The causal structure of lenition: A case for the causal precedence of durational shortening

Uriel Cohen Priva, Emily Gleason

Language, Volume 96, Number 2, June 2020, pp. 413-448 (Article)

Published by Linguistic Society of America
DOI: *https://doi.org/10.1353/lan.2020.0025*


LANGUAGE

THE CAUSAL STRUCTURE OF LENITION: A CASE FOR THE CAUSAL PRECEDENCE OF DURATIONAL SHORTENING

Uriel Cohen Priva
Brown University

Emily Gleason
Brown University

Studies of variable lenition patterns converge on two phonetic properties as characteristic of lenition: reduced duration and increased intensity. However, the causal precedence of the two factors remains unclear. We focus on the causal structure of variable lenition. Study 1 examines the relationship between three correlates of lenition—speech rate, stress, and low information content—and their effect on reduced duration and increased intensity. We find that though increased intensity is more prototypically viewed as the core aspect of lenition, the effect of the three correlates on intensity is mediated by duration. Study 2 shows that all frequent lenition processes in the Buckeye corpus involve durational reduction. The contribution of this article is a proposal with a fairly simple principle, with few auxiliary assumptions: reduced duration precedes increased intensity in variable lenition.*

Keywords: consonant lenition, American English, causal precedence, mediation analysis, duration, intensity

1. Introduction.

1.1. Goal and overview. Consonant lenition refers to a list of seemingly unrelated processes that are grouped together by their tendency to occur in similar environments (e.g. intervocalically) and under similar conditions (e.g. in faster speech). These processes typically include degemination, voicing, spirantization, approximantization, tapping, debuccalization, and deletion (Hock 1986, Kirchner 1998). In recent years, studies of nonphonologized lenition patterns have converged on two phonetic properties as characteristic of lenition: reduced duration and increased intensity (Lavoie 2001, Warner & Tucker 2011, Bouavichith & Davidson 2013; see also Katz 2016). However, the relative precedence of the two factors and their relation to established causes of lenition, such as fast speech, remain unclear. This article focuses on the causal structure of nonphonologized lenition processes that occur variably (as a trend, rather than a rule), which we label variable lenition. We use mediation analysis to study the causal chain that correlates extrinsic factors with reduced duration and increased intensity in variable lenition in the Buckeye corpus (Pitt et al. 2007). The results show that even though increased intensity is more prototypically viewed as the core aspect of lenition, capturing the property of being more vowel-like (Szigetvári 2008) and having increased sonority (Parker 2002, Smith 2008), reduced duration causally connects extrinsic causal factors with increased intensity.

Phonological processes, though motivated by phonetic and cognitive factors, are abstract. Once a phonological process is part of speakers' grammars, the functional pressures that led to its existence do not necessarily have an active role in determining output. For instance, Kiparsky (2006) notes that a process such as /k/ → [tʃ] before front vowels is a typologically frequent rule in synchronic grammars, but is not phonetically motivated except as the culmination of a chain of sound-change processes [k] → [c] → [t̠ʲɕ] → [tʃ]. Each step along the chain is phonetically motivated, but the rule in the resulting grammar is not phonetically motivated in the strict sense. In the context of lenition,

* The work on this article owes much to many. Steven Sloman, Jonah Katz, and Chelsea Sanker provided significant input on earlier versions, and we discussed several aspects of this paper with Bertram Malle, Elinor Amit, Roman Feiman, Scott AnderBois, Emily Strand, and many others. The referees and the editorial board of *Language* contributed significantly with their helpful comments.

Kirchner (1998) expects that no strident fricatives would arise from stop lenition, and indeed such fricatives are rarely observed in variable lenition (e.g. Bouavichith & Davidson 2013). However, phonological rules in which stops spirantize to strident fricatives (or change their place of articulation) do exist, both synchronically and diachronically (e.g. some stage of Classical Hebrew synchronically: Gesenius 1910; German diachronically: Honeybone 2002). We therefore argue that the key to understanding the functional pressures that result in lenition is to focus on the effect different factors have on the duration, intensity, and articulatory properties of variable lenition. Several existing studies have taken this route (Turk 1992, Lavoie 2001, Warner & Tucker 2011, Bouavichith & Davidson 2013, Cohen Priva 2015, among others) and established that lenited forms tend to have shorter duration and higher intensity than nonlenited forms, that the environments that host lenition processes tend to host other nonphonologized reduction processes, that fast speech is correlated with greater reduction rates, and that low information content results in the reduction and omission of segments.

Two of these properties, reduced duration and increased intensity, are closely related to two overarching descriptions of lenition. Reduced duration corresponds to Vennemann's famous personal communication in Hyman 1975 that lenition is a stage toward deletion. Increased intensity corresponds to several theoretical approaches to lenition that view lenited forms as more vowel-like (Lass 1984, Szigetvári 2008) or sonorous (Smith 2008).[1] However, the causal relationship between these two factors and the established correlates of lenition is still unknown. Thus, it is difficult to understand the phenomena phonologists describe as lenition processes, or to even establish whether there are common properties among all lenition processes.[2]

One solution is to establish a foothold using factors that correlate with lenition but are independently motivated by forces unrelated to lenition, and proceed from those factors to the elements that characterize lenition. In this article, we consider three such factors: fast speech, proximity to stress, and low information content. In STUDY 1 we use the three independent correlates of lenition and perform a two-way mediation analysis using the Buckeye corpus (Pitt et al. 2007) between the three correlates and each of the characteristics of lenition (duration or intensity), using the other characteristic as a mediator. Though lenition has been argued to involve both durational reduction and increased intensity, we show that the correlates of lenition do not predict increased intensity directly in American English. Some or all of the established correlates of lenition correlate with durational reduction, but their correlation with increased intensity is mediated by reduction in duration. This finding unpacks two of the phonetic properties that are involved in lenition and implies that durational reduction, rather than increased intensity, is the necessary step en route to lenition. Such a finding changes the way phonetic reduction of low information content should be interpreted.

For completeness, we investigate the predictive power of treating reduced duration as a crucial step en route to lenition. STUDY 2 shows that all of the frequent lenition processes in the Buckeye corpus involve durational shortening: there are no variable lenition processes that involve durational lengthening, though most but not all fortition processes involve durational lengthening. Together these studies provide converging evidence that duration reduction is a likely cause rather than an effect of lenition processes.

[1] Sonority has been independently argued to correspond to intensity (Parker 2002), though the relationship is probably more complex (cf. Gordon et al. 2012).

[2] In order to avoid the causality implied by PREDICTORS and other such terms, we use the term CORRELATES to mean 'correlated with'. We do not use it as a short form of 'acoustic correlates'.

1.2. WHAT IS LENITION? Intuitively, lenition is a family of processes that involve reduction or weakening of segments in similar environments and under similar conditions, arranged in some sort of hierarchy or chain.[3] We focus here on phonological processes that seem to be in relative consensus, including degemination, voicing, spirantization, debuccalization, approximantization, tapping, and deletion. These processes are illustrated in Figure 1 (adapted from Hock 1986, Bauer 2008). We explicitly exclude processes that would not be considered lenition by many authors, such as affrication (Honeybone 2002) and final devoicing (Bauer 2008). We also exclude processes that are triggered by assimilation to a small set of segments, such as processes that are triggered by high vowels. While these processes may share many properties with lenition processes, they differ in several aspects, which has led many researchers to exclude them. The results of this study could be used to argue for or against the classification of such processes as lenition processes, but this is outside the scope of this article.

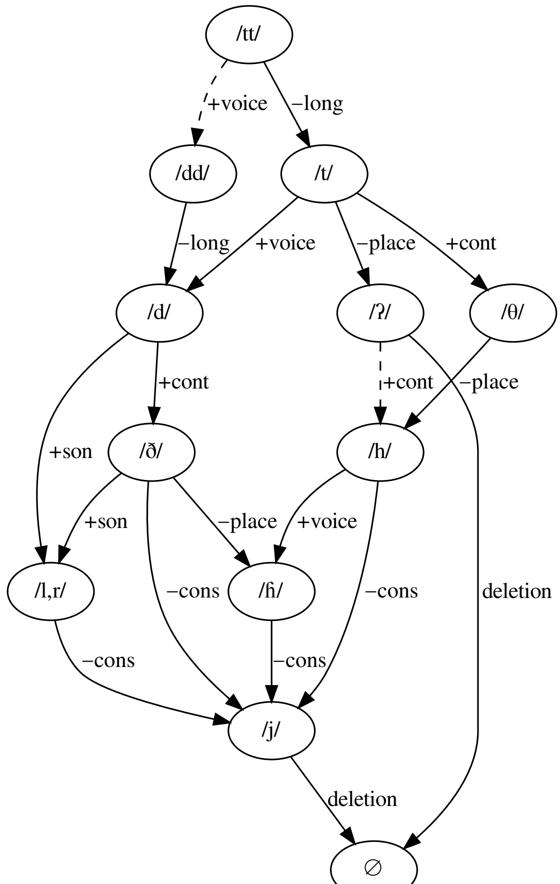

FIGURE 1. A range of processes commonly accepted as lenition processes. Dashed lines indicate typologically missing processes. IPA symbols represent a class of symbols; for example, /t/ stands for voiceless stops, /j/ for any glide. Feature values are added to ease interpretation. A lenition process may skip intermediate nodes (e.g. /h/ deletion). Adapted from Hock 1986, Bauer 2008. Vocalization patterns are not shown, though liquids may vocalize rather than delete (e.g. in Spanish; Proctor 2011).

[3] An excellent review of the history of the term *lenition* can be found in Honeybone 2008, which provides a more thorough discussion of the use of the term in linguistics as well as various definitions and supporting theories.

What unites these commonly agreed-upon disparate processes, aside from linguists' intuitions? The question has been the topic of a great deal of research and has split into two related discussions. First, what phonological properties characterize lenition? And second, what phonetic, communicative, or functional pressures cause lenition?

The focus of this article is the causal structure from which lenition follows, specifically the transition from the various conditions that promote lenition-type processes to the phonologization of lenition processes. However, to cover this discussion it is important to begin with the generally accepted properties of lenition and lenited forms. The prototypical environment for lenition, investigated in multiple recent studies, is the intervocalic environment (e.g. Lavoie 2001, Warner & Tucker 2011, Bouavichith & Davidson 2013). The influence of environment is so prominent that occurrence in intervocalic environment is considered by some to be a defining property of lenition (Katz 2016:49). Under a broader interpretation, which we use here, lenition most likely occurs in intervocalic positions but may also occur in prevocalic and postvocalic positions (Kirchner 1998, Kingston 2008). Studies such as Turk 1992 have shown that reduction also occurs in these environments for segments for which no phonologized lenition process exists, lending support to the idea that these environments are leniting environments.

Variable lenition processes have a number of common characteristics that have been observed by previous researchers.[4] In an in-depth study of the characteristics of lenited forms, Lavoie (2001) found that higher intensity, greater sonority, and shorter duration were all stable characteristics of lenited forms. Of these, duration was the most consistent indicator. However, studies with varying arguments focus on intensity as the main correlate of lenition. While we do not commit to any of these particular theories, this evidence together points to intensity and duration as the main perceptual correlates of lenition. For example, because lenition typically occurs in intervocalic positions, one recurring idea in the literature is that lenition makes a segment more vowel-like, with qualities such as increased intensity, sonority, and openness, similar to but separate from assimilation processes (Lass 1984, Smith 2008, Szigetvári 2008). Similarly, Harris (2003) argues that lenition is a loss of information, which blends the segment more with the background 'carrier signal', causing less disturbance between vowel sounds. Kingston (2008) builds off of this proposal and explicitly argues that lenition increases intensity of segments within prosodic constituents, reducing interruption of the speech stream. This argument is supported by evidence that the likelihood of lenition is affected by the identity of surrounding consonants (which vary greatly in intensity) but not the identity of surrounding vowels (which vary greatly in openness but not intensity). Katz (2016) uses OPTIMALITY THEORY (OT) to argue that a combination of intensity and duration constraints can result in both lenition inside prosodic constituents and fortition at prosodic boundaries. Studies that focus similarly on both factors find different results for intensity and duration in different environments and find that the two are not perfectly correlated (Lavoie 2001, Warner & Tucker 2011, Bouavichith & Davidson 2013). The absence of a perfect correlation is important, because it implies that the two may be independent of one another, or that one causally precedes the other.

At least two families of factors have been associated with promoting lenition: fast speech and low information content. Fast (and casual) speech is repeatedly mentioned as a condition promoting lenition (Kirchner 2004, Gurevich 2011, Warner & Tucker 2011, Cohen Priva & Gleason 2018). Fast speech may promote lenition for mechanistic

[4] Phonologized lenition processes are excluded from analysis here, not because they are EXPECTED to differ from variable lenition processes, but because they MAY involve additional changes.

reasons, for example, by creating difficulty in coordinating the articulators in a precise manner or limiting the time in which articulators have to move from one position to another.[5] Low information content is an umbrella term we use here for frequent and predictable contexts. Contextually predictable information can be reduced due to communicative or mechanistic pressures (see Jaeger 2010:50–51). Communicative pressures include, for example, the assumption that predictable linguistic material is easier to recover and can therefore be more reduced than less predictable material. Predictable material should also take less time to produce if language users attempt to keep the amount of information per second roughly smooth or uniform (Aylett & Turk 2004, Levy & Jaeger 2007). Mechanistic reasoning can apply to the reduction of contextually predictable linguistic material as well: predictable material may be more easily retrieved (Bell et al. 2009) and therefore produced in less time. Frequentist accounts of reduction may follow from the same principles that guide the reduction of predictable materials, as everything else being equal, frequent linguistic material would be more predictable. Alternatively, frequently used materials may be reduced for more mechanistic reasons that focus on usage, storage, and the quick activation of frequently used linguistic materials (Pierrehumbert 2001, Bybee 2002, Bell et al. 2009), all leading to more reduced articulation.

1.3. Causal reasoning and mediation analysis. Increased intensity and reduced duration are highly correlated with one another, and with a range of other correlates of lenition such as fast speech.[6] We adopt here a commonly held view, explored extensively by Hausman and Woodward (1999), Pearl (2000), Spirtes, Glymour, and Scheines (2000), and others, that causal relationship links can be inferred from correlations, if certain conditions apply. We focus on a key principle in such theories: conditional independence.

