Yevgeniy Melguy
Susan Lin, Brian Smith
M.A. Qualifying Paper
9 December 2018

Strengthening, weakening and variability: The articulatory correlates of hypo- and hyper-articulation in the production of English dental fricatives

1. INTRODUCTION

A number of influential approaches to understanding phonetic and phonological variation in speech have highlighted the importance of functional factors (Blevins, 2004; Donegan & Stampe, 1979; Kiparsky, 1988; Kirchner, 1998; Lindblom, 1990). Under such approaches, speaker- and listener-oriented principles—ease of articulation vs. perceptual clarity—often work in opposite directions with respect to consonantal articulation. Minimization of effort is thought to drive a general "weakening" of consonants (resulting in decreased articulatory constriction and/or duration) which often makes them more articulatorily similar to surrounding sounds. This can result in assimilation, lenition, and ultimately deletion, and generally comes at the expense of clarity. By contrast, maximization of clarity drives consonantal "strengthening" processes (resulting in increased articulatory constriction and/or duration) that makes target segments more distinct from neighboring sounds, which can result in fortition. Clear speech generally involves more extreme or "forceful" articulations, and usually comes at the expense of requiring more articulatory effort from the speaker.

One important difference within such functionalist approaches, however, is whether they aim primary to explain phonetic or phonological patterns in speech, where the former focuses on continuous, *gradient* variation in the contextual realization of a given sound, while the latter seeks to account for categorical alternations between discrete sounds.[1] For instance, in Hypo- and Hyper-articulation (H&H) Theory, which attempts to explain phonetic gradience, speech is claimed to occur "along a continuum with more forcefully articulated 'hyper' forms at one end

[1] Note that the term *categorical* has also been used in the literature to refer to phonological processes that apply 100% of the time in a given context, whereas I use it to refer to a non-gradient change involving the substitution of one phonological target with another. Crucially, this includes probabilistic alternation between two sounds that is contextually-conditioned.

and less energetic 'hypo' forms at the other." (Lindblom, 1996, p. 1687). Hyper-articulation involves more "extreme" productions of the target sound that are maximally distinct from neighboring sounds, while hypo-articulation will result in less extreme productions that overlap to a greater degree with flanking segments. By contrast, approaches such as Natural Phonology explicitly focus on categorical phonological alternations, where fortition is claimed to be driven by the principle of clarity, while lenition is driven by minimization of effort (Donegan & Stampe, 2009, p.2). Crucially, while H & H theory does not discount the possibility of fortition and lenition being

Yet, despite such differences, the patterns of phonetic weakening and strengthening predicted by H&H theory mesh closely with the distributions of lenition and fortition that other functionalist accounts predict (Donegan & Stampe, 1979; Kiparsky, 1988; Kirchner, 1998). This suggests that the phonological processes of *fortition* (the "strengthening" of a phone to a manner of articulation with greater articulatory constriction and/or duration, e.g., fricative > stop) and *lenition* (the "weakening" of a sound to a manner of articulation with less constriction and/or reduced duration, e.g., stop > fricative) can be conceived of as related but opposite processes along a spectrum of consonantal strength. The idea is appealing because it captures the fact that lenition and fortition phenomena across languages tend to be in complementary distribution—contexts that favor the one tend to disfavor the other (Kiparsky, 1988; Kirchner, 1998). For instance, lenition is quite frequent word-medially and in unstressed syllables, while fortition tends to be more common word-initially and in stressed syllables, which are the same contexts where we would expect hypo- and hyper-articulation, respectively.

In fact, this close link between phonetic weakening and strengthening and their phonological cousins, lenition and fortition, has led some scholars to propose that the latter are just special cases of strengthening and weakening (Blevins, 2004) that have been phonologized. It is thus possible that what have been frequently analyzed in the literature as phonological phenomena (i.e., lenition vs. fortition) may actually be better explained as the outcome of gradient phonetic processes—that is, just two possible endpoints along a hypo- to hyper-articulation spectrum. This study investigates this possibility by examining the role of functional factors—economy of effort vs. perceptual clarity— in the production of the (inter-)dental fricatives /θ ð/ in Standard American English. These sounds are known to frequently undergo *th-stopping*, a phenomenon in which they are realized as dental or alveolar stops [t d] (Mesthrie &

Bhatt, 2008). In particular, I am interested in examining the hypothesis that the production of /θ ð/ as stops can be explained as the endpoint of a gradient gestural strengthening process—a type of hyper-articulation—rather than a categorical phonological one.

Analyzing the production of these sounds within an H&H-theoretic framework allows us to make specific predictions about both the contextual distribution of stopping and about gradient phonetic variation in the realization of /θ ð/ across contexts. First, we would expect minimization of speaker effort to inhibit stopping (just as it favors lenition), while maximization of perceptual clarity facilitates it. Second, we would expect to see *gradient* variation in the realization of /θ ð/, meaning that even when these sounds are they are *not* realized as stops, they should show the phonetic correlates of hyper- and hypo-articulation: /θ ð/ in contexts that favor stopping should be produced with increased consonantal strength, and vice versa. We would also expect to see no major change in place of articulation, since the latter would suggest a change in the articulatory target, which would be more consistent with the alternative hypothesis that th-stopping is the result of a phonological process.

