Theory*: University of Edinburgh dissertation.

Corretge, Ramon. 2012. Praat Vocal Toolkit. http://www.praatvocaltoolkit.com/index.html.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7–44.

Flemming, Edward. 2008. The Realized Input. Ms., Massachusetts Institute of Technology, Cambridge, MA.

Flemming, Edward & Hyesun Cho. 2017. The phonetic specification of contour tones: Evidence from the Mandarin rising tone. *Phonology* 34. 1–40.

Fox, J. & S. Weisberg. 2019. *An R Companion to Applied Regression*. Thousand Oaks, CA: Sage 3rd edn.

Fox, John. 2003. Effects Displays in R for Generalised Linear Models. *Journal of Statistical Software* 8. 1–27.

Fox, John & Sandford Weisberg. 2018. Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals. *Journal of Statistical Software* 87. 1–27.

Gordon, Matt. 2002. A factorial typology of quantity insensitive stress. *Natural Language and Linguistic Theory* 20. 491–552.

Gouskova, Maria & Kevin Roon. 2013. Gradient clash, faithfulness, and sonority sequencing effects in Russian compound stress. *Laboratory Phonology* 4. 383–434.

Green, Thomas & Michael Kenstowicz. 1995. The Lapse Constraint. Ms., MIT. ROA-101, Rutgers Optimality Archive.

Hay, Jennifer. 2004. *Causes and consequences of word structure*. Routledge.

Hayes, Bruce. 1984. The phonology of rhythm in english. *Linguistic Inquiry* 15(1). 33–74.

Hayes, Bruce. 1986. Assimilation as Spreading in Toba Batak. *Linguistic Inquiry* 17. 467–499.

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago/London: The University of Chicago Press.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.

Hayes, Bruce, Colin Wilson & Ben George. 2009. Maxent grammar tool. Software package.

Hooper, Joan B. 1978. Constraints on schwa-deletion in american english. In Jacek Fisiak (ed.), *Recent developments in historical phonology*, 183–207. The Hague: Mouton.

Karvonen, Daniel. 2008. Explaining nonfinality: Evidence from Finnish. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 306–314. Somerville, MA: Cascadilla Proceedings Project.

Kuznetsova, Alexandra, Per B. Brokhoff & Rune H. B. Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82. 1–26.

Li, Ping & Yaushiro Shirai. 2000. *The acquisition of lexical and grammatical aspect*. Berlin & New York: Mouton de Gruyter.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.

Lunden, Anya. 2013. Renanalyzing final consonant extrametricality. *The Journal of Comparative Germanic Linguistics* 16. 1–31.

MacWhinney, Brian. 2000. *The CHILDES project*. Mahwah, NJ: Lawrence Erlbaum 3rd edn.

Martínez-Paricio, Violeta & René Kager. 2015. The binary-to-ternary rhythmic continuum in stress typology: layered feet and non-interventino constraints. *Phonology* 32.

McCarthy, John J. & Alan Prince. 1993. Generalized alignment. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1993*, 79–153. Dordrecht: Springer.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, 249–384. Amherst, MA: GLSA.

Mołczanow, Janina & Beata Łukaszewicz. 2021. Metrical Structure and Licensing: An Argument from Ukrainian. *Linguistic Inquiry* 52. 551–577.

Nespor, Marina & Irene Vogel. 1989. On clashes and lapses. *Phonology* 6. 69–116.

Pater, Joe. 2000. Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* 17. 237–274.

Polgárdi, Krisztina. 2015. Syncope, syllabic consonant formation, and the distribution of stressed vowels in English. *Journal of Linguistics* 51. 383–423.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. In *Chicago linguistics society*, vol. 26, 355–398.

Prince, Alan S. 1983. Relating to the grid. *Linguistic Inquiry* 14(1). 19–100.

Raffelsiefen, Renate. 2004. Absolute ill-formedness and other morphophonological effects. *Phonology* 91–142.

Ryan, Kevin M. 2016. Attenuated spreading in Sanskrit retroflex harmony. To appear in *Linguistic Inquiry*.

Stanton, Juliet. 2019. Phonetic lapse in American English -ative. *Glossa* 4. 1–37.

Stanton, Juliet & Donca Steriade. 2014. Stress Windows and Base Faithfulness in English suffixal derivatives. Talk presented at the 22nd Manchester Phonology Meeting.

Stanton, Juliet & Donca Steriade. in prep. English stress and the cycle. Ms., NYU and MIT.

Steriade, Donca. 1997. Phonetics in Phonology: The Case of Laryngeal Neutralization. Ms., University of California, Los Angeles.

Steriade, Donca. 1999. Lexical Conservatism. In *Linguistics in the Morning Calm, Selected Papers from SICOL 1997*, 157–179. Linguistic Society of Korea, Hanshin Publishing House.

Steriade, Donca & Juliet Stanton. 2020. Productive pseudo-cyclicity and its significance. Talk presented at LabPhon 17.

Suzuki, Keiichiro. 1998. *A typological investigation of dissimilation*. Tucson, AZ: The University of Arizona dissertation.

Turk, Alice & Stefanie Shattuck-Hufnagel. 2020. *Speech Timing. Implications for Thoeiroes of Phonology, Phonetics, and Speech Motor Control*. Oxford: Oxford University Press.

Yoo, Seung-Nam. 1992. *Subsidiary stress in russian compound words*. Champaign, IL: University of Illinois at Urbana-Champaign dissertation.