In causal frameworks, such as those based on Pearl 2000, when two factors *X* and *Y* are correlated, they are considered causally related: perhaps one causes the other (directly or indirectly), or they have a shared cause. In contrast, when two factors are independent from one another, such frameworks would argue that the two do not cause one another and do not have a shared cause. Furthermore, two factors *X* and *Y* can also be conditionally independent, when a third factor *Z* is held constant. In such cases, any causal relationship that may exist between *X* and *Y* is assumed to be indirect and apply via *Z*. Such evidence does not reveal whether *X* causes *Y* via *Z*, *Y* causes *X* via *Z*, or whether *Z* causes both *X* and *Y*. The three possible causal structures are illustrated in Figure 2. Disambiguating the three remaining causal structures requires different types of evidence, such as prior knowledge about the causal properties, temporal precedence, or direct manipulation.

Conditional independence lies at the heart of mediation analysis (Baron & Kenny 1986, Pearl 2012). The methods used in mediation analysis differ from the logical structure described above—since it is not possible to hold a factor constant, as in the case of natural speech, controlling for a factor (adding the factor to the model) is used as a proxy for holding it constant. The goal of a mediation analysis is to rule out specific causal

[5] We use the term mechanistic to refer to outcomes that do not follow from optimization-related processes but rather occur as byproducts of the architectural properties of speech. Thus, to argue that frequent words have shorter duration because speakers conserve effort (Zipf 1935) assumes goal-driven and possibly volitional optimization, whereas to argue that frequent words have shorter duration because they are more easily retrieved (Bell et al. 2009) assumes mechanistic reasons.

[6] Here and elsewhere, when the term correlation is used, there is no commitment to linear correlation.

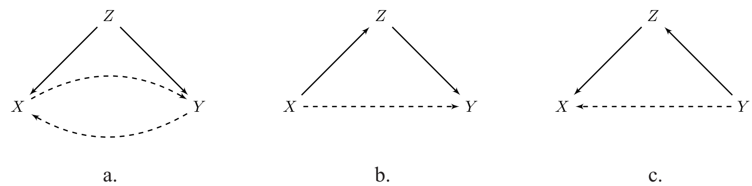

Figure 2. Three possible causal structures for three correlated factors, *X*, *Y*, and *Z*, if *X* and *Y* are not correlated when *Z* is held constant. Dashed lines are causal pathways that are ruled out as impossible by the conditional independence. The figure shows only the direct causation case, but the causal paths are not necessarily direct. In (a) *Z* causes *X* and *Y*, and there is no direct causal link between *X* and *Y*. In (b) *X* causes *Z*, and *Z* causes *Y*, but *X* does not cause *Y* except through *Z*. In (c) *Y* causes *Z*, and *Z* causes *X*, but *Y* does not cause *X* except through *Z*.

structures between three (or more) factors. The factor being controlled for is called a mediator, and the analysis is meant to assess whether that factor mediates the correlation between the other two. The procedure has three steps. First, each of the three factors is shown to be correlated with the other two. Second, a regression to predict one factor is performed, with both the mediator and the remaining factor as predictors. Finally, the results of the regression are evaluated. Complete mediation is found if the mediator affects the predicted factor, but the other remaining factor does not affect the predicted factor. This means that, conditional on the mediator, the two factors are independent, and therefore any causal relationship between them is dependent on the mediator.

We adopt this approach to examine the causal structure of lenition, focusing on its two prominent characteristics, short duration and increased intensity. This structure can be revealed by examining whether one characteristic mediates the effects of known correlates of lenition on the other characteristic. We focus on three factors that are correlated with lenition and have frequently been associated with causing its appearance: fast speech, proximity to stress, and low information content. Using the three factors we test whether one characteristic of lenition (reduced duration or increased intensity) mediates the other characteristic. In the rest of this section we discuss these three correlates of lenition and the ways they may be causally linked to reduced duration and increased intensity using existing evidence.

Fast speech is correlated with independent factors, such as age, task, and gender (Duchin & Mysak 1987, Kendall 2009, Jacewicz et al. 2010), and voluntary control of speech rate is a factor used in many studies (e.g. Dellwo et al. 2004, Malisz et al. 2018), which implies that lenition is not the sole predictor of fast speech. Fast speech has been demonstrated to contain more lenitions, segments with shorter duration, and more deletions than slower or more careful speech (Gay 1981, Dalby 1984, Browman & Goldstein 1990, Flemming 2004, Cohen Priva 2015), and Kirchner (2004) views fast speech as a cause for segment lenition. Fast speech is closely (almost definitionally) correlated with one of the characteristics of lenition, short segment duration.[7] We are therefore reluctant to decide whether fast speech causes shorter segment duration or the other way around. The same does not hold for the increase in segment intensity: it is not clear why increasing segment intensity should result in fast speech, but several causal accounts would predict increased intensity in fast speech (increasing likelihood of undershoot in undershoot-based accounts; increasing effort in effort-based accounts). We therefore argue that it is

[7] It should be noted that speech rate is not identical to consonant duration. For instance, fast speech predicts less than 40% of the variance in consonant duration in study 1, because consonant duration is affected by additional factors, and because fast speech may affect vowels more than it affects consonants.

unlikely that increased consonant intensity would causally lead to fast speech, but we make no causal argument with respect to speech rate and segment duration.

Leniting environments, such as intervocalic and postvocalic environments, are phonological environments that typically host lenition processes, though the specific effects may not be universal (Honeybone 2002). Proximity to stress in such environments has been shown to be critical as well for variable lenition processes (Turk 1992, Lavoie 2001, Warner & Tucker 2011, Bouavichith & Davidson 2013). Most researchers do not consider phonological environments to be dependent on lenition, and therefore would not consider phonological environments to be likely effects of lenition, but rather the reverse. Stress does appear to be sensitive to segment duration, not in terms of the actual segment duration but in terms of phonological weight (singleton vs. geminate, CV vs. CVC). Some evidence suggests that more fine-grained distinctions (e.g. low vowels vs. high vowels) could matter as well, particularly for meter (Ryan 2011). No current account would predict that increased intensity would attract stress, and current theories of stress assignment (e.g. Hayes 1995) do not include such pressures. Given the current understanding of stress and phonological environments, we are inclined to view syllable structure and the proximity to stress as causally leading (directly or indirectly) to lenition, and not the other way around. Throughout this article we discuss this factor as proximity to stress.

Low information content of segments has been the center of much research in the past two decades (e.g. van Son & Pols 2003, Pluymaekers et al. 2005, Aylett & Turk 2006, Raymond et al. 2006, Surendran & Niyogi 2006, Kuperman et al. 2007, Wedel et al. 2013, Hall et al. 2016), and its predecessors go back at least to Zipf (1929) and Hockett (1955). The basic logic is that frequent and predictable segments, or segments that contribute little to the contrast between a word and other words, can be safely 'reduced' articulatorily. Frequency is the segment's or word's frequency in language use, predictability is predictability in context (how surprised we should be to learn that a token starting with /bæ-/ ends with /t/ rather than /k/), and contrast from other words depends on the number and frequency of minimally different competitors ('minimal pairs'; Wedel et al. 2013). Every aspect of information content relies only on the structure of the lexicon and its usage: words, morphemes, and their frequency, rather than on the likelihood of individual segments to lenite in some context or another. Word-frequency-centric accounts (Pierrehumbert 2001) similarly argue that the reduced duration of frequent words stems from usage, not from existing lenition processes, making words that contain them more suitable for frequent use. It is therefore difficult to argue that low information content is driven by the existence of variable lenition. The converse, however, is well motivated, as discussed above, for both mechanistic and communication-oriented reasons.

Current research in linguistics supports the view of proximity to stress and low information content as causing lenition, rather than being caused by lenition. For fast speech, we argue, only one causal direction has been justified in the literature with respect to increased intensity, though for reduced segment duration there seems to be little support for causal directionality. These directional claims facilitate the argument we make in this article, but the analysis does not depend on them. In the discussion of study 1 we spell out the implications of the findings for those who reject the arguments made here.

The three factors that correlate with lenition could correlate with both increased intensity and reduced duration, but that does not have to be the case. Instead, either duration or intensity may causally precede the other: it may function as a precondition for the existence of the other. We are not implying that one must temporally precede the other to causally precede it—changes in duration and intensity likely happen at the same time. Rather, we raise the possibility that one is determined by the properties of the other, but not necessarily the other way around. If that is the case, then correlations with the other characteristic would not hold in mediation analysis. For instance, fast speech, proximity to stress, and low information content may not correlate with intensity when duration is controlled for. In such a case, causal structures that include direct influence between intensity and the three correlates of lenition would be ruled out. Figure 3 illustrates a possible difference between a causal model in which the independent correlates are causally linked to both increased intensity and reduced duration, and a different model in which they are causally linked only to reduced duration, and to increased intensity only via reduced duration.

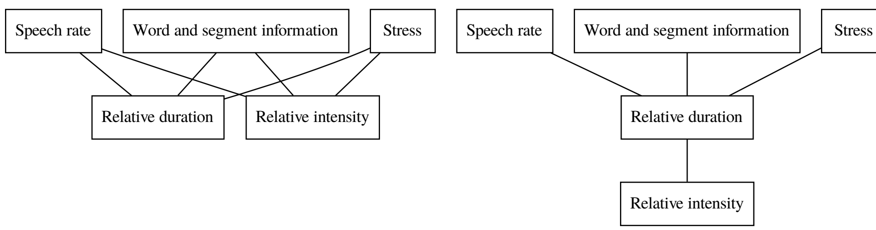

FIGURE 3. An illustration of two of the many possible causal alternatives between three correlates of lenition and the phonetic characteristics of lenited forms. On the left, relative duration and relative intensity are independently linked with the three correlates. On the right, the three correlates are causally linked with relative duration, but relative intensity is causally linked only with relative duration and with the three correlates indirectly. The graphs are undirected to express lack of information about the direction of causation.

1.4. DURATION AS A LIKELY CAUSE FOR LENITION. Both increased intensity and reduced duration have been repeatedly mentioned in the discussion of lenition. For example, Sievers (1876, via Honeybone 2008:43) argues that fortis and lenis forms are distinguished by 'lower energy and shorter duration'. Is there any reason to believe that one should precede the other? Historically, researchers of lenition have not made this distinction.

Lavoie (2001) classifies a group of lenition definitions as 'lenition as a decrease in the duration and magnitude of articulatory gestures', connecting gestural-reduction accounts directly to duration (e.g. as was posited directly in Lindblom 1963). Lavoie (2001:159) sums up the findings on duration:

> A major phonetic result of this research is that the main acoustic correlate of lenition is decreased duration. While the usual phonological correlate of lenition is said to be voicing, my data do not uniformly show additional vocal fold vibration in segments, but rather shorter duration. The shorter duration has been shown by other researchers to give rise to the percept of voicing. Additionally, the shorter durations may not provide enough time for speakers to reach articulatory targets, resulting in target undershoot. Shorter segments may also lack sufficient pressure build-up to produce stop bursts.

Lavoie and others (Fougeron & Keating 1997) reject the idea of a perfect correlation between duration reduction and gestural reduction, and do not directly argue for duration as a causal influence. Bauer (2008), who argues for lenition as articulatory undershoot, also states that changes that involve duration reduction should be considered lenition processes. In the context of effort reduction, reduced duration directly contributes to effort estimates (Kirchner 1998): reducing the need to hold contact is the very thing that makes some segments require less effort than other segments. However, Kirchner seems to assume that speakers' grammars choose different outputs for which both duration and intensity are affected 'at once'. Katz (2016), who argues for a prosodic-driven account of lenition, explicitly argues for two kinds of constraints leading to lenition, signifying distinct pressures to increase intensity and to reduce duration. However, in a study of Gurindji stops, Ennever, Meakins, and Round (2017) found no independent contribution of prosodic environment going beyond what is already captured by changes in duration (contrasting word-initial and word-medial contexts), suggesting a likely mediation of the effect of prosodic environment via change in duration (the prosodic environment they tested is not the same as the one used here and may not generalize to other languages; cf. Katz & Pitzanti 2019).

We believe increased intensity and reduced duration are not equally important contributors to lenition. First, fast speech may cause both reduced duration and increased intensity, but there is no doubt that reduced duration is immediately affected by speech rate, whereas increased intensity would only follow by intervening factors: for example, in the case of undershoot, intensity increase could be due to reduced time to get the articulators in place. Similarly, reduced duration follows straightforwardly from low-information-content conditions: at the word level, duration-reducing effects have been repeatedly demonstrated for high frequency (Bell et al. 2009), high predictability (Bell et al. 2009, Seyfarth 2014), and low informativity (Seyfarth 2014). At the segmental level, reduced duration has been demonstrated for both high predictability and low informativity (Cohen Priva 2015). An argument for increased intensity, as with fast speech, has to be more roundabout: there is no straightforward answer for why low information content would lead to increased intensity, except through, for example, more careless articulation. Reduced duration is therefore motivated directly, while increased intensity is motivated indirectly.

Though the idea that reduced duration causally precedes increased intensity in lenition is mechanism-agnostic, its most natural association is with views that regard lenition as undershoot (especially Bauer 2008). If the time allotted for the articulation of a segment is reduced, there is a simple and direct reason for the speaker to undershoot that segment (less time resulting in partial gestures). It is important to note, though, that time plays an important role in other theories of lenition, such as lenition as effort reduction (Kirchner 1998), in which shorter-duration gestures are regarded as less effortful.

There are therefore several independent reasons to believe that reduced duration is in some sense more basic compared to other aspects of lenition. The studies in the next section aim to investigate this possibility.

2. STUDIES.

2.1. OVERVIEW. Following from the discussion in §1.3 on mediation analysis, the hypothesis that short duration causally precedes other aspects of lenition, such as intensity, makes at least two predictions. These are expected to hold categorically or as statistical trends:

(i) The three independent correlates of lenition SHOULD NOT be correlated (or should show a weak correlation) with relative intensity if the relative duration of the segment is included as a predictor in the model. In contrast, the independent causes SHOULD be correlated with relative duration, even if relative intensity is controlled for.

(ii) Nonphonologized lenition processes should involve reduced duration. Phonologized lenition processes could be exempt from this requirement, if subsequent processes have further changed the output (e.g. /p/ → $_{\text{variable}}$[ɸ] → $_{\text{categorical}}$[f]).

Investigating these questions requires a detailed corpus of conversational speech, and for this reason we use the Buckeye corpus (Pitt et al. 2007).