Results provide mixed support for the hypothesis that stopping is a result of clear speech—hyper-articulation. On the one hand, the distribution of stopping is consistent with the predictions that H&H theory generates about where we should expect hyper- vs. hypo-articulation. Results show that stopping is inhibited in contexts where more effort would be required to achieve articulatory constriction, and facilitated in contexts where perceptual clarity is especially important. However, there is no clear evidence for the kinds of gradient weakening and strengthening we would expect if stopping were just an endpoint of a gradient phonetic process. Although, as predicted, dental place of articulation is preserved in stopped fricatives, results do not show a consistent correlation between phonetic correlates of strength (constriction/duration) and the likelihood of stopping. This suggests that th-stopping in Standard American English is an instance of *categorical* phonological alternation between stops and continuants, rather than the endpoint of a *gradient* phonetic process.

2. LITERATURE REVIEW

Weakening/strengthening vs. lenition/fortition

The broad theoretical question which is the focus of this study concerns the nature of the relationship between gradient phonetic processes and the corresponding categorical outcomes that are usually attributed to phonology. H&H Theory acknowledges the tight relationship between the two when it states that the distinction between lenition and fortition is compatible with the distinction between ‘‘hypo’’ and ‘‘hyper’’ forms,” but notes that its own explanatory framework is derived from phonetic principles rather than phonological data, and that its aim is to “describe ‘on-line’ phonetic properties of speech,” rather than phonological patterns (Lindblom, 1996, p. 1689).

The question of how exactly such “on-line phonetic properties” relate to phonological patterns is especially salient in the case of fortition, since there are several key ways in which the latter type of change appears inconsistent with the functionalist goal of enhancing listener comprehension. First, there is something unintuitive about the notion of a stop being a maximally clear production of an underlying continuant (e.g., [d] being the ideal form of /ð/). The fact that hyper-articulation aims for an ideal or citation form of a sound suggests that a hyper-articulated fricative may be exaggerated in certain ways (e.g., longer duration, increased amplitude) but will still preserve its manner of articulation. Second, fortition poses a potential problem for listener comprehension in cases where it leads to neutralization of a phonological contrast, which should impede rather than improve comprehension. For instance, in a dialect of English that exhibits th-stopping, the contrast between alveolar stops /t d/ and dental fricatives /θ ð/ may be lost, with the consequence that minimal pairs like *thought* vs. *taught* or *there* vs. *dare* would become homophones. Crucially, strengthening would not neutralize a contrast provided it were not extreme enough to result in a manner change in the target segment.

One possible account of the relationship between phonetic strengthening and fortition is that the latter is simply one possible phonetic outcome of a more general process of gestural strengthening. Rather than aiming for a distinct articulatory target in such cases, speakers may simply be “applying” general phonetic correlates of clear speech. Thus, fortition could—at least in certain cases—be better understood as the phonetic ‘endpoint’ of a continuum of consonantal strength. This explanation appears to be consistent with the predictions of H&H theory, which states that “as the performance level increases from hypo to hyper, both the duration and the amplitude of articulatory gestures tend to increase, whereas their temporal overlap tends to decrease” (Lindblom, 1996, p. 1687). That is, we should expect more ‘extreme’ articulatory gestures in hyper-articulated consonants, which could in certain cases result in complete occlusion of articulatory constriction—resulting in a stop or affricate.

This explanation is consistent with a broad range of experimental studies suggesting that what are commonly analyzed as categorical phonological phenomena (e.g., deletion) may in fact be better explained as outcomes of gradient phonetic processes For example, some scholars have argued that English /t d/-flapping (Vaux, 2000) is a by-product of prosodically-conditioned articulatory variability, which can result in acoustic shortening of the target sound that leads to a categorical percept (De Jong, 1998; Fukaya & Byrd, 2005). On the bases of these results, the authors argue that rather than being the outcome of a categorical rule, /t d/-flapping is a gradient phonetic process. A similar sort of argument has been made for /t d/ deletion: articulatory research has shown that in some cases of apparent deletion, the tongue-tip gesture for the target sounds is preserved, but simply hidden due to articulatory overlap with surrounding sounds (Browman & Goldstein, 1992). Similarly, Raymond, Dautricourt, and Hume (2006) conclude from their analysis of /t d/-deletion in spontaneous speech that syllable-initial deletion of these sounds is just one possible outcome of gradient gestural reduction in this context. Finally, experimental studies of vowel reduction in English (Davidson, 2006) suggest that what has been frequently attributed to a categorical deletion rule in fact looks more like a gradual reduction process, where the vowel is frequently not fully deleted but simply considerably shortened, leading to the impression that it has been dropped completely.