2.2. The corpus. For all of the studies reported below, we use the Buckeye Corpus of Conversational Speech (Pitt et al. 2007), which provides data collected from forty speakers at The Ohio State University conversing 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 by the procedure detailed in the appendix. The goal of the procedure was to align underlying dictionary forms with their surface realization, as transcribed in the corpus. For instance, if the word *backs* /bæks/ surfaced as [bɜz], the procedure would align /b/ with [b], /æ/ with [ɜ], and /s/ with [z], and would regard /k/ as deleted. To do so, the algorithm was trained on the entire corpus to deduce which correspondences and deletions were more likely than others. The underlying representations provided by Buckeye were not used and were replaced with their CMU equivalents (*The CMU pronouncing dictionary*; Weide 2008), as Buckeye does not provide information about stress, and previous studies of segmental duration (e.g. Lavoie 2001, Bouavichith & Davidson 2013) have shown that such information is relevant to lenition and duration.

As is the case with most edit-distance implementations (that are based on a variant of Wagner & Fischer 1974), the algorithm we used could only handle segment-to-segment correspondences. Because of this, if an underlying segment was realized using two surface segments (e.g. /ɚ/ → [ɪɹ]), the more common substitution was considered the surface form, and the other segment was considered to have been epenthesized.[8] Similarly, when two underlying segments merged, the merged output was considered to be the surface realization of one of them, and the other segment was considered to have been deleted (the choice depended on which substitution is more probable and which segment is more likely to delete). The method is very similar to the one used in Cohen Priva 2015, except our algorithm learned the likely and unlikely substitutions from the corpus and was not provided with a list of likely substitutions. The main reason for this difference is that the corpus could be (and is) richer than any list of substitutions that we imagined we could find, and we did not want to presuppose which processes were more probable.

2.3. Measuring intensity. A fundamental methodological question when dealing with segment intensity revolves around the means by which irrelevant factors such as speakers' baseline intensity or distance from the microphone can be removed from the measurement. Previous researchers such as Warner and Tucker (2011) have done so by comparing a consonant's minimum measured intensity with the surrounding vowels' maximal intensity. This method deals with many of the shortcomings mentioned above, but it is limiting in this case. First, the intensity of the surrounding vowels can change independently from the intensity of the consonant. Stress, word frequency, and speech rate are all likely to affect vowel intensity regardless of their effect on consonant intensity. For instance, if frequent words have lower-intensity vowels, the intensity of the adjacent consonants would be measured as relatively more intense regardless of whether their intensity actually changed. A different issue is the absence of per-phoneme baselines: fricatives are overall more intense and have shorter duration than stops, every-

[8] This limitation does not apply to the affricates /tʃ/ and /dʒ/, which are represented by a single symbol in Buckeye and in the CMU dictionary. Cases of a single segment being realized as two segments almost always involved a rhotic vowel being realized as a nonrhotic vowel + [ɹ]. In these cases, it was likely that the rhotic vowel was aligned with [ɹ], and the surface vowel was regarded as epenthesized.

thing else being equal. Pulling them together into a single regression model would predict that short duration correlates with higher intensity regardless of whether the intensity of stops or fricatives increases when their duration is reduced, simply because overall fricatives demonstrate these two properties.

To overcome these two limitations, we measured segment minimum intensity by comparing each segment with the mean minimum intensity of all other tokens of the same underlying phoneme produced by that speaker in the same Buckeye file. If a certain speaker's file contained three values for the intensity of intervocalic /g/: −50, −48, and −46 dB, then the first /g/ would be compared relative to a baseline of −47 (the mean of −48 and −46, excluding the token itself), the second /g/ would be compared relative to a baseline of −48 (mean of −50 and −46), and the final /g/ would be compared relative to a baseline of −49 (mean of −50 and −48).

To test whether the method we use is better at predicting actual values than previous methods were, we compared the CONSISTENCY of the methods. For a method to be consistent, it should have a high correlation between its measured values and the mean (or mode, or median) of those values. The correlation between the values and their mean is zero when the values are random, and one when there is no variance in the measured values. The higher the correlation, the better the method. For the method we used, the correlation was quite high (Pearson $r = 0.761$), higher ($p < 0.001$) than the correlation for the method used in previous studies (Pearson $r = 0.656$), suggesting that our approach was reliable. It is important to note that our approach reaches high consistency only if values are measured PER-FILE (or per-speaker): if intensity is averaged per-phoneme without taking file or speaker into account, the consistency of our method drops significantly (Pearson $r = 0.569$), worse than the previous standard ($p < 0.001$).

2.4. STUDY 1: INTERVOCALIC LENITION PROCESSES.

INTRODUCTION. Intervocalic environments are the prototypical environments for consonant lenition (Katz 2016). We consider here three independent correlates of lenition: (i) fast speech, (ii) proximity to stress, and (iii) low information content. Each of those should correlate with both shorter duration and increased intensity if short duration and increased intensity are independent from one another. However, if one causally preceded the other, we should observe a correlation with only one of the two when the effects of the other factor are controlled for.

To approach this problem we use mediation analysis. The general outline of the procedure we use follows Baron and Kenny (1986, and Pearl 2012 for linear models), with a few noteworthy differences. First, traditional mediation analysis assumes that there are empirical reasons to identify each of the three variables as cause, possible mediator, and effect, which is not the case here, as we do not know whether reduced duration is likely to causally precede increased intensity or the other way around. We therefore conduct the analysis in two different ways: in one, we consider the possibility that duration mediates intensity, and in the other, we consider the possibility that intensity mediates duration. Second, the original Baron and Kenny (1986) approach applies to simple correlations between three factors, whereas our study deals with multiple correlates with many repeated measures. Our basic models are therefore not models of simple correlations, but mixed-effect linear regressions (as in Freeman et al. 2017). Finally, we add several diagnostics for the strength of the mediation, which generalize and add to the traditional approaches, by modeling the regressions jointly, using the brms package (Bürkner 2017, 2018) in R (R Core Team 2018).

For each of the two characteristics of lenition (duration and intensity), we attempt to check whether one characteristic causally precedes the other by checking whether one of them mediates the effect of the known correlates of lenition on the other.

Method and materials. We collected all nondeleted intervocalic obstruents in the Buckeye corpus that had the following additional properties: they (i) were not in function words, (ii) were in utterances with at least four words to exclude short backchannels, (iii) had either a primary stressed vowel or unstressed vowel following, and (iv) were in words with a CMU dictionary (Weide 2008) representation (for stress information). This resulted in fourteen different phonemes and ~12,600 tokens (/b/: 8%, /tʃ/: 2%, /ð/: 7%, /f/: 6%, /g/: 3%, /dʒ/: 3%, /k/: 11%, /p/: 18%, /s/: 11%, /ʃ/: 5%, /θ/: 1%, /v/: 18%, /z/: 6%, /ʒ/: 1%), from ~620 word types and forty speakers. Information-theoretic properties were calculated using word counts from the Buckeye (Pitt et al. 2007), Switchboard (Godfrey & Holliman 1997; using Mississippi State (MS-State) corrected annotations: Harkins et al. 2003), and Fisher (Cieri et al. 2004, Cieri et al. 2005) corpora. Information-theoretic variables were not smoothed. Individual variables are listed below.

- Segment duration: Hand-corrected annotations made in Buckeye.
- Segment mean duration: Calculated for each underlying segment as the geometric mean duration of its surface tokens, excluding the particular segment token for the calculation of the mean.
- Segment relative duration: Calculated as the log ratio between the segment's actual duration and mean duration of segments with the same underlying form.
- Minimum segment intensity: Calculated using Praat (Boersma & Weenink 2008) using the default settings for each word separately, and assigned for each segment using the timestamps provided in Buckeye.
- Segment mean minimum intensity: Calculated for each segment token as the mean minimum intensity of all surface tokens associated with the same intervocalic underlying segment produced by the same speaker, excluding the token segment itself (see explanation in §2.3).
- Segment relative minimum intensity: Calculated as the difference between the segment's minimum intensity and its mean minimum intensity. Differences were used rather than log ratios because intensity is on a log scale.
- Stress: Binary; indicates whether the following vowel had primary stress. Stress was assigned using the CMU dictionary (Weide 2008).
- Pointwise speech rate: The ratio between the word's actual duration and the mean duration of all instances of that word (excluding the word token in question) in the entire corpus (Cohen Priva 2017b). This means that slow speech rate would have high values: a word whose mean duration was 400 ms produced in 500 ms would have a pointwise speech rate of 1.25, but if its duration had been 300 ms, its pointwise speech rate would have been 0.75. Pointwise speech rate was log-transformed.
- Word frequency: Calculated as the number of times each word was observed in the three corpora, and log-transformed, as is standard practice in information-theoretic approaches to linguistics (cf. Jaeger & Buz 2017).
- Segment contextual probability: Calculated as the negative log probability of observing that segment conditioning on the previous segments. For example, the conditional probability of /t/ in /bæt/ was calculated as the number of times the string /$[_w$bæt/ occurred in the three corpora, divided by the number of times /$[_w$bæ/ occurred in the three corpora, negative log-transformed following standard practice. The corpora are big enough to yield reliable estimates for contextual probability (Cohen Priva & Jaeger 2018).
- Number of minimal pairs: Calculated using the underlying representations from the CMU dictionary, using only words that occurred in one of the three corpora.

For each word, we calculated the number of words that differed only by a substitution of the segment in question. For instance, for the /t/ in *meeting*, that number was 1 (due to *meaning*). For the /s/ in *basic*, that number was 0. For the purpose of the regressions, 1 was added to each number before it was log-transformed.

- Distance from word edges: Calculated using the underlying representation, and log-transformed (both distances start at 1).

Mixed-effects linear regressions were trained in R (R Core Team 2018) using the lmerTest package (Kuznetsova et al. 2017), which encapsulates the lme4 package (Bates et al. 2015) and supplies degrees of freedom for the calculation of *p*-values using Satterthwaite's method.[9] Both models used the segment's predictability in context, log word frequency, pointwise speech rate, and whether the following vowel was stressed as fixed variables of interest. The log distances from both word edges were used as additional controls, as both stress and contextual predictability are sensitive to their position within a word. Word, speaker, and underlying segment identity were used as random intercepts. All variables were standardized.[10] Standardization helps in the evaluation of the mediated effects and facilitates model convergence.

The data does not support using a maximal random-effects structure (Barr et al. 2013) due to the large number of predictors and the small number of data points for some individual groups. An important concern here is that if the random-effects structure is overspecified, it would keep us from identifying the variables of interest: the Baron and Kenny (1986) procedure relies on the lack of significance of the suspect causes when controlling for the mediator, and we would not want to inflate the odds of type 2 errors, which may happen when an overspecified random-effects structure is used (Matuschek et al. 2017). We therefore followed the procedure outlined below to identify the richest random-effects structure that could be supported by the data, using model comparison. First, the two fully specified models (using all of the correlates and controls) were evaluated with all random intercepts (word, speaker, and phoneme), but without random slopes. Second, by-speaker and by-phoneme random slopes were added for each variable if they significantly improved the model, without calculating the correlation between random effects. The slopes were added in a greedy manner, by adding the slope that significantly improved the model the most (using likelihood ratio tests; see Matuschek et al. 2017).[11] Third, if including a variable and all associated random slopes did not significantly improve the model (in a likelihood ratio test), the associated random slopes were removed from the model, and only the fixed variable was kept. Finally, if a minimally different model, in which correlations between random effects were evaluated, showed significant improvements, we used that model. The model-comparison criterion was evaluated using the rather stringent $\alpha = 0.05$, since greater alpha levels would make it easier to find complete mediation. Using a less stringent criterion of $\alpha = 0.2$ (argued for in Matuschek et al. 2017) did result in a model with

[9] We also refitted the models such that degrees of freedom were calculated using Kenward-Roger's method, but the models were nearly identical and are not reported here.

[10] For binary variables, standardization involved replacing True with 1 and False with 0, prior to standardization.

[11] The procedure we use for adding random slopes is a greedy forward-selection method, unlike the greedy backward-selection method used by Matuschek et al. (2017). We were forced to use forward-selection rather than backward-selection because the full model did not converge. The goals of the two procedures are identical: to find the maximal random-effects structure that is supported by the data. The following steps are similar to steps 4 and 5 in the model-selection procedure described by Bates et al. (2018) and have the same goal.

additional slopes, but the results were essentially the same. The two resulting models are provided below using the lme4 syntax. The duration model does fit the correlations between the random effects in formula 1, but the full intensity model was not improved by the inclusion of correlation between random effects in formula 2. We refer to these two models as SATURATED, as they contain the most variables that the data can support, and to distinguish them from other models.

(1) segment relative duration ~ segment relative intensity +
segment predictability in context +
word frequency +
minimal pairs +
speech rate +
stress follows +
distance from word beginning +
distance from word end +
(1 | word) +
(1 + distance from word end + word frequency + stress follows | speaker) +
(1 + speech rate + stress follows + segment relative intensity | segment)

(2) segment relative intensity ~ segment relative duration +
segment predictability in context +
word frequency +
minimal pairs +
speech rate +
stress follows +
distance from word beginning +
distance from word end +
(1 | word) +
(1 | speaker) +
(1 | segment) +
(0 + segment relative duration | segment) +
(0 + segment relative duration | speaker)

To establish possible mediation, we also calculated minimally different models, in which the other characteristic (duration for intensity, intensity for duration) was removed from the saturated model, along with any associated random slopes. We label these models UNSATURATED to distinguish them from the saturated models.

We use three measurements to evaluate mediation. In order to make the explanation clearer, we describe only the procedure for the model in which we test whether reduced duration mediates increased intensity. We also performed the inverse analysis to test whether increased intensity mediates reduced duration, and both results are reported below.

(i) THE BARON AND KENNY (1986) PROCEDURE: We used the two unsaturated models to establish whether each correlate of lenition predicts reduced duration and increased intensity. We then used the intensity-predicting saturated model in formula 2 to check whether the correlates predict relative intensity even when relative duration is controlled for.

(ii) CALCULATING MEDIATION SIZE: We calculated the ratio between the mediated effect and the direct effect. The direct effect is simply the coefficient of the correlate in the saturated model (e.g. the direct effect of word frequency on intensity is the regression estimate of word frequency in the saturated model that predicts intensity). The indirect effect is calculated as the product of the coefficient of the correlate in the unsaturated model to predict the mediator and the coefficient of the mediator in the saturated model predicting the outcome (e.g. the indirect effect of word frequency on intensity is the estimate of word frequency in the unsaturated model to predict relative duration multiplied by the estimate of duration in the saturated model to predict relative intensity). High ratios would signal complete mediation. To calculate the ratio jointly, we recalculated the saturated relative intensity and unsaturated relative duration models jointly using the brms package (Bürkner 2018) and used the sampled regression coefficients to calculate the 95% credible interval (CI) of the ratio. These values cannot be straightforwardly calculated in case of sign reversal, and when a coefficient changed its sign we discarded it.[12] Results are shown in percentages, rather than ratios, to conform with existing literature and ease interpretation.