Gestural strengthening and fortition

However, studies investigating the phonetic underpinnings of apparently phonological processes tend to concentrate on processes of gestural weakening rather than strengthening. To test the question of whether fortition can be explained as a consequence of gradient phonetic strengthening (hyper-articulation), the present study focuses on a well-documented instance of apparent fortition in Standard American English, th-stopping. Although this phenomenon is generally treated as a categorical phonological process, there is evidence which suggests that stopping is just one possible outcome of gestural strengthening. First, like other phenomena which have been argued to result from gradient phonetic mechanisms (e.g., /t/d/-deletion, schwa-deletion), th-stopping is a *variable* process, meaning that it does not apply 100% of the time, but rather occurs with varying likelihood across a range of different prosodic and segmental contexts. On this basis, some scholars have argued that it is driven by general phonetic processes (Zhao, 2007). In addition, there is evidence that the phonetic realizations of /θ ð/ in dialects of English exist on a spectrum: they can occur as stops (Zhao, 2010), as affricates (Rose (2006), or even delete entirely in spontaneous speech (Jurafsky, Bell, Fosler-Lussier, Girand, & Raymond, 1998), supporting the idea that the realization of these segments is gradient rather than being a cut-and-dry case of categorical alternation between a fricative and stop. Experimental work also provides evidence that place of articulation is preserved in stopped /θ ð/ in Standard American English. For instance, Zhao (2010) found that the acoustic characteristics of stopped /ð/ were distinct from those of /d/, suggesting that there is no change in place of articulation during stopping. If present, a place change would provide evidence for a phonological explanation, since it would suggest that stopping involves substituting a distinct phonological representation with a distinct articulatory target. Finally, as we shall see, th-stopping shows evidence of the same sorts of contextual conditioning as both categorical and gradient strengthening phenomena. Thus, it is possible that stopping is better understood as just one outcome of gradient phonetic strengthening, rather than involving a distinct phonological target. If this is true, then we would predict to see a continuum of possible phonetic outcomes for /θ ð/, where complete occlusion may be just one possible manifestation of such strengthening.

Hyper- and Hypo-articulation (H&H) Theory

The following overview summarizes some of the literature on the role of functional principles— economy of effort vs. clarity enhancement— in conditioning consonantal weakening and strengthening. Beginning with the general architecture of H&H theory and the predictions it generates about phonetic variation in speech, it then reviews some of the basic mechanisms by which functional principles have been argued to lead to consonantal weakening or strengthening, and discusses to what extent these findings are compatible with what has been observed in the th-stopping literature.

As we have already noted, H & H theory predicts that speakers' articulations of a given sound will span a continuous range between hypo- and hyper-articulation. Hyper-articulation is driven by a speaker's desires to produce speech that is optimally intelligible for listeners, and involves more extreme articulations that are assumed to require increased speaker effort. Hypo-articulation, by contrast, is driven by speakers' desires to minimize their own articulatory effort, and is assumed to result in speech that is potentially unintelligible. H&H Theory thus pits these weakening and strengthening processes against each other in a constant tug-of-war: speakers are always aiming to maximize the clarity of their speech while minimizing effort.

H&H theory predicts that hyper-articulated speech will result in gestures that are more 'extreme' and maximally distinct from surrounding sounds, whereas hypo-articulated speech shows more contextual influence. The result is that coarticulation and reduction are more likely in hypo-speech, while hyper-speech involves articulations that are closer to their target values. (Lindlbom 1996, p. 1687)." Phonetically, consonantal hyper-articulation has been shown to manifest as longer segmental duration and increased linguopalatal contact in nasal and oral stops (Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2004), and as exaggerated VOT contrasts in oral stops (Jun, 1996). Hypo-articulation, by contrast, manifests in shorter durations and weakened or shallower consonantal constrictions, and in some cases leads to wholesale deletion of phones (Byrd & Tan, 1996; Lindblom, Sussman, & Agwuele, 2009; Sussman, Hoemeke, & Ahmed, 1993). Of particular relevance for this study are two well-documented forms of weakening and strengthening that seem consistent with the two extremes of the continuum that H&H theory posits: coarticulatory undershoot and domain-initial strengthening. Both of these phenomena have been argued to be functionally motivated, and both appear be relevant in conditioning the realization of /θ ð/ as stops.

Economy of effort, weakening, and hypo-articulation

The idea that various weakening processes may be grounded in general constraints of articulatory economy has a long history within speech production research. Frequently, such phenomena have been attributed to coarticulatory "undershoot"— failure to reach an articulatory target due to the influence of surrounding sounds (Lindblom, 1963)—which manifests as reduced articulatory contact and/or duration in consonant production. Undershooting a target can save speaker effort, which may explain why it has been observed to be more likely at faster speech rates, where there is increased time pressure to produce a given sound (Byrd & Tan, 1996; Lindblom et al., 2009), where it also tends to be more extreme (Krull, 1989).

Moreover, the likelihood and degree of undershoot has been shown to be sensitive to segmental context: target consonants flanked by more open segments, such as low vowels, tend to see greater gestural reduction than those flanked by more closed segments, such as high vowels. (Romero, 1996; Sussman et al., 1993).

These observations are compatible with H&H theoretic claims about the phonetic correlates of hypo-articulation, which, as we have already noted, predicts economy of effort to result in coarticulation and reduction. A concise explanation of the mechanisms involved here is put forward by Kirchner (1998), who argues that the amount of articulatory effort a speaker must expend to produce a given sound (defined as the degree of tongue displacement required to reach the articulatory target) is the primary conditioning factor in the propensity of a consonant to see weakening: the more effort required to achieve consonantal constriction, the more likely we are to see lenition (Kirchner, 1998, p. 194). This observation accounts for the fact that intervocalic position is a very common lenition site both diachronically (Hyman, 1975) and synchronically (for an overview, see Kirchner, 1998, p. 182). An effort-based account also captures the fact that lenition is often sensitive to speech rate and occurs more often in rapid speech. As speech rate increases, Kirchner notes, overlap between vowel and consonant gestures (coarticulation) also increases, meaning that the articulator has less time to move from open to closed position, resulting in undershoot if the degree of articulatory effort expended is held constant.

An illustrative case study of this can be seen in Florentine Italian, where consonants have been shown to undergo a broad range of possible lenition outcomes (Giannelli & Savoia, 1978). Italian lenition patterns demonstrate that weakening tends to be not only more likely in contexts where greater articulatory effort would be required to achieve target constriction, but also more extreme. In natural speech, for example, voiceless stops generally lenite to approximants intervocalically, but to close fricatives (or not at all) elsewhere (Kirchner, 1998, p. 188). So, all else being equal, Italian stops they see a greater likelihood and degree of weakening in contexts where surrounding segments are more open—i.e., where increased effort would have to be expended to achieve target constriction. This effect is magnified as the amount of time a speaker has to achieve target constriction and the pressure to maintain clear speech both decrease.

Similar kinds of scalar contextual effects, where the openness of segments surrounding the target conditions not only whether lenition occurs, but the degree of weakening that we see, have also been observed experimentally in other languages. For example, electromagnetic articulography (EMA) studies examining the effects of vowel context on the realization of Spanish stops have shown that although the voiced stops /b d g/ see weakening in virtually all positions, the degree of weakening is more extreme when the flanking vowel is low, resulting in a systematically reduced articulatory constriction (Romero, 1996). As Kirchner (1998) points out, this sort of contextual conditioning may explain the different reduction patterns that have been observed in Spanish past-participle allomorphs, where /-ido/ lenites to a fricative [iðo] or approximant [ið̞o], while /-ado/ often sees total loss of the consonant [ao] in casual speech (Resnick, 1975).

These kinds of patterns demonstrate that segmental context affects the degree of effort that a speaker would need to expend in order to achieve target constriction, and that this in turn affects that degree of gestural weakening that we tend to see across contexts. Turning to th-stopping, we see that the contextual distribution of stopping appears consistent with an effort-based account. The highest rates of stopping are usually observed after a plosive (Childs et al., 2010; Newlin-Lukowicz, 2013; Rose, 2006; Zhao, 2007), followed by after a fricative/affricate (Newlin-Lukowicz, 2013; Zhao, 2007), and with the lowest rates of stopping generally found after a preceding vowel or liquid (Bell & Gibson, 2008; Childs et al., 2010; Newlin-Lukowicz, 2013; Rose, 2006; Zhao, 2007). This basic pattern of results suggests that the more articulatorily open the segmental context, the more strongly it disfavors stopping. Thus, post-vocalic /θ ð/ see the lowest rates of stopping, potentially due to the increased expenditure of effort that would be needed to achieve sufficient consonantal constriction in this context.

Clarity-enhancement, strengthening, and hyper-articulation

On the other side of the spectrum, explanations of consonantal strengthening processes frequently rely on the role of perceptual clarity (Donegan & Stampe, 2009; Kiparsky, 1988; Kirchner, 1998). There are two main observations in favor of this idea. The first is that such strengthening results in increased articulatory and perceptual contrast of the target sound relative to neighboring sounds. For instance, in discussing the functional motivations for the stop-fricative alternation in Tamil, Donegan and Stampe (1979) note that “perceptually, stops represent a sharper contrast with adjacent vowels [than fricatives]” (p. 129). The second observation is that phonological strengthening processes like fortition tend to occur primarily in prosodically prominent positions such as phrase- or word-initially and in stressed syllables (cf. Kirchner, 1998, pp. 10-11, for a cross-linguistic survey of prosodically-conditioned fortition patterns). A possible reason for this distribution, some scholars have suggested, is that more prosodically prominent positions carry greater importance for listener comprehension (Kirchner, 1998; Kiparsky, 1988). This is consistent with the basic assumption of H&H theory that speakers hyper-articulate "primarily to communicate what they mean and to promote correct lexical access" (Lindblom, 1996, p. 1689).

Very similar effects of prosodic position on consonant strength have been shown to take place at the phonetic level. Sounds with lexical or phrasal stress, for instance, have been shown to generally be hyper-articulated—produced with increased amplitude and duration compared to unstressed sounds (see Shattuck-Hufnagel and Turk (1996) for an overview). Similarly, consonants at the beginning of a prosodic boundary (e.g., syllable-, word-, or phrase-initial) have been shown to exhibit increased linguopalatal contact and longer segmental durations (Keating, 2006) as well as exaggerated VOT contrasts (Jun, 1996). This type of hyper-articulation, known as domain-initial strengthening, has been shown to occur at multiple prosodic levels, and appears to be cumulative. Fougeron & Keating (1997), for instance, measured the duration and constriction of /n/ in English across four different prosodic domains—the phonological word, the phonological/intermediate phrase, the intonational phrase, and the utterance. They found that consonants were longer and had more linguopalatal contact in domain-initial position than domain-medially or -finally, and this effect increased proportionally with height on the prosodic hierarchy. Subsequent work has replicated this basic pattern of results for coronal stops across a range of different languages, including French, Korean, and Taiwanese (Cho & Keating, 2001; Keating, Cho, Fougeron, & Hsu, 2004).