(iii) MEDIATION MODEL COMPARISON: We used the brms package to perform model comparison between pairs of models. We contrasted the mediation model used in the calculation of mediation size (which had an unsaturated duration model and a saturated intensity model) with a FULL MEDIATION model in which the saturated intensity model was replaced by an intensity model in which all fixed effects except relative duration were removed, as well as their random slopes. Full mediation represents the possibility that the correlates of lenition do not contribute to the prediction of relative intensity if relative duration is included in the model. The standard mediation model and the full mediation model were both evaluated in brms, and we report the difference between them using LEAVE-ONE-OUT cross-validation (Vehtari et al. 2017, Vehtari et al. 2018). This approach is based on piecewise structural equation modeling (Shipley 2000).

In the following three sections we present the results. We first introduce the unsaturated models, which serve to establish which of the correlates of lenition predicts reduced duration and increased intensity. We then introduce the two competing saturated mediation models. In the first, relative duration serves as the mediator, and in the second, relative intensity serves as the mediator.

UNSATURATED MODELS RESULTS. When relative duration was not controlled for, many of the correlates of lenition significantly correlated with obstruent intensity. These included an increase in intensity for frequent words ($\beta = 0.06$, $SE = 0.02$, $t = 2.7$, $p = 0.006$), a decrease for slow speech rate ($\beta = -0.13$, $SE = 0.009$, $t = -14.7$, $p < 0.001$; high pointwise speech rate means slower speech), a decrease when stress followed ($\beta = -0.09$, $SE = 0.01$, $t = -6.4$, $p < 0.001$), and an unexpected increase the further the segment was from word end ($\beta = 0.04$, $SE = 0.02$, $t = 2.5$, $p = 0.01$). Other predictors did not have a significant effect on intensity. These include predictability in context ($\beta = -0.0003$, $SE = 0.02$, $t = -0.02$, $p = 0.98$), number of minimal pairs ($\beta = 0.02$, $SE = 0.02$, $t = 0.9$, $p = 0.36$), and distance from word beginning ($\beta = -0.02$, $SE = 0.02$, $t = -1.0$, $p = 0.30$). The results are summarized in Table 1.

Similarly, when relative intensity was not controlled for, many of the correlates of lenition significantly correlated with obstruent duration. Unpredictable segments in context correlated with longer relative duration ($\beta = 0.04$, $SE = 0.02$, $t = 2.2$, $p = 0.03$), frequent words correlated with shorter relative duration ($\beta = -0.10$, $SE = 0.03$, $t = -3.8$,

[12] Discarding cases of sign reversal is conservative in this case, because the direct effects were always more likely to be switch sign. This is not always the case in mediation analysis.

	EST	*SE*	*df*	*t*-VALUE	*p*-VALUE
Segment predictability in context	−0.00034	0.0174	779.20	−0.02	0.9844
Word frequency	0.05854	0.0213	567.15	2.74	0.0063
Minimal pairs	0.01622	0.0178	565.98	0.91	0.3614
Pointwise speech rate	−0.12942	0.0088	12359.68	−14.70	< 0.0001
Stress follows	−0.08617	0.0134	849.70	−6.44	< 0.0001
Distance from word beginning	−0.01979	0.0191	581.92	−1.03	0.3013
Distance from word end	0.04198	0.0168	652.81	2.49	0.0129

TABLE 1. Summary table for study 1 model predicting relative intensity without controlling for relative duration.

$p < 0.001$), slow speech correlated with longer relative duration (β = 0.39, $SE = 0.03$, $t = 14.7$, $p < 0.001$), and following stress correlated with longer relative duration (β = 0.26, $SE = 0.06$, $t = 4.7$, $p < 0.001$). As with the intensity models, distance from word end unexpectedly led to shorter relative duration (β = −0.11, $SE = 0.02$, $t = -5.1$, $p < 0.001$). The number of minimal pairs and distance from word beginning had no significant effect on relative duration (β = −0.01, $SE = 0.02$, $t = -0.7$, $p = 0.49$; β = 0.009, $SE = 0.02$, $t = 0.4$, $p = 0.67$, respectively). The results are summarized in Table 2.

	EST	*SE*	*df*	*t*-VALUE	*p*-VALUE
Segment predictability in context	0.0411	0.019	936.64	2.21	0.0272
Word frequency	−0.1024	0.027	354.65	−3.84	< 0.0001
Minimal pairs	−0.0140	0.020	580.80	−0.69	0.4925
Pointwise speech rate	0.3867	0.026	9.50	14.72	< 0.0001
Stress follows	0.2638	0.056	11.05	4.68	< 0.0007
Distance from word beginning	0.0089	0.021	799.00	0.42	0.6739
Distance from word end	−0.1070	0.021	210.24	−5.08	< 0.0001

TABLE 2. Summary table for study 1 model predicting relative duration without controlling for relative intensity.

Though distance from word end correlated with both intensity and duration, we are not aware of intervocalic processes that depend on distance from word end (though there are many processes that occur AT word-final position). This factor, which we added as a control, could be causally linked to effects such as COMPRESSION (White 2002), due to its correlation with word length. We do not consider this factor as one of the established correlates of lenition and therefore keep our discussion of its patterns with respect to intensity and duration minimal.

The results confirm the precondition of the mediation analysis: speech rate, whether the following vowel was stressed, and one aspect of information content are correlated with both reduced duration and increased intensity. The question now is whether these correlates of lenition would still be correlated with intensity and duration when the other characteristic of lenition is controlled for in mediation analysis.

DURATION-AS-MEDIATOR RESULTS. The regression model for relative intensity that controlled for relative duration had dramatically different results. Segment relative duration was highly predictive of relative intensity, with shorter productions correlating with increased intensity, as expected (β = −0.29, $SE = 0.04$, $t = -8.1$, $p < 0.001$). However, none of the other predictors had a significant effect on relative intensity, including predictability in context (β = 0.01, $SE = 0.02$, $t = 0.6$, $p = 0.53$), word frequency (β = 0.02, $SE = 0.02$, $t = 1.3$, $p = 0.21$), number of minimal pairs (β = 0.006, $SE = 0.02$, $t = 0.4$, $p = 0.71$), speech rate (β = −0.01, $SE = 0.01$, $t = -1.5$, $p = 0.14$), following stress (β = 0.005, $SE = 0.01$, $t = 0.4$, $p = 0.72$), and distance from word beginning (β = −0.02, $SE = 0.02$, $t = -1.1$, $p = 0.25$) or end (β = 0.02, $SE = 0.02$, $t = 1.3$, $p = 0.20$). These results are summarized in Table 3.

	EST	*SE*	*df*	*t*-VALUE	*p*-VALUE
Segment relative duration	−0.2941	0.0362	16.68	−8.12	< 0.0001
Segment predictability in context	0.0102	0.0161	871.15	0.64	0.5252
Word frequency	0.0243	0.0192	497.77	1.27	0.2061
Minimal pairs	0.0060	0.0159	470.56	0.38	0.7072
Pointwise speech rate	−0.0143	0.0098	10627.19	−1.46	0.1438
Stress follows	0.0046	0.0127	874.53	0.36	0.7166
Distance from word beginning	−0.0202	0.0177	672.01	−1.14	0.2526
Distance from word end	0.0198	0.0153	596.27	1.29	0.1971

TABLE 3. Summary table for study 1 model predicting relative intensity while controlling for relative duration.

One possible concern is that lack of statistical significance may not mean that the correlates had no effect on the outcome. There is an important distinction between an effect being small and not significantly different from zero, and large and not significantly different from zero. The coefficient for word frequency in the duration-mediated model (the direct effect of word frequency) is not significantly different from zero, but is still approximately half the size of the correlation in the unsaturated model ($\beta = 0.024$ vs. $\beta = 0.059$). In contrast, the coefficient for stress on the following vowel in the duration-mediated model (the direct effect of stress) changed its sign relative to the original unsaturated model ($\beta = 0.0046$ vs. $\beta = -0.086$). Though neither effect is significantly different from zero in the duration-mediated model, it would be easier to argue that the effect of stress on the following vowel was completely mediated by reduced duration than to make the same argument for word frequency. To measure the degree of mediation, we followed current practice (e.g. Freeman et al. 2017) and calculated the ratio between the direct effect and the mediated effect (as explained in §2.4). For speech rate and whether the following vowel was stressed, the percentage of the explained variance via the indirect path (through duration) was decidedly larger than in the direct path (median: 88.1%, CI: [75.1%, 98.8%]; median: 91.2%, CI: [69.3%, 99.6%], respectively), suggesting a strong case of complete mediation.[13] The same was not true for word frequency (median: 53.3%, CI: [26.8%, 94.3%]) and distance from word end (median: 59.4%, CI: [34.2%, 94.4%]), for which, in the majority of the cases, the indirect path was larger, but the model samples contained many cases in which the relative significance was reversed. This suggests that despite the absence of a significant effect for word frequency and distance from word end, they may still be causally linked to changes in intensity. Figure 4 shows the density plots of the direct and indirect effects.

Finally, we used model comparison to compare the original duration-mediated model (which compared indirect and direct effects) with a full mediation model (in which only relative duration, relative duration slopes, and random intercepts could affect intensity), as described in §2.4. We used leave-one-out (LOOIC; Vehtari et al. 2017) to perform model comparison. The models were only numerically different from one another, as the standard error of the LOOIC difference was greater than the measured difference (ELPD diff: −2, $SE = 3.3$).[14] This means that the full mediation model is not better or worse than the model in which all of the correlates of lenition were used to predict relative intensity.[15] If a parsimony bias is applied then it should be assumed that changes in duration completely mediate changes in intensity in variable lenition.

[13] On his website, David Kenny suggests treating mediations of 80% and above as 'complete' (http://davidakenny.net/cm/mediate.htm, retrieved April 2019).

[14] ELPD: expected log predictive density

[15] We also tried to compare the full mediation model with two other semi-complete mediation models. In one, only word frequency and distance from word end were left in the model, alongside relative duration. In

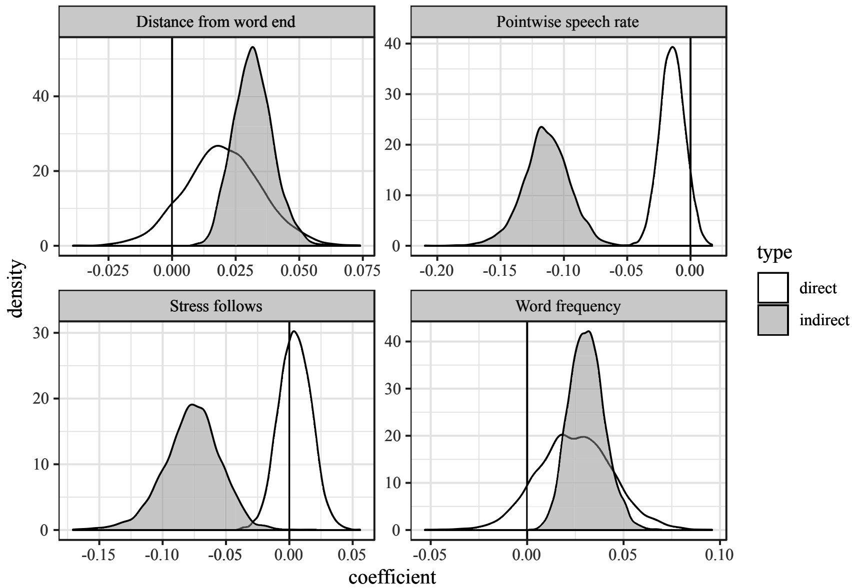

FIGURE 4. Density plots for the direct and indirect effects of distance from word end, speech rate, whether the following vowel was stressed, and word frequency on relative intensity. Though for all measures the indirect effect is more robust, only speech rate and stress exhibit complete mediation patterns. The indirect path is marked in gray, the direct path in white, and the dark vertical line marks zero (having no effect).

These results strongly support the possibility that increase in intensity, which is typical of variable lenition but also of phonologized lenition, does correlate with fast speech rate, prosodic position (prestress), and information-theoretic redundancy (at the word level). However, these results are mediated by changes in duration: when changes in duration are controlled for, these established correlates of lenition are no longer significantly correlated with intensity. For segment predictability, speech rate, and whether the following vowel was stressed, there is likely no residual direct effect, and changes in intensity depend only on changes in duration. For word frequency and distance from word end, there may be a case for a weak remaining direct effect.

Furthermore, the saturated duration-mediated model was not better than a model in which none of the correlates could affect relative intensity (complete mediation). To sum up, there is little evidence for direct interaction between the correlates of lenition and changes in segment intensity, though word frequency and distance from word end may have a small residual correlation.

INTENSITY-AS-MEDIATOR RESULTS. Unlike the results presented for duration as a mediator of intensity, the inclusion of segment relative intensity did not have such a dramatic effect on the prediction of relative duration. Even when intensity was controlled for, most of the predictors significantly affected segment duration in the expected direction. High intensity correlated with shorter relative duration ($\beta = -0.19$, $SE = 0.02$, $t = -8.9$, $p < 0.001$), which is to be expected regardless of the causal direction between relative duration and relative intensity. Among the correlates of lenition and controls, unpredictable

the other, word frequency was omitted, but both word-edge controls were included, alongside relative duration. Both models were not distinguishable from the reported mediation model (which does not omit any factors).

segments in context correlated with longer relative duration ($\beta = 0.04$, $SE = 0.02$, $t = 2.5$, $p = 0.01$), frequent words correlated with shorter relative duration ($\beta = -0.09$, $SE = 0.03$, $t = -3.6$, $p < 0.001$), slow speech correlated with longer relative duration ($\beta = 0.36$, $SE = 0.03$, $t = 13.7$, $p < 0.001$), and following stress correlated with longer relative duration ($\beta = 0.22$, $SE = 0.05$, $t = 4.4$, $p < 0.001$). As with the intensity models, distance from word end unexpectedly led to shorter relative duration ($\beta = -0.10$, $SE = 0.02$, $t = -4.9$, $p < 0.001$). The number of minimal pairs and distance from word beginning had no significant effect on relative duration ($\beta = -0.01$, $SE = 0.02$, $t = -0.6$, $p = 0.53$; $\beta = 0.004$, $SE = 0.02$, $t = 0.2$, $p = 0.84$, respectively). The results are summarized in Table 4.