Prominent prosodic contexts have also been shown to favor th-stopping. Dental fricatives found in post-pausal position, corresponding to the onset of a large phrase boundary, for instance, are frequent targets of stopping across many distinct dialects of English (Bell & Gibson, 2008; Childs et al., 2010; Newlin-Lukowicz, 2013; Rose, 2006; Zhao, 2007). Moreover, stopping has been shown to be more likely utterance-initially than utterance-medially (Newlin-Lukowicz, 2013; Rose, 2006) while analyses of stopping in prosodically-labeled speech corpora database have shown even more fine-grained effects. Zhao (2007), for instance, found that th-stopping in the American English Map Task Database was most likely at a large phrase boundaries, followed by a small phrase boundary and normal word-boundary, respectively. Finally, there is also evidence that other prosodic factors such as the presence of lexical stress may favor th-stopping (Bell & Gibson, 2008). The general pattern that we observe with prosodic conditioning of th-stopping thus looks consistent with what we get in the case of gradient cases of gestural strengthening, suggesting that th-stopping may be due hyper-articulation.

3. THE PRESENT STUDY

However, although several scholars have hypothesized articulatory strengthening as a possible mechanism for stopping (Newlin-Lukowicz, 2013; Zhao, 2007), such claims have not been experimentally explored. The goal of the present study is to investigate the hypothesis that certain cases of fortition may be the result of gradient phonetic mechanisms, rather than a categorical phonological alternation between "weak" and "strong" sounds. To test this question, I examine the role of functional factors associated with weakening and strengthening in conditioning the phonetic realization of inter-dental fricatives in Standard American English. I do so within the framework of H&H theory, which predicts a continuum of possible phonetic realizations for a given sound, where minimization of effort leads to gestural weakening, while maximization of clarity should result in gestural strengthening. If the production of /θ ð/ as stops is the result of a gradient process of gestural strengthening (hyper-articulation), then we can make several concrete predictions about the distribution of stopping across contexts as well as the articulatory and acoustic characteristics of /θ ð/ even when they are *not* stopped.

Hypotheses

If th-stopping is a result of hyper-articulation, then contexts where perceptual clarity is of greater importance should favor stopping, while contexts where articulatory ease dominates should disfavor it. This means that prosodically prominent positions—domain-initially and in stressed syllables—should see higher likelihood of stopping, with more stopping observed higher on the prosodic hierarchy. H&H theory also predicts that /θ ð/, like any sound, should see weakening in contexts favoring the hypo- end of the continuum. Thus, stopping should be less likely in articulatorily open contexts, due to the increased co-articulatory distance required to achieve target constriction. Post-vocalic environments should disfavor stopping compared to other segmental contexts, and more open (low) vowels should disfavor it more strongly than more closed (high) vowels. In addition to such contextual conditioning, we would also expect lexical factors that favor hypo-articulation, such as high word frequency (Baker & Bradlow, 2009) to decrease the likelihood of stopping. Again, this is because we predict such factors to shift articulation toward the hypo- end of the continuum, where gestural weakening predominates, making articulatory occlusion less likely.

At the same time, if stopping is a categorical outcome that results from gradient phonetic mechanisms, then we should expect to see a correlation between constriction degree/duration and likelihood of stopping. Namely, those environments that show evidence of such phonetic strengthening should also have higher rates of stopping, *even when* strengthening does not result in complete occlusion and lead to a stop percept. And, conversely, those environments which disfavor stopping should systematically see decreased constriction/duration. Finally, we should expect to see no difference in the place of articulation between stopped and fricated /θ ð/. Acoustic studies of Th-stopping in Standard American English are consistent with this hypothesis, suggesting that stopped /θ ð/ are not produced at a different place of articulation (Zhao, 2010), but this has not yet been corroborated via articulatory research. If stopping involves a systematic change in the place of articulation of the constriction, this suggests the presence of distinct targets for stop vs. fricative /θ ð/, more consistent with the view that stopping is a categorical phonological alternation than with the view that it is just one end-point of an articulatory weakening-strengthening continuum.

Summary of hypotheses

In summary, if th-stopping is a result of hyper-articulation, we would expect to see the following patterns in the production of inter-dental fricatives in Standard American English speech:

First, stopping ought to be conditioned by both minimization of effort and maximization of contrast, predicting the following contextual distribution:

- Higher likelihood of stopping in prosodically prominent contexts, including domain-initially and in stressed syllables, since these ought to shift articulations closer to the hyper-articulation end of the continuum.

- Lower likelihood of stopping in contexts which require more articulatory effort to achieve target constriction (i.e., less stopping in low-vowel vs. high-vowel contexts), as well as in in words with higher lexical frequency, since these ought to shift articulations closer to the hypo-articulation end of the continuum.

Second, we should see gradience in the phonetic realization of /θ ð/, and no major changes in place of articulation. In particular, we would predict:

- Gestural strengthening (longer segmental durations and increased tongue-tip constriction) in environments that favor stopping, even where tokens are *not* stopped.
- Gestural weakening (shorter segmental durations and decreased tongue-tip constriction) in environments that disfavor it, even where tokens are *not* stopped.
- Preservation of dental place of articulation

Methods

Data

To test these questions, I rely on the X-ray Microbeam Speech Production Database (XRMBDB), a large, open-source database of Standard American English speech consisting of articulatory data synchronized with the acoustic signal (Westbury, Milenkovic, Weismer, & Kent, 1990). Due to the high spatial resolution of the data, the XRMBDB presents a promising resource for investigating articulatory variability in the production of /θ ð/. This can allow us to detect both differences in place of articulation as well as fine-grained differences in constriction degree. At the same time, the accompanying acoustic data can be used to determine when and where strengthening of /θ ð/ results in occlusion, i.e., clearly perceptible stops.

The XRMBDB utilizes a point-tracking system to monitor the position of small (~3mm) gold pellets attached to the tongue, lips, and other areas in the vocal tract, and the resulting data are synchronized with the acoustic signal. The participants, 48 speakers of Wisconsin English, performed oral motor tasks (e.g., swallowing, repetition of dummy syllables) and read words in isolation, sentences, and short passages. Due to technical issues in extracting point-tracking data for 7 subjects, the present study excludes those participants, making for a total of 41 speakers. Many sentences and passages found in the XRBMDB were selected directly from previous speech production corpora such as the TIMIT database (Garofolo et al., 1993).

Procedure

Acoustic recordings from the XRMBDB were automatically transcribed using the Penn Forced Aligner (Yuan & Liberman, 2009) and subsequently hand-corrected by research assistants trained in phonetic transcription. A burst detector (Johnson, 2004) was then run on the acoustics within each token of /θ ð/ in the data, as delineated by the hand-corrected segmental boundaries. This script detects burst-like transients in the speech signal by comparing successive 5ms windows in the spectrum and waveform, searching for the areas of maximum difference. In the waveform, these are the largest valleys (corresponding to pressure peaks), and in the spectrum, they are areas of maximal spectral change. Based on both waveform and spectral measures, the script then assigns a burst-strength score using a linear discriminate function trained on stop burst in the TIMIT corpus. This is a measure of how 'burst-like' the burst is. If no burst is detected, the script assigns a score of zero. For /θ/ and /ð/, this yielded respective burst-strength values ranging from 0 to 6.09 (M = 0.95, SD = 1.25) and 0 to 7.02 (M = 0.90, SD = 1.25). For reference, respective burst-strength scores for /t/ and /d/ ranged from 0 to 7.49 (M = 1.62, SD = 1.73) and 0 to 7.50 (M = 1.03, SD = 1.82).

A threshold for burst strength was then selected and tokens of /θ ð/ were automatically classified as stopped or not, based on whether they met this cutoff. This was done in order to be sure that tokens classified as stopped consistently corresponded to clearly perceptible stops in the data, following previous studies of Th-stopped which used the presence of a burst as a marker of stopping (Newlin-Lukowicz, 2013; Rose, 2006; Zhao, 2007). The procedure for determining an appropriate burst-score threshhold was as follows: a single subject was selected and all tokens of /θ ð/ in their data for which a burst score was returned were divided into 0.5 increment windows. Tokens in each window were manually hand-checked, looking for the presence of 1) a visible burst in the waveform, defined as sharp spike in amplitude 2) a visible burst in the spectrum, defined as a clearly delineated, dark, vertical band, and 3) an audible burst in the sound file. Each token was then classified as stopped if met all three criteria, or not stopped, otherwise.[2] Beginning with the lowest burst scores and proceeding window-by-window, the threshold was

[2] In future work, this procedure could be made more reliable by having multiple individuals trained in phonetic transcription perform a similar task, and only classify as stopped those tokens for which there is inter-transcriber consensus.

then set at the lowest value of burst strength above which 70 percent or more of tokens were classified as being stopped. This corresponded to the 1.0 – 1.5 burst-score window, within which 14/19 tokens (74%) met the criteria for stop categorization, and the threshold was thus set at a burst score of 1.0.[3] For comparison, only about half of the tokens in the immediately preceding (0.5 – 1.0) window were classified as stopped (15/27 or 56%), while the immediately following (1.5 – 2.0) window was near-ceiling (27/31 or 87% stopped). At the selected threshold, 36% of /ð/ and 37% of /θ/ tokens were classified as stopped (compare to 50% of /d/ and 51% of /t/ tokens). Each token of /θ ð/ was also coded for constriction degree. This was done by first calculating the tongue-tip aperture (the distance between the tongue-tip pellet (T1) and the palate[4], and then extracting the minimum value of this measure for each segment.