	EST	*SE*	*df*	*t*-VALUE	*p*-VALUE
Segment relative intensity	−0.1859	0.021	13.13	−8.87	< 0.0001
Segment predictability in context	0.0428	0.017	924.16	2.45	0.0146
Word frequency	−0.0891	0.025	345.89	−3.58	< 0.0001
Minimal pairs	−0.0120	0.019	568.99	−0.63	0.5283
Pointwise speech rate	0.3595	0.026	9.37	13.72	< 0.0001
Stress follows	0.2211	0.051	11.53	4.38	< 0.0010
Distance from word beginning	0.0042	0.020	786.47	0.21	0.8351
Distance from word end	−0.0978	0.020	194.58	−4.89	< 0.0001

TABLE 4. Summary table for study 1 model predicting relative duration while controlling for relative intensity.

The Baron and Kenny procedure therefore provides no evidence for complete mediation of duration by intensity. However, it is still interesting to investigate the extent to which the other two measures we used to study duration as a mediator would apply to intensity as a possible mediator. For all of the correlates of lenition and controls that were correlated with duration in the intensity-mediated model, the percentage of the mediated effect was very small. These include segment predictability in context (median: 4.7%, CI: [0.2%, 23.9%]), speech rate (median: 6.3%, CI: [4.3%, 8.6%]), whether the following vowel was stressed (median: 6.7%, CI: [3.9%, 12.8%]), word frequency (median: 10.5%, CI: [3.1%, 24.7%]), and distance from word end (median: 7.5%, CI: [2%, 15.9%]). These results show that changes in intensity do not mediate changes in duration in variable lenition, and the 'direct' effects in the original unsaturated intensity model are likely due to what Judd and Kenny (2010) call 'specification error': performing mediation analysis on a false premise of a causal structure. Figure 5 shows the density plots of the direct and indirect effects. Model comparison between the intensity-mediated model and its full mediated counterpart showed that the full mediated model's predictive power is much reduced (ELPD diff: −1314, *SE*: 52.5). In other words, all three methods decisively show that changes in intensity do not mediate changes in duration.

DISCUSSION. The results of the mediation analysis indicate that shorter duration screens off or mediates the correlation between three known correlates of lenition and increased intensity. They suggest that one of three causal structures is possible, for each of the correlates: (a) reduced duration causes both increased intensity and the correlate, (b) increased intensity causes reduced duration, and reduced duration causes the correlate, or (c) the correlate causes reduced duration, and reduced duration causes increase in intensity (see discussion in §1.3). The alternatives are illustrated in Figure 6. The causal structures of the three correlates are not independent of one another, as some causal structures would not be compatible with others. For instance, if (b) holds for any of the correlates of lenition, then (a) and (c) cannot hold for any other factor, as the causal direction between changes in intensity and changes in duration would need to be reversed. In contrast, it is possible for (a) to coexist with (c) for different correlates; for

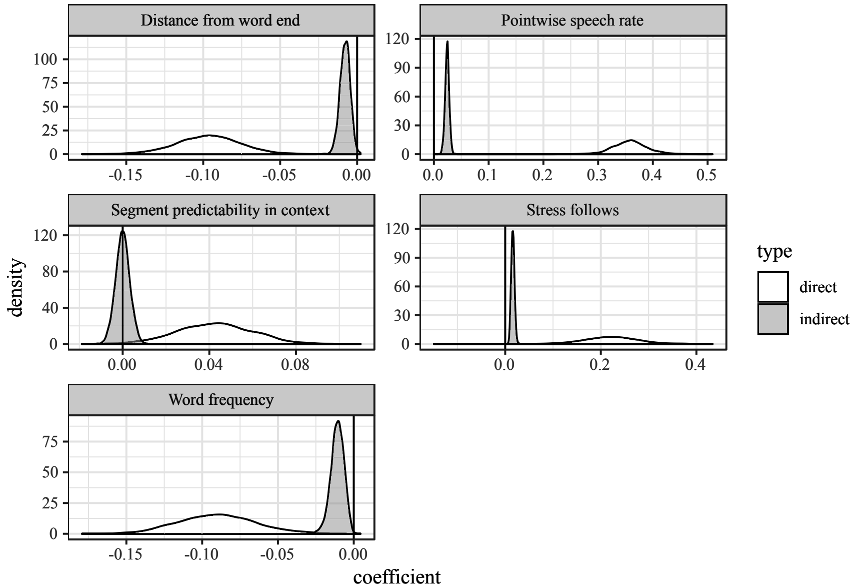

FIGURE 5. Density plots for the direct and indirect effects of distance from word end, speech rate, segment predictability in context, whether the following vowel was stressed, and word frequency on relative duration. For all measures the direct effect is more robustly different from zero. The indirect path is marked in gray, the direct path in white, and the dark vertical line marks zero.

example, perhaps low information causes reduced duration, and increased intensity indirectly, but reduced segment duration causes fast speech and increased intensity. It is important to note that if (b) holds, then the correlates of lenition would be indirect effects of changes in duration and intensity, rather than causes. All existing causal accounts argue that some external factor causes increased intensity and reduced duration, while for (b), changes in intensity and duration are not caused by any of the correlates. For example, current research does not regard changes in duration as a likely cause for the assignment of stress and word frequency, which would be assumed if (b) holds for any of the correlates.

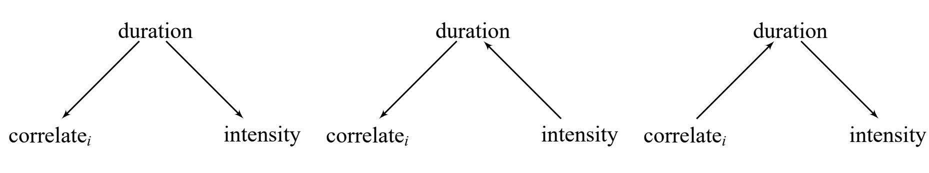

FIGURE 6. The three possible causal structures given the results of study 1, for each of the correlates. 'Correlate$_i$' stands for any of the correlates of lenition, 'duration' for changes in duration, and 'intensity' for changes in intensity. In each alternative, the lines represent (jointly) a possible causal relationship. In (a), changes in duration cause the correlate and changes in intensity. In (b), changes in intensity cause changes in duration, and changes in duration cause changes in the correlate. In (c), the correlate causes changes in duration, and changes in duration cause changes in intensity. For different correlates, correlate$_k$ and correlate$_j$, (a) and (c) are compatible, as the causal direction between duration and intensity does not change. However, (a) and (c) are not compatible with (b) because the causal direction between direction and intensity is reversed.

If we accept that (a) or (c) hold for all of the correlates of lenition, it would suggest that short duration is in some sense more fundamental to variable lenition than increased intensity is. Our studies provide support for existing accounts that claim that fast speech, low information, and prosodic nonboundary environments promote lenition, but call for a reinterpretation of such arguments: there is evidence that all of these factors correlate fundamentally with shorter duration, while they are indirectly correlated with increased intensity, if at all.

There are four potential objections to our conclusions, but we believe that these objections also have implications, which we spell out below. First, we rely on existing research to argue that the causal path should flow from duration to intensity and not the other way around (we offer no evidence for the causal directions between duration and the correlates of lenition). One could accept the mediation analysis but reject our argument based on existing research that causal chain (b) is not correct. However, that would mean that they reject the causal direction for all of the correlates of lenition, and they would have to explain why changes in segment duration cause fast speech, low information, and prosodic smoothing, as we explain above, which seems inconceivable to us.

Second, one could argue that the absence of an effect for a direct causal relationship between the correlates and increased intensity is due to a lack of statistical power. This is an obvious possibility, but it nevertheless raises the question of how strong that correlation is, given that duration is still very highly correlated with several correlates of lenition while controlling for intensity (Table 4), while the converse is not true for intensity and any of the correlates of lenition (Table 3). The two methods we used on top of the Baron and Kenny (1986) procedure are somewhat less susceptible to the statistical power argument: we compared the explanatory power of the effects directly, ignoring statistical significance, and contrasted the predictive power of the duration-mediated model against a full mediation model, finding no difference between the two.

Third, we assume minimality and stability (in the Pearl 2000 sense—that is, we assume the true model is simple), which some may object to. We do not consider highly complex (and uncommon) structures, such as ones in which A and B are not correlated because A causes B to increase via one type of influence, and to decrease through another kind of influence, and the two effects cancel each other out. Such assumptions are made following the scientific process, or Occam's razor. Future research may argue for more complex relationships between the correlates of lenition and duration and intensity, which may necessitate a revision of our findings. However, without evidence or prior research pointing to the existence of such complex structures, we feel comfortable in assuming that the structure is indeed simple.

Finally, the causal argument we offer presupposes that there are no missing factors that cause, for example, both increased intensity and reduced duration (without a directed causal link between the two). However, we do not reject the possibility of there being additional underlying factors that we did not include in our analysis. For instance, it is possible that degree of undershoot would predict both reduced duration and increased intensity, such as in an account based on articulatory undershoot. Even in such causal models though, the three known correlates of lenition would have to be more directly linked to reduced duration than to the underlying cause for there to be no correlation between the correlates of lenition and increased intensity when reduced duration is controlled for. In this case, future research should reveal the nature of the missing factors.

The results raise the interesting possibility that variable lenition could be interpreted as duration-reducing processes first, and only secondarily as, for example, effort-reducing (Kirchner 1998) or undershooting (Bauer 2008) processes. To test the explanatory power of this proposal we conducted an additional study. Study 2 tests whether lenition processes transcribed in the Buckeye corpus are duration-reducing.

2.5. Study 2: lenition processes and duration reduction.

Introduction. In Lavoie's (2001) durational findings, and as evident in additional studies of duration (e.g. Umeda 1977, Crystal & House 1988), voiced obstruents are typically shorter than their voiceless counterparts, fricatives are typically shorter than same-place stops, and approximants are shorter than obstruents.[16] This means that spirantization, voicing, and approximantization are duration-reducing, everything else being equal. Degemination and deletion trivially reduce duration. Though Lavoie (2001) did not present data for glottal segments, surface [h] and [ʔ] in the Buckeye corpus (Pitt et al. 2007) have short duration. [h] is the shortest fricative (77 ms) and [ʔ] is the shortest stop (43 ms), providing an explanation for debuccalization. In this study we check whether all lenition processes in American English can be characterized as duration-reducing. Lenition processes should not result in segments that are naturally longer: if *A* lenites to *B*, the duration of *B* should be overall shorter than the duration of *A*, or at the very least not longer. This question is the focus of study 2.

Methods and materials. To find and contrast as many lenition processes as possible we tried to be more inclusive in the choice of phonological environments than we were in study 1. Study 1 was limited to intervocalic environments in order to control for factors that previous studies investigated. Additionally, we were curious about changes that would affect fortition processes as well, and those are more likely in nonintervocalic environments. Rather than look at word-medial intervocalic contexts, in study 2 we restricted ourselves to environments in which at least one of the flanking segments was a vowel (including segments from the following or previous word), resulting in ~270,000 segments. For each segment, we calculated how often it surfaced as some other segment. To avoid processing noisy alignments and to provide enough data points for the analysis, we excluded changes in which an underlying segment surfaced as that outcome less than 2% of the time. That is, if /t/ surfaced as [t] 57% of the time, as [ɾ] 31% of the time, as [ʔ] or [ˀt] 6% of the time, as [d] 5% of the time, and as [tʃ] 1% of the time, we ran the analysis described below against /t/ → [ɾ], /t/ → [ʔ], /t/ → [d] only, using [t] as a baseline, and excluding changes that were too infrequent (/t/ → [tʃ]).

Buckeye annotations use a transcription scheme that does not support all of the distinctions the IPA allows, which makes it impossible to tell whether some sounds were lenited. /ɾ/ and the nasal tap ([ɾ̃]) are transcribed, but velar fricatives and approximants, which Lavoie (2001) lists as likely outputs for /k/-lenition and /g/-lenition, do not have corresponding surface forms in Buckeye. This means that such lenition processes would be impossible to detect using the method employed here. A different issue has to do with /ʔ/, which may stand for a glottal stop, but also for glottalized coronal stops. This means that in this study, which relies on surface annotations, both debuccalization and partial glottalization would be considered lenition (cf. Chong and Garellek (2018), who focused on codas). Study 1 does not use the surface annotations except for the alignment of underlying forms with their likely surface forms.

We used a mixed-effects linear regression over a subset of the data that included only the two segments in question (those that surfaced as the underlying or surface forms) to predict the duration of each segment instance. The lmerTest and lme4 packages

[16] Though the data provided by Umeda (1977) does not control for word frequency or other variables that study 2 controls for, it does suggest that the durations of voiceless sibilants and /f/ are likely longer than those of same-place stops. This is consistent with the absence of such processes in our data set (Table 5 below): if lenition is initially duration-reducing and /p/ → [f] processes are duration-increasing, then they are predicted not to surface in variable lenition (though such processes are common in synchronic and diachronic change, perhaps after an intermediate /p/ → [ɸ] stage).

(Kuznetsova et al. 2017, Bates et al. 2015, respectively) were used to fit the model in R (R Core Team 2018) and provide *p*-values. The main variable of interest was the identity of the segment (binary). The significance of this variable reflects whether the identity of the segment impacts duration significantly. We included the variables in 3 as additional fixed effects, and the variables in 4 as random intercepts. Phrase-initial and phrase-final positions were treated as special segments for the purpose of calculating random intercepts.

(3) Fixed-effects variables
  a. Segment identity (binary; variable of interest)
  b. Log word count in the corpus, calculated as described in study 1 using the Buckeye, Fisher (Cieri et al. 2004, Cieri et al. 2005), and Switchboard (Godfrey & Holliman 1997) corpora
  c. Pointwise speech rate of the word, calculated as described in study 1. Measuring speech rate in the more conventional number of segments per second did not yield different results.
  d. Distance from the beginning of the word, in segments, as described in study 1
  e. Distance from the end of the word, in segments

(4) Random intercepts
  a. Word-type identity
  b. Speaker identity
  c. Phonological environment: each of the previous and following contexts was recorded as one of seven categories: obstruent, approximant, nasal, unstressed vowel, secondary stress vowel, primary stress vowel, phrase boundary. The two values were combined to a single random intercept (forty-nine possible combinations).

We considered three alternative methods. First, we considered the SURFACE METHOD, in which we contrasted all relevant surface forms, regardless of whether their underlying form was the segment in question. For instance, for /t/ → [d], we contrasted the durations of all surface [t] with all surface [d], regardless of the actual underlying form. These are the results we present below. In the second PROCESS METHOD we considered only output forms that were aligned underlyingly with the segment of interest. Thus, for /t/ → [d], we considered only surface [t] and [d] that were underlyingly /t/. Finally, in the FAITHFUL METHOD we contrasted only segments that surfaced unchanged, which means that for /t/ → [d], we contrasted /t/ that surfaced as [t] with /d/ that surfaced as [d]. The third method is not applicable to lenited forms that never correspond to an identical underlying form in American English, such as [ɾ] and [ʔ]. The differences in the results obtained using the three methods are minute for the purpose of the current study, and though we focus on the surface method, any difference in predictions from either of the other two methods will be discussed.

Since this study is composed of many individual regression models for each method, we report the full results only of the variable of interest (the identity of the segment), and provide the direction of the effects of the control variables. The identity of the segment is always a binary variable that takes the value 0 (that is, the baseline) when the segment is the same as the reference segment, and 1 when the segment is the same as the output of the lenition process. For example, for /t/ → [ɾ], the reference segment is [t] and would be modeled as a baseline, and [ɾ] is the process output and would be treated as the nondefault value. We also provide summary information about each of the other variables.

Results and discussion. Table 5 provides the full results, and the controls are included in summary form in Table 6. Though the causal argument we are making here predicts only that lenition processes will be duration-reducing, and makes no predictions with respect to fortitions and other changes, we provide results for all of the processes.

S1	S2	process type	diff. β	diff. *SE*	dur. diff. *p*	*n*. S1	*n*. S2	%	example	
b	v	lenition	−6.2	0.9	< 0.001	7,639	6,261	2.6	[pɹɑvli]	'probably'
d	ɾ	lenition	−17.8	0.5	< 0.001	11,280	8,135	37.1	[sʌmbəɾi]	'somebody'
ð	ɾ	lenition	−16.7	0.5	< 0.001	1,379	8,135	2.5	[fɑɾɝ]	'father'
dʒ	ʒ	lenition	−9.6	1.6	< 0.001	4,169	528	6.0	[ɪnʒɝi]	'injury'
n	ň	lenition	−14.3	0.5	< 0.001	24,599	4,850	21.8	[gʌňə]	'gonna'
p	b	lenition	−22.6	0.9	< 0.001	10,805	7,639	2.4	[pibə]	'people'
t	d	lenition	−15.6	0.6	< 0.001	14,362	11,280	4.6	[sɛvəndi]	'seventy'
t	ɾ	lenition	−29.3	0.7	< 0.001	14,362	8,135	30.9	[lɪɾl̩]	'little'
t	ʔ	lenition	−14.4	0.7	< 0.001	14,362	3,201	5.9	[gɑʔn̩]	'gotten'
θ	ð	lenition	−22.2	1.3	< 0.001	5,181	1,379	9.6	[sʌmðɪŋ]	'something'
ŋ	n	other	−5.6	0.6	< 0.001	8,207	24,599	2.3	[θɪn]	'think'
s	ʃ	other	6.2	0.9	< 0.001	23,382	3,500	2.9	[dʒʌʃ]	'just'
ʒ	z	other	2.0	1.7	0.238	528	6,646	4.2	[juzuli]	'usually'
ð	θ	fortition	22.2	1.3	< 0.001	1,379	5,181	3.0	[wɪθɛn]	'within'
ʃ	tʃ	fortition	−0.4	1.2	0.744	3,500	3,236	2.1	[ətɛntʃɪn]	'attention'
z	s	fortition	15.8	0.7	< 0.001	6,646	23,382	15.5	[just]	'used'
ʒ	ʃ	fortition	25.8	1.6	< 0.001	528	3,500	3.8	[juʃli]	'usually'

Table 5. A comparison of surface segments' duration using a linear mixed-effects regression as described in study 2. 'S1' and 'S2' are the two surface segments whose duration is being compared (e.g. when S1 is *b* and S2 is *v*, the row describes the results of a mixed-effects regression that compares the duration of all surface [b] and [v]). 'Process type' is the standard classification for a process that changes S1 to S2. Process type is 'lenition' for processes that are regarded as lenition processes (e.g. voicing, spirantization, debuccalization, tapping), 'fortition' for processes that are regarded as fortition processes (devoicing, occlusion), and 'other' for processes that are neither lenition nor fortition. 'Difference β' is the difference between the duration of S1 and S2 in the model (the coefficient of the segment identity variable), in milliseconds. The coefficient is positive when S2 is longer than S1, and negative when S2 is shorter than S1. 'Difference *SE*' and 'Duration difference *p*' are the standard error of the modeled difference and the *p*-value, respectively. '*n*. S1' and '*n*. S2' are the numbers of surface S1 and S2 used for the comparison. '%' is the percentage of the times an underlying S1 surfaced as S2. 'Example' is an example of the relevant process, taken from the corpus.

S1	S2	slower speech rate	dist. from [$_{\#}$	dist. from]$_{\#}$
s	ʃ	lengthen ***	shorten ***	shorten **
ʒ	ʃ	lengthen ***	shorten *	
ʃ	tʃ	lengthen ***		shorten *
z	s	lengthen ***	shorten ***	shorten **
ð	θ	lengthen ***	shorten ***	
ʒ	z	lengthen ***	shorten *	
dʒ	ʒ	lengthen ***	shorten ***	
p	b	lengthen ***	shorten ***	shorten ***
ŋ	n	lengthen ***		shorten ***
b	v	lengthen ***	shorten ***	shorten ***
t	d	lengthen ***	shorten ***	lengthen ***
θ	ð	lengthen ***	shorten ***	
t	ʔ	lengthen ***	shorten ***	lengthen ***
d	ɾ	lengthen ***	shorten ***	lengthen ***
ð	ɾ	lengthen ***	shorten *	shorten ***
t	ɾ	lengthen ***	shorten ***	lengthen ***
n	ň	lengthen ***		shorten ***

Table 6. The direction and significance of the controls from each of the regression models used in the study. 'Shorten' means significantly reducing duration, and 'lengthen' means significantly increasing duration. '*' stands for $p < 0.05$, '**' for $p < 0.01$, and '***' for $p < 0.001$.

Ten frequent lenition processes were detected, as well as four frequent fortition processes and three processes that involved a change in place of articulation, rather than lenition or fortition. The range of processes is visualized in Figure 7, which compares the duration differences of all processes. Lenition processes were classified as such if they were one of the processes listed in the first section of this article, but in our data set included only voicing, spirantization (of stops or affricates), tapping, and debuccalization. Fortition processes were classified as such if they involved the inverse phonological process, but in our data set included only devoicing and occlusion (including fricative affrication). 'Other' processes in our data set involved change in place of articulation.

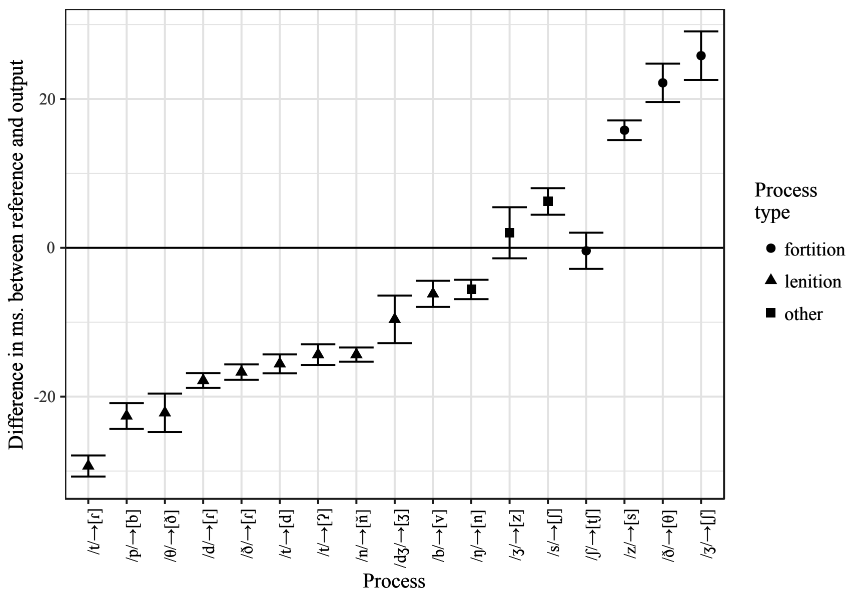

FIGURE 7. The modeled difference in duration between the surface forms of the two segments involved in a process. The reference segment is treated as a baseline. Negative values indicate that the duration of the output form has shorter duration. Error bars are two standard errors from the mean. Shape indicates the process type (lenition, fortition, or unclassified). The comparison uses the surface method, which ignores the underlying forms of the two segments.

All of the lenition processes have a significant effect of duration reduction when switching from the reference segment to the process output, all at $p < 0.001$. Three of the fortition processes involved significant increase in duration, but one process (/ʃ/ → [tʃ]) did not trend significantly in any direction. In the alternative process-based method in which only surface forms that correspond to the reference underlying form were compared, all of the lenition processes remained significantly duration-reducing, but one of the fortition processes (/ʒ/ → [ʃ]) did not involve a significant increase in duration, unlike the surface method. The faithful method could not be used for five processes because their surface form is never underlying. In this method one of the fortition processes, /ʃ/ → [tʃ], involved a significant REDUCTION of 6 ms in duration, which was not detected as such in the other two methods.

All of the lenition processes show a significant reduction in duration, confirming the proposition that lenition involves a reduction in segment duration. Furthermore, though the theory aims to explain lenition processes alone, several cases of fortition feature a statistically significant increase in duration, and only one fortition process, in only one of the three methods, involved statistically significant decrease in duration. Thus, the pattern of fortition processes provides converging evidence supporting the view that lenition is first and foremost duration-reducing.

3. Discussion.

3.1. Causal precedence in lenition. The mediation analysis in study 1 shows that reduction in duration and increase in intensity do not have the same causal status with respect to lenition. The results indicate that the correlation between changes in established correlates of lenition and increased intensity is completely mediated by changes in duration, though there may be small direct links between word frequency and change in intensity. Given the size of the corpus and the strength of the predictors with respect to reduced duration, the absence of a more direct link between the correlates of lenition and changes in intensity does not seem to follow from lack of sufficient data. Though the prototypical lenition processes often involve increased intensity, so much so that phonologized lenition is viewed by many researchers as an increase in sonority (associated with increased intensity; Parker 2002), the causal relationship between increased intensity and fast speech, proximity to stress, and low information content is mediated by reduced duration.

The mediation results do not provide, by themselves, the direction of the causal effect: they do not reveal what is the cause and what is the effect. However, they do crucially require that the causal direction between changes in duration and changes in intensity be consistent for all causal paths: either changes in duration cause changes in intensity, or the other way around. The second alternative, that changes in intensity cause changes in duration, would imply that changes in duration cause all of the correlates: differences in word frequency, speech rate, and proximity to stress, which we find highly implausible given existing research.

As was shown in the past in smaller-scale studies, in study 2, reduced duration was near-categorical in its effect on variable lenition: every lenition process in the Buckeye corpus was duration-reducing, regardless of the method we used to compare the segments' durations. Though phonologized lenition processes may deviate from this generalization, the findings provide evidence with respect to the existing variation in the intermediate stages. Phonologization would apply then to these intermediate stages, possibly leading to duration-increasing outputs.[17] Though fortition processes may not have the exact opposite trend of lenition processes, it is interesting to note that variable fortition processes were often duration-increasing, suggesting that some aspects of fortition may be indeed the flip-side of lenition.

We stipulate that forms with reduced duration, possibly accompanied by increased intensity as a byproduct of duration reduction, may gradually affect the representation of the phonemes used in the context of fast speech, post-stress intervocalic positions, and low-information-content contexts, and reshape them such that phonologization of variable processes would be affected by these properties, resulting in processes whose output forms naturally have shorter duration or increased intensity.

3.2. Implications for existing accounts of lenition. Hypoarticulation accounts posit that lenition occurs due to the speaker's failure to make the correct articulatory

[17] For example, although spirantization frequently affects voiced stops crosslinguistically, Lavoie (2001) found that voiced stops in American English lenite only to approximants, and never to fricatives. It is possible that the first stage of voiced stop spirantization is a change to an approximant, and the approximant can later change further to a fricative. In our data, both lenition to approximants and voiced stop spirantization (except /d/ → [z]) would be duration-reducing.

gesture, which can in turn be caused by a variety of factors. While general hypoarticulation accounts (e.g. Lindblom 1990) focus on missing or hitting the articulatory target and can therefore account for both undershooting and overshooting the target, undershoot accounts focus specifically on gestures that fall short of reaching the intended articulatory target (e.g. Bauer 2008). Articulatory phonology accounts (e.g. Browman & Goldstein 1992) rely on the overlap of the timing of gestures or the reduction of specific gestures, which leads to the production of lenited forms. While there could be multiple reasons to undershoot an articulatory target (e.g. carelessness), reduced duration could be a prominent contender: given less time to reach an articulatory gesture, gestures would be less likely to be complete, and would be more likely to overlap. The same is not true for intensity, however, and it is not clear why increased intensity would cause undershooting an articulatory target or articulatory overlap. Therefore, in many ways our findings provide support for lenition-as-undershoot accounts. One advantage of combining our findings with existing undershoot accounts is that doing so could help curb the overgeneration of undershoot accounts regarding final devoicing: though final devoicing may be considered an undershoot due to the elimination of the voicing target, it likely does not involve reduced duration. Duration reduction also provides a straightforward way to regard debuccalization as lenition: debuccalized forms do not involve undershoot, but they do involve reduced duration. The causal path that explains debuccalization may involve an articulatory target whose duration is so short that the output is misperceived as other targets of shorter duration, leading subsequently to the perception of debuccalization processes.

The idea of effort reduction as a causal mechanism behind weakening is highly intuitive and has been discussed at great length. For some time researchers have assumed or speculated that lenited forms (the output of lenition processes) should be less effortful than their input (Peile 1875, Zipf 1935, Hock 1986, Lindblom 1990, Boersma 1998). In Kirchner 1998, effort can be reduced explicitly by reducing the duration of sustained effort. However, one could imagine that reduced duration could also cause greater acceleration to achieve a given gesture, making reduced duration more effortful. This discussion is quite speculative though, because there is currently no agreed-upon method to measure articulatory effort. Disagreements exist in the literature, for example, between Kirchner (1998), who regards fricatives as less effortful than stops, and Kingston (2008), who regards fricatives as more effortful than stops. There is therefore a way for our findings to be compatible with effort-reduction accounts, but this would depend on the specific model by which effort is determined. In contrast to effort, duration is measured directly.

Kingston (2008) and Katz (2016) propose that the goal of lenition is to reduce disruption of the speech stream (the carrier signal in Harris 2003) when there is no prosodic boundary, creating prosodic smoothing. This is achieved by changing the intensity of the target segment such that lenited forms have increased intensity compared to the input. The opposing fortition processes (strengthening) increase disruption in the presence of a prosodic boundary by reducing the intensity of the target segment. Katz and Fricke (2018) show that listeners expect reduced forms to occur in nonboundary positions in an artificial language experiment. We found no evidence that stress affects intensity directly, but that evidence does not directly oppose prosody-based accounts. Rather, our data would suggest that duration reduction is the property that follows from the need to signal prosodic boundaries, and increased intensity follows from reduced duration.