Likelihood of stopping, constriction degree, and duration were analyzed with mixed-effects regression modeling (Baayen, 2008; Levshina, 2015; Jaeger, 2008) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2014) in the statistical software R (R Core Team, 2017). Logistic regression was used to predict the likelihood of stopping, while the continuous measures (constriction degree and duration) were analyzed using linear models, fitted to the subset of data that were not stopped, since stopped tokens will necessarily be more constricted then fricatives.[5] All three models included random intercepts for speaker and word. Fixed effects included word position (word-initial, -medial, and -final), preceding context (low, mid, or high vowel; liquid, fricative/affricate (=ref) liquid, nasal, stop, or pause/silence), following context (low, mid, or high vowel), lexical stress (stressed or unstressed, based on phonetic transcriptions in the textgrids), lexical class (function or content word) and lexical frequency (using log-transformed frequency measures obtained from the SUBTLEX corpus: Lg10WF). Segmental context was coded based on the phonemic status of the preceding or following segment, rather than its phonetic realization. Although the lexical factors (word and word frequency) are not of primary interest for the present study, they were included as predictors in the regression analyses since they are theoretically motivated by previous studies of Th-stopping (Childs et al., 2010;

[3] Thus, a total of 136 tokens were analyzed, corresponding to four windows where tokens fell into the 0 – 2.0 burst-score range. Since stopping in the 1.5 – 2.0 window was near-ceiling, no further analysis of tokens above 2.0 burst strength was conducted.

[4] Tongue-tip aperature was calculated with a Python script written by Susan Lin which uses the following equation: $\sqrt{((T1x\text{-}Palx)^2 + (T1y\text{-}Paly)^{2)}}$. which takes the x and y coordinates from the T1 pellet and the palate

[5] For this analysis, a more stringent criterion was used: only the data for which a burst strength of 0 was returned were considered. This was done in order to minimize the influence of stopped segments on the results, since these will necessarily involve greater constriction than fricatives.

Dubois & Horvath, 1998; Newlin-Lukowicz, 2013; Van Herk et al., 2009). They were thus left in the model as controls, even if they failed to significantly improve model fit.

Due to differences in stopping distribution between /θ/ and /ð/ in the XRMBDB (e.g., preceding nasals favored stopping in /θ/ but disfavored it in /ð/), separate (but otherwise identical) models were initially fitted for both /θ/ and /ð/). However, presumably due to the comparatively small number of observations of /θ/ in the dataset (1229, compared to 7486 for /ð/), attempting to fit the full model to the /θ/ data resulted in convergence issues, and a pruned model that successfully converged did not approach significance for any of the included predictors. Therefore, statistical results are only presented for /ð/ in this paper, although minor place of articulation differences between the two fricatives and /t d/ are still discussed.

Results

Results of a likelihood ratio test showed that both prosodic and segmental context were significant predictors of stopping (Table 1). This test works by removing factors from the model one by one and comparing the reduced model to the full model to see whether that factor contributes to a significantly improved goodness of fit. LRT results showed that both preceding and following segmental context were significant ($X^2(7) = 559.25$, $p < 0.001$ and $X^2(2) = 9.85$, $p < 0.01$, respectively), as was word position ($X^2(2) = 6.39$, $p < 0.05$). Of the three lexical factors included in the model, word frequency emerged as significant ($X^2(1) = 35.34$, $p < 0.001$), but not lexical stress ($X^2(1) = 0.27$, $p = 0.603$) or lexical class ($X^2(1) = 1.63$, $p = 0.202$). Predictors that favored stopping included a preceding silence/pause (marking a large phrase boundary) and a preceding stop, while factors that disfavored it included high lexical frequency, a preceding vowel, liquid, or nasal disfavored it (Figure 1).[6] Moreover, preceding low vowels were found to disfavor stopping more strongly than mid-vowels, while high vowels had no significant effect (see Table 1). For following segments, low vowels were found to disfavor stopping relative to high vowels, while mid vowels neither significantly favored nor disfavored stopping. Finally, /ð/ was significantly more likely to be stopped in word-final position than word-medially. Word-initial position approached significance in favoring stopping relative to word-medial position.

[6] Note that the reference level for this predictor is a preceding fricative/affricate.

Table 1. Coefficients for the mixed-effects logistic regression model of stopping likelihood of /ð/: estimates, standard error, z-values, and corresponding and p-values. Note that a positive coefficient indicates increased odds of stopping, while a negative one indicates decreased odds. Significance codes: 0 ‘***’, 0.001 ‘**’, 0.01 ‘*’, 0.05 ‘.’, 0.1 ‘ ’. See Appendix A for full model output.

	Estimate	Std. Error	z-value	Pr(>|z|)	
(intercept)	0.594922	0.390857	1.522	0.127985	
word stress = stressed	-0.047646	0.091614	-0.52	0.603016	
preceding context = high vowel	0.070663	0.168929	0.418	0.675728	
preceding context = liquid	-0.463699	0.131553	-3.525	0.000424	***
preceding context = low vowel	-0.731714	0.297157	-2.462	0.013802	*
preceding context = mid vowel	-0.337065	0.177772	-1.896	0.057953	.
preceding context = nasal	-0.97437	0.146301	-6.66	2.74E-11	***
preceding context = pause/silence	1.018932	0.095345	10.687	< 2e-16	***
preceding context = stop	1.41857	0.111802	12.688	< 2e-16	***
following context = low vowel	-0.335227	0.114634	-2.924	0.003452	**
following context = mid vowel	-0.005822	0.068581	-0.085	0.932347	
log word frequency	-0.477944	0.080422	-5.943	2.80E-09	***
part of speech = function word	0.491116	0.385412	1.274	0.202571	
word position = final	1.29729	0.519336	2.498	0.01249	*
word position = initial	0.738846	0.384881	1.92	0.054899	.