Hume (2008) argues that high predictability leads to segment instability, and Cohen Priva (2017a) explicitly suggests that low average information content (informativity) is the underlying reason that certain segments lenite in multiple environments and in multiple varieties of a particular language. For example, English /t/ provides relatively low information content compared to /t/ in other languages, which means that it is more predictable in the context in which it occurs. Zipf (1935) makes a similar argument, building on frequency instead. However, the data provided in Cohen Priva 2017a:§5.2 shows that fricatives usually have higher average information content than stops. This makes the claim somewhat paradoxical if taken as an argument about optimization: suppose that the crosslinguistic average of the information /t/ provides is two bits, and /θ/ three bits. If some language's /t/ provides one bit of information, it would be under pressure to lenite. However, spirantization would make things worse: a /θ/ in the same context would still provide one bit, and would be even further from its crosslinguistic average of three bits. This problem could be solved if the direct causal route from low information content to lenition is abandoned. If low information content only causes shorter duration, then the processes that follow are in response to the shorter duration, rather than low information content directly. In other words, spirantization is a solution to a duration constraint, not a pressure to resolve disparities between information in a language and average information. In this way relatively low information content and short duration can coexist as causal factors for lenition processes. Cohen Priva 2012:Ch. 1 asks what makes /t/ prone to lenite in English varieties. We propose that the answer is that unusually short duration makes a segment prone to lenite. That short duration, in turn, is likely explained by low information content.

3.3. Possible limitations. The results presented in this article have a few limitations. First, our results have not tested the full causal chain leading from independent causal forces to a lenition process. Our results showed that fast speech, proximity to stress, and low information content were correlated with duration differences, but did not rule out additional potential causal factors that may still have a direct effect on intensity or duration. In a similar vein, our account is not specific to any given mechanism or explanation, and is in this sense compatible with existing accounts. Effort reduction may be one of the causes for reduced duration, or undershoot may be the mechanism by which reduced duration results in increased intensity.

A different limitation is that the results presented here are for only one language, American English. We do not know whether equivalent results would be found in other languages. There are two studies of the effect of prosodic environment and duration on intensity in other languages: a study of Gurindji stops (Ennever et al. 2017) and one of Campidanese Sardinian (Katz & Pitzanti 2019), both using related but different measures of intensity. Both studies find that the effect of prosodic position is largely or wholly mediated by duration. Ennever et al. (2017) did not use mediation analysis per se, but provided the crucial regression model by including both duration and prosodic position in multiple regression. They showed that in the presence of duration, prosodic environment carried no additional explanatory power. Katz and Pitzanti (2019) did perform a mediation analysis along the lines of Baron & Kenny 1986, and with the exception of word boundary, the effects of prosodic environments were mediated by duration. Study 1 focused on intervocalic environments, and therefore makes no prediction with respect to the independent effect word boundary may have on intensity. However, as Katz and Pitzanti (2019) note, different articulatory mechanisms are likely involved. Together, these findings suggest that the mediation of prosodic effects is unlikely to be American English-specific.

An important American English-specific finding is shown in the duration of interdental fricatives. The data in Table 5 suggests that the duration of [ð] and [θ] in American English is of the same scale or longer than that of [d] and [t], respectively. In the proposed account, this difference would forbid /d/ → [ð] and /t/ → [θ] processes, which are attested in other languages. One possible solution is that American English [d] and [t] may have unusually short duration, or American English [ð] and [θ] may have unusually long duration (differences in durations were found crosslinguistically by Lavoie (2001) and Torreira and Ernestus (2011), though trends were kept). If this is true, the proposed causal account would explain why these two processes are not attested in American English, even though they may be found in other languages.

We do not provide an exact articulatory mechanism that would lead from reduced duration to increased intensity. There is much work that has tested such effects specifically and has found, for example, that shorter duration leads to passive voicing (Westbury & Keating 1986). Other explanations would likely follow the causal route suggested by undershoot accounts: given less time to reach a particular articulatory target, a speaker's performance would exhibit the failed attempt to reach the articulatory target. Another possibility, implied by Katz (2016), is that speakers or grammars may 'choose' a compromise between perceptual similarity and constraints on short duration. For example, [ʔ] could be the easiest voiceless stop to produce given less time than it takes to produce [t], predicting /t/ → [ʔ]. Unlike the account proposed by Katz (2016), our results do not predict the conditions under which lenition would result in, for example, voicing as opposed to spirantization.

Finally, since the account we provide is not complete and relies on existing accounts for the mechanisms that lead from reduced duration to lenited forms, it is difficult to check whether it would overgenerate. Bauer (2008) argues that lenition follows from undershooting the articulatory target, and he is forced to define processes such as final devoicing as lenition-type processes because they too follow from failure to reach a particular articulatory target. For theoretical reasons, Smith (2008) makes a similar argument that draws parallels between lenition and neutralization. Our account may overgenerate by regarding all duration-reducing processes (e.g. /d/ → /ʔ/) as lenition, but the absence of such processes is explained by other accounts that our proposal is compatible with. These include processes that are ruled out on perceptual similarity grounds (Steriade 2008), which could follow from the output being taken as not representative of the input), and processes that require unlikely articulatory changes (e.g. as explained by undershoot accounts).

4. Summary. Lenition processes seem to share very little from the articulatory standpoint, involving different gestures and outputs. However, two properties—reduced duration and increased intensity—were identified as characteristic of such processes almost 150 years ago and have received much attention in current work on lenition. We show here that the two do not have the same causal status. Rather, reduced duration causally precedes increases in intensity.

Study 1 used the Buckeye corpus to show that in variable lenition processes in American English, three independent correlates of lenition—fast speech, proximity to stress, and low information content—are not associated with increased intensity if the actual duration of the segment is controlled for. The converse did not hold: even when intensity is controlled for, the three independent correlates of lenition are still strongly associated with reduced duration. Study 2 showed that all variable lenition processes in the Buckeye corpus are significantly associated with reduced duration, while several fortition processes are associated with an increase in duration.

The results presented here are for consonant lenition in American English, and future work should extend the current findings to additional languages, as our account proposes a crosslinguistic solution to consonant lenition. The focus on reduced duration as causally preceding other aspects of lenition is compatible with existing causal accounts for lenition, but offers several advantages. Duration is readily measurable, the additional assumptions are few, and the requirement of reduced duration helps prevent existing accounts from overgenerating or making counter-functional arguments.

Lenition as a topic crosses traditional boundaries between phonetics, phonology, historical change, and information-theoretic linguistics. There is a lively ongoing discussion on the causes and characteristics of lenition processes. The contribution of this article is a proposal with a fairly simple principle, with few auxiliary assumptions: lenition is initially a reduction in duration. Such an account has the flexibility to work with previous causal proposals but the significance to impact the way we view lenition.

Appendix: The alignment of the buckeye corpus

A1. The structure of the data sources. The Buckeye corpus provides two types of records that are relevant to the alignment procedure. One type is the word record, which lists the following for each word.

(a) Its position relative to the sound file, in seconds. Start and end positions can be straightforwardly calculated from the sequence of records.
(b) The identity of the word
(c) An idealized underlying phonemic representation, which does not contain information about lexical stress
(d) The surface representation
(e) Part-of-speech information (not used here)

The second type is the segment record, which contains only the identity (as determined by the transcription process) and time-alignment information for surface segments (equivalent to (a)–(b) above).

Since the availability of stress is crucial for the data presented in this article, we used a different source for the phonemic representations, the CMU dictionary (Weide 2008). For words in the Buckeye corpus without CMU equivalents, we attempted to turn the Buckeye representations into CMU-like phonemic representations. Crucially, such words were not used in the final studies, but were used only during the alignment process. This was only an approximation that does not attempt to reach true CMU representation, as these data were not going to be used in the analysis. The sole purpose was to allow the alignment of the rest of the data to proceed uninterrupted. This process included:

- Turning all syllabic nasals and liquids to schwa + liquid/nasal equivalent
- Reverting all ‘phonemic’ taps and flaps to the appropriate phoneme
- Changing all ‘phonemic’ glottal stops to /t/
- Assigning primary stress to all vowels
- Turning nonstandard symbols to their standard equivalents (e.g. in a few places Buckeye uses ‘h’ to represent /h/, though the standard ARPAbet representation used elsewhere in the corpus and in the CMU dictionary is ‘hh’)

A2. The alignment model. The alignment used an edit-distance procedure in which the underlying phonemic form was intended to align with the surface representation through a series of ‘edit’ operations. The possible edits were substitution, in which an underlying element corresponded to a surface element, deletion, in which an underlying element did not have a surface equivalent, and insertion, in which the surface representation did not have an underlying representation.

An edit path was defined as a sequence of edit operations that related an underlying representation with a surface one. For instance, the string *a* could correspond to the string *b* through any of the following paths: [substitute(a, b)], [delete(a), insert(b)], and [insert(b), delete(a)].[18]

A path of edit operations represents a simple generative process that generates the underlying and surface strings at the same time: delete(u) appends *u* to the underlying string (but does not modify the surface string), insert(s) appends *s* to the surface string (but does not modify the underlying string), and substitute(u, s) appends *u* to the underlying string and *s* to the surface string. Positive costs were assigned to each operation, and

[18] It should be noted that even though [delete(a), insert(b)] and [insert(b), delete(a)] lead to the same correspondence between underlying and surface forms ({a→Ø, Ø→b}), they are considered different from one another. This is due to the first path describing insertion of an element before *a*, and the latter an insertion of an element following *a*.

the cost of a path of edit operations is the sum of costs of the edit operations that comprise it. Given such costs, an edit-distance algorithm would choose the least costly path for a given underlying string and a given surface string. We trained the model to assign costs that were designed to approximate the negative log probability of observing that operation relative to other operations. Therefore, the cost of a path is the negative log probability of all of the edit operations that generate the underlying and surface strings, and an edit-distance algorithm given such costs would be choosing the most probable path generating the underlying and surface strings at once. For example, an alignment of *ab* to *bc* involving [delete(a), substitute(b, b), insert(c)] would have the sum of the cost of the three operations. The sum represents the probability of observing the underlying string *ab* cooccurring with the surface *bc* string, with this particular path of edit operations.[19]

A3. TRAINING THE ALIGNMENT MODEL. The goal of training the alignment model is to choose the probabilities that would make the pairs of underlying and surface forms most probable: frequent edit operations (e.g. substitute(/n/,[n]), delete(/t/)) should be highly probable, and therefore have relatively low cost, while infrequent or nonexistent edit operations (e.g. substitute(/m/,[k]), insert([ʃ])) should have very low probability, and therefore very high cost.

At the initial stage all operations were equally likely. The alignment training iterated over underlying to surface correspondences and chose the most likely alignment of each underlying phonemic representation with its corresponding surface form, counting all of the operations of each kind. The counts were subsequently transformed into probabilities with some added noise using NumPy's random.dirichlet() function, and the probabilities were negative log-transformed to act as weights in the following iterations. Ten iterations over the entire corpus sufficed to get accurate alignments.

A4. ALIGNING THE CORPUS. The individual files of the corpus were not always perfectly aligned. That is, the segment file could have aligned a segment start or end point as lying somewhat outside the start and end points of the word it was assigned to. Therefore, the sequences of underlying and surface sounds was split based on word edges that were identical to segment edges and aligned to one another several words at a time. By using this method, segments that began or ended somewhat before or after word edges could be aligned with the underlying segments associated with a word, though misalignments of more than 25 ms were not permitted.

Underlying-segment-to-surface-segment alignment used the counts discovered during the training of the corpus, and 0.1 was added to all possible operations to permit substitutions that were not observed in the word-to-word training. Probabilities were otherwise calculated as the number of times the edit operation was observed, divided by the number of all edit operations. The full alignment and the counts associated with the training model are available at https://github.com/ucpresearch/moredata.

REFERENCES

AYLETT, MATTHEW, and ALICE TURK. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47.31–56. DOI: 10.1177/00238309040470010201.

AYLETT, MATTHEW, and ALICE TURK. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America* 119.3048–58. DOI: 10.1121/1.2188331.

BARON, REUBEN M., and DAVID A. KENNY. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51.1173–82. DOI: 10.1037/0022-3514.51.6.1173.

BARR, DALE J.; ROGER LEVY; CHRISTOPH SCHEEPERS; and HARRY J. TILY. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–78. DOI: 10.1016/j.jml.2012.11.001.

BATES, DOUGLAS; REINHOLD KLIEGL; SHRAVAN VASISHTH; and R. HARALD BAAYEN. 2018. Parsimonious mixed models. arXiv:1506.04967v2 [stat.ME]. Online: https://arxiv.org/abs/1506.04967v2.

[19] Notice that this is not the probability of observing *ab* as an underlying string and *bc* as a surface string with ANY edit path. *ab* could cooccur with *bc* as a product of other edit paths—for example, substitute(a, b) and substitute(b, c)—and the probability of observing *ab* with *bc* would be the sum of all the processes that can generate both strings at once.

Bates, Douglas; Martin Mächler; Ben Bolker; and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48. DOI: 10.18637/jss.v067.i01.

Bauer, Laurie. 2008. Lenition revisited. *Journal of Linguistics* 44.605–24. DOI: 10.1017/S0022226708005331.

Bell, Alan; Jason M. Brenier; Michelle Gregory; Cynthia Girand; and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60.92–111. DOI: 10.1016/j.jml.2008.06.003.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Boersma, Paul, and David Weenink. 2008. Praat: Doing phonetics by computer. Version 6.0.37. Online: http://www.praat.org/.

Bouavichith, Dominique, and Lisa Davidson. 2013. Segmental and prosodic effects on intervocalic voiced stop reduction in connected speech. *Phonetica* 70.182–206. DOI: 10.1159/000355635.

Browman, Catherine P., and Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, ed. by John Kingston and Mary E. Beckman, 341–76. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511627736.019.

Browman, Catherine P., and Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49.155–80. DOI: 10.1159/000261913.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14.261–90. DOI: 10.1017/S0954394502143018.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80.1–28. DOI: 10.18637/jss.v080.i01.

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10.395–411. DOI: 10.32614/RJ-2018-017.

Chong, Adam J., and Marc Garellek. 2018. Online perception of glottalized coda stops in American English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 9:4. DOI: 10.5334/labphon.70.

Cieri, Christopher; David Graff; Owen Kimball; Dave Miller; and Kevin Walker. 2005. Fisher English training part 2, transcripts. Philadelphia: Linguistic Data Consortium.

Cieri, Christopher; David Miller; and Kevin Walker. 2004. The Fisher corpus: A resource for the next generations of speech-to-text. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC '04)*, 69–71. Online: http://www.lrec-conf.org/proceedings/lrec2004/pdf/767.pdf.

Cohen Priva, Uriel. 2012. *Sign and signal: Deriving linguistic generalizations from information utility*. Stanford, CA: Stanford University dissertation. Online: https://searchworks.stanford.edu/view/9698524.

Cohen Priva, Uriel. 2015. Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6.243–78. DOI: 10.1515/lp-2015-0008.

Cohen Priva, Uriel. 2017a. Informativity and the actuation of lenition. *Language* 93.569–97. DOI: 10.1353/lan.2017.0037.

Cohen Priva, Uriel. 2017b. Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition* 160.27–34. DOI: 10.1016/j.cognition.2016.12.002.

Cohen Priva, Uriel, and Emily Gleason. 2018. The role of fast speech in sound change. *Proceedings of the 40th annual meeting of the Cognitive Science Society*, 1512–17. Online: http://mindmodeling.org/cogsci2018/papers/0293/index.html.

Cohen Priva, Uriel, and T. Florian Jaeger. 2018. The interdependence of frequency, predictability, and informativity in the segmental domain. *Linguistics Vanguard* 4(s2): 20170028. DOI: 10.1515/lingvan-2017-0028.

Crystal, Thomas H., and Arthur S. House. 1988. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America* 83.1553–73. DOI: 10.1121/1.395911.

Dalby, Jonathan M. 1984. *Phonetic structure of fast speech in American English*. Bloomington: Indiana University dissertation.

Dellwo, Volker; Bianca Aschenberner; Petra Wagner; Jana Dankovicova; and Ingmar Steiner. 2004. BonnTempo-Corpus and BonnTempo-Tools: A database for the study of speech rhythm and rate. *Proceedings of Interspeech 2004*, 777–80. Online: https://www.isca-speech.org/archive/interspeech_2004/i04_0777.html.

Duchin, Sandra W., and Edward D. Mysak. 1987. Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders* 20. 245–57. DOI: 10.1016/0021-9924(87)90022-0.

Ennever, Thomas; Felicity Meakins; and Erich R. Round. 2017. A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8:20. DOI: 10.5334/labphon.18.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. *Phonetically based phonology*, ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade, 232–76. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486401.008.

Fougeron, Cécile, and Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101.3728–40. DOI: 10.1121/1.418332.

Freeman, Daniel; Bryony Sheaves; Guy M. Goodwin; Ly-Mee Yu; Alecia Nickless; Paul J. Harrison; Richard Emsley; Annemarie I. Luik; Russell G. Foster; Vanashree Wadekar; et al. 2017. The effects of improving sleep on mental health (OASIS): A randomised controlled trial with mediation analysis. *The Lancet Psychiatry* 4.749–58. DOI: 10.1016/S2215-0366(17)30328-0.

Gay, Thomas. 1981. Mechanisms in the control of speech rate. *Phonetica* 38.148–58. DOI: 10.1159/000260020.

Gesenius, Heinrich Friedrich Wilhelm. 1910. *Gesenius' Hebrew grammar*. Oxford: Clarendon.

Godfrey, John J., and Edward Holliman. 1997. Switchboard-1 release 2. Philadelphia: Linguistic Data Consortium.

Gordon, Matthew; Edita Ghushchyan; Brad McDonnell; Daisy Rosenblum; and Patricia A. Shaw. 2012. Sonority and central vowels: A cross-linguistic phonetic study. *The sonority controversy*, ed. by Steve Parker, 219–56. Berlin: De Gruyter Mouton.

Gurevich, Naomi. 2011. Lenition. *The Blackwell companion to phonology, vol. 3: Phonological processes*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume, and Keren Rice, 1559–75. Malden, MA: Blackwell.

Hall, Kathleen Currie; Elizabeth Hume; T. Florian Jaeger; and Andrew Wedel. 2016. The message shapes phonology. PsyarXiv preprint. DOI: 10.31234/osf.io/sbyqk.

Harkins, Dan; David Feinstein; Troy Lindsey; Sarah Martin; and Greg Winter. 2003. Switchboard MS-State manually corrected word alignments. Online: https://www.isip.piconepress.com/projects/switchboard/.

Harris, John. 2003. Grammar-internal and grammar-external assimilation. *Proceedings of 15th International Congress of Phonetic Sciences (ICPhS)*, Barcelona, 281–84. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_0281.html.

Hausman, Daniel M., and James Woodward. 1999. Independence, invariance and the causal Markov condition. *The British Journal for the Philosophy of Science* 50.521–83. DOI: 10.1093/bjps/50.4.521.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hock, Hans Henrich. 1986. *Principles of historical linguistics*. Berlin: Mouton de Gruyter.

Hockett, Charles F. 1955. *A manual of phonology.* Baltimore: Waverly.

Honeybone, Patrick. 2002. *Germanic obstruent lenition: Some mutual implications of theoretical and historical phonology*. Newcastle upon Tyne: University of Newcastle upon Tyne dissertation.

Honeybone, Patrick. 2008. Lenition, weakening and consonantal strength: Tracing concepts through the history of phonology. *Lenition and fortition*, ed. by Joaquim Brandão de Carvalho, Tobias Scheer, and Philippe Ségéral, 9–92. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110211443.1.9.

Hume, Elizabeth. 2008. Markedness and the language user. *Phonological Studies* 11.295–310.

Hyman, Larry M. 1975. *Phonology: Theory and analysis*. Fort Worth: Harcourt Brace Jovanovich.

Jacewicz, Ewa; Robert Allen Fox; and Lai Wei. 2010. Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America* 128.839–50. DOI: 10.1121/1.3459842.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61.23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Jaeger, T. Florian, and Esteban Buz. 2017. Signal reduction and linguistic encoding. *The handbook of psycholinguistics*, ed. by Eva M. Fernández and Helen Smith Cairns, 38–81. Hoboken, NJ: Wiley-Blackwell. DOI: 10.1002/9781118829516.ch3.

Judd, Charles M., and David A. Kenny. 2010. Data analysis in social psychology: Recent and recurring issues. *Handbook of social psychology, vol. 1: The science of social psychology*, 5th edn., ed. by Susan T. Fiske, Daniel T. Gilbert, and Gardner Lindzey, 115–39. New York: Wiley. DOI: 10.1002/9780470561119.socpsy001004.

Katz, Jonah. 2016. Lenition, perception and neutralisation. *Phonology* 33.43–85. DOI: 10.1017/S0952675716000038.

Katz, Jonah, and Melinda Fricke. 2018. Auditory disruption improves word segmentation: A functional basis for lenition phenomena. *Glossa: a journal of general linguistics* 3(1):38. DOI: 10.5334/gjgl.443.

Katz, Jonah, and Gianmarco Pitzanti. 2019. The phonetics and phonology of lenition: A Campidanese Sardinian case study. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10:16. DOI: 10.5334/labphon.184.

Kendall, Tyler S. 2009. *Speech rate, pause, and linguistic variation: An examination through the sociolinguistic archive and analysis project*. Durham, NC: Duke University dissertation. Online: https://hdl.handle.net/10161/1097.

Kingston, John. 2008. Lenition. *Selected proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, 1–31. Online: http://www.lingref.com/cpp/lasp/3/paper1711.pdf.

Kiparsky, Paul. 2006. The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics* 32.217–36. DOI: 10.1515/TL.2006.015.

Kirchner, Robert. 1998. *An effort-based approach to consonant lenition*. Los Angeles: University of California, Los Angeles dissertation.

Kirchner, Robert. 2004. Consonant lenition. *Phonetically based phonology*, ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade, 313–45. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486401.010.

Kuperman, Victor; Mark Pluymaekers; Mirjam Ernestus; and Harald Baayen. 2007. Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America* 121.2261–71. DOI: 10.1121/1.2537393.

Kuznetsova, Alexandra; Per B. Brockhoff; and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82.1–26. DOI: 10.18637/jss.v082.i13.

Lass, Roger. 1984. *Phonology: An introduction to basic concepts*. Cambridge: Cambridge University Press.

Lavoie, Lisa M. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. London: Psychology Press.

Levy, Roger, and T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems (NIPS)* 19.849–56. Online: https://papers.nips.cc/paper/3129-speakers-optimize-information-density-through-syntactic-reduction.

Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35.1773–81. DOI: 10.1121/1.1918816.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modeling*, ed. by William J. Hardcastle and Alain Marchal, 403–39. Dordrecht: Kluwer.

Malisz, Zofia; Erika Brandt; Bernd Möbius; Yoon Mi Oh; and Bistra Andreeva. 2018. Dimensions of segmental variability: Interaction of prosody and surprisal in six languages. *Frontiers in Communication* 3:25. DOI: 10.3389/fcomm.2018.00025.

Matuschek, Hannes; Reinhold Kliegl; Shravan Vasishth; Harald Baayen; and Douglas Bates. 2017. Balancing type I error and power in linear mixed models. *Journal of Memory and Language* 94.305–15. DOI: 10.1016/j.jml.2017.01.001.

Parker, Stephen G. 2002. *Quantifying the sonority hierarchy*. Amherst: University of Massachusetts Amherst dissertation.

Pearl, Judea. 2000. *Causality*. Cambridge, MA: Cambridge University Press.

Pearl, Judea. 2012. The causal mediation formula—A guide to the assessment of pathways and mechanisms. *Prevention Science* 13.426–36. DOI: 10.1007/s11121-011-0270-1.

Peile, John. 1875. *An introduction to Greek and Latin etymology*. London: Macmillan and Company.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 137–57. Amsterdam: John Benjamins.

Pitt, Mark A.; Laura Dilley; Keith Johnson; Scott Kiesling; William Raymond; Elizabeth Hume; and Eric Fosler-Lussier. 2007. Buckeye corpus of conversational speech. 2nd release. Columbus: Department of Psychology, The Ohio State University. Online: http://www.buckeyecorpus.osu.edu/.

Pluymaekers, Mark; Mirjam Ernestus; and R. Harald Baayen. 2005. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62.146–59. DOI: 10.1159/000090095.

Proctor, Michael. 2011. Towards a gestural characterization of liquids: Evidence from Spanish and Russian. *Laboratory Phonology* 2.451–85. DOI: 10.1515/labphon.2011.017.

R Core Team. 2018. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: https://www.r-project.org/.

Raymond, William D.; Robin Dautricourt; and Elizabeth Hume. 2006. Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 18.55–97. DOI: 10.1017/S0954394506060042.

Ryan, Kevin M. 2011. Gradient syllable weight and weight universals in quantitative metrics. *Phonology* 28.413–54. DOI: 10.1017/S0952675711000212.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133.140–55. DOI: 10.1016/j.cognition.2014.06.013.

Shipley, Bill. 2000. *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference*. Cambridge: Cambridge University Press.

Sievers, Eduard. 1876. *Grundzüge der Lautphysiologie zur Einführung in das Studium der lautlehre der indogermanischen Sprachen*. Leipzig: Breitkopf und Härtel.

Smith, Jennifer L. 2008. Markedness, faithfulness, positions, and contexts: Lenition and fortition in optimality theory. *Lenition and fortition*, ed. by Joaquim Brandão de Carvalho, Tobias Scheer, and Philippe Ségéral, 519–60. Berlin: Mouton de Gruyter.

Spirtes, Peter; Clark N. Glymour; and Richard Scheines. 2000. *Causation, prediction, and search.* Cambridge, MA: MIT Press.

Steriade, Donca. 2008. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. *The nature of the word: Studies in honor of Paul Kiparsky*, ed. by Kristin Hanson and Sharon Inkelas, 151–80. Cambridge, MA: MIT Press. DOI: 10.7551/mitpress/9780262083799.003.0007.

Surendran, Dinoj, and Partha Niyogi. 2006. Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. *Competing models of linguistic change: Evolution and beyond*, ed. by Ole Nedergaard Thomsen, 43–58. Amsterdam: John Benjamins.

Szigetvári, Péter. 2008. Two directions for lenition. *Lenition and fortition*, ed. by Joaquim Brandão de Carvalho, Tobias Scheer, and Philippe Ségéral, 561–92. Berlin: Mouton de Gruyter.

Torreira, Francisco, and Mirjam Ernestus. 2011. Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology* 2.331–53. DOI: 10.1515/labphon.2011.012.

Turk, Alice. 1992. The American English flapping rule and the effect of stress on stop consonant durations. *Working Papers of the Cornell Phonetics Laboratory* 7.103–33. Online: https://conf.ling.cornell.edu/plab/paper/wpcpl7-Turk.pdf.

Umeda, Noriko. 1977. Consonant duration in American English. *The Journal of the Acoustical Society of America* 61.846–58. DOI: 10.1121/1.381374.

van Son, R. J. J. H., and Louis C. W. Pols. 2003. How efficient is speech? *Proceedings of the Institute of Phonetic Sciences* 25.171–84.

Vehtari, Aki; Jonah Gabry; Yuling Yao; and Andrew Gelman. 2018. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.2.0. Online: https://CRAN.R-project.org/package=loo.

Vehtari, Aki; Andrew Gelman; and Jonah Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27.1413–32. DOI: 10.1007/s11222-016-9696-4.

Wagner, Robert A., and Michael J. Fischer. 1974. The string-to-string correction problem. *Journal of the ACM* 21.168–73. DOI: 10.1145/321796.321811.

Warner, Natasha, and Benjamin V. Tucker. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America* 130.1606–17. DOI: 10.1121/1.3621306.

Wedel, Andrew; Abby Kaplan; and Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128.179–86. DOI: 10.1016/j.cognition.2013.03.002.

Weide, R. L. 2008. *The CMU pronouncing dictionary*. Release 0.7a. Pittsburgh: Carnegie Mellon University. Online: http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Westbury, John R., and Patricia A. Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22.145–66. DOI: 10.1017/S0022226700010598.

White, Laurence. 2002. *English speech timing: A domain and locus approach*. Edinburgh: The University of Edinburgh dissertation.

Zipf, George Kingsley. 1929. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 15.1–95. DOI: 10.2307/310585.

Zipf, George Kingsley. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Boston: Houghton Mifflin.

[uriel_cohen_priva@brown.edu]
[emily_gleason@alumni.brown.edu]

[Received 17 August 2018;
revision invited 3 March 2019;
revision received 3 July 2019;
revision invited 28 October 2019;
revision received 8 January 2020;
accepted 9 January 2020]