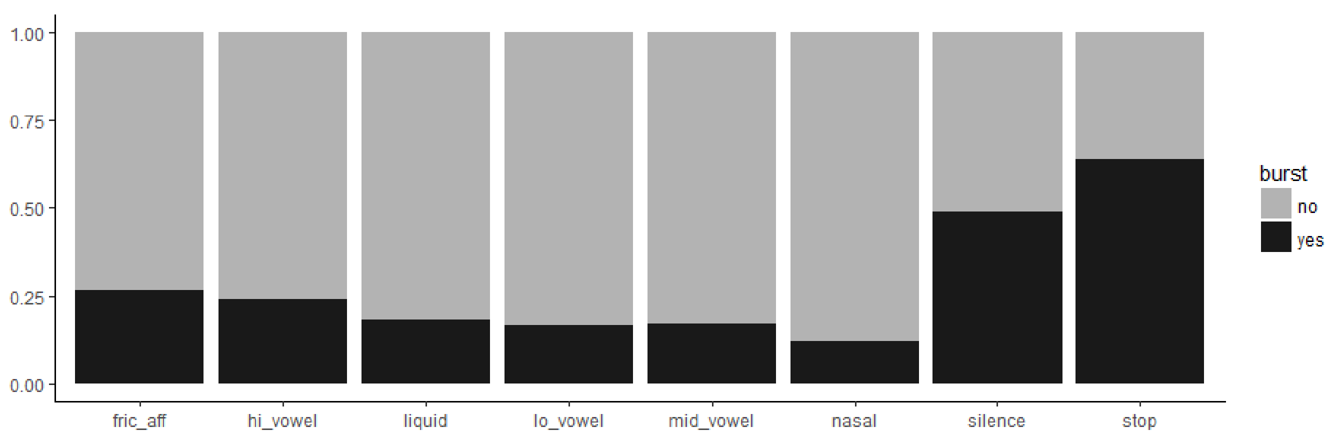

Figure 1. Proportion /ð/ classified as having a burst, by preceding context.

Results of the mixed-effects linear model for constriction degree (tongue-tip aperture) also showed significant effects of prosodic and segmental context, as well as segmental properties of /ð/ (see Table 2). Significant predictors included preceding context ($X^2(7) = 54.90$, $p < 0.001$), following context ($X^2(2) = 29.50$, $p < 0.001$), and duration ($X^2(1) = 22.91$, $p < 0.001$). Word position approached significance $X^2(2) = 4.34$, $p = 0.114$, while non-significant predictors included word stress ($X^2(1) = 0.12$, $p = 0.723$), lexical class ($X^2(1) = 0.002$, $p = 0.967$), and frequency ($X^2(1) = 1.16$, $p = 0.282$).

Table 2. Coefficients for the mixed-effects linear regression model of tongue-tip aperture (constriction degree) of /ð/: estimates, standard error, t-values, and corresponding p-values (t-tests use Satterthwaite's method). Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '. See Appendix B for full model output.

	Estimate	Std. Error	t-value	Pr(>|t|)	
(intercept)	3.17E+00	4.28E-01	7.402	3.70E-05	***
duration	-3.25E-03	6.78E-04	-4.787	1.80E-06	***
word stress = stressed	2.43E-02	6.85E-02	0.354	0.72316	
preceding context = high vowel	3.12E-02	1.19E-01	0.263	0.79238	
preceding context = liquid	-2.00E-01	8.98E-02	-2.225	0.02621	*
preceding context = low vowel	-3.79E-01	1.82E-01	-2.088	0.03992	*
preceding context = mid vowel	-2.77E-01	1.14E-01	-2.422	0.01551	*
preceding context = nasal	-1.16E-01	8.08E-02	-1.435	0.15134	
preceding context = pause/silence	-4.86E-02	7.92E-02	-0.613	0.54005	
preceding context = stop	-6.69E-01	1.07E-01	-6.243	9.62E-10	***
following context = low vowel	4.94E-01	1.53E-01	3.223	0.00834	**
following context = mid vowel	3.39E-01	6.51E-02	5.212	3.20E-07	***
log word frequency	1.03E-01	9.56E-02	1.076	0.3268	
part of speech = function word	1.30E-02	3.10E-01	0.042	0.96701	
word position = final	-6.32E-01	4.42E-01	-1.43	0.16888	
word position = initial	-6.91E-01	3.32E-01	-2.084	0.04863	*

Segmental duration significantly varied inversely with tongue-tip aperture, meaning that longer segments were associated with closer tongue-tip constrictions. For prosodic context, word-initial /ð/ showed decreased aperture compared to word-medial /ð/ (see Figure 3), but no significant effect of phrasal position was found. For segmental context, a preceding stop predicted the largest decrease in tongue-tip aperture (or increase in constriction size), followed by a low vowel, mid vowel, and liquid. The reverse pattern was observed for following vowels: in both cases, low and mid vowels predicted increased tongue-tip aperture relative to high vowels, with a greater increase in aperture observed for low vowels (see Figure 2).

Segmental duration was also significantly affected by prosodic and segmental context (Table 3). Significant predictors included preceding context ($X^2(7) = 348.040$, $p < 0.001$), following context ($X^2(2) = 27.525$, $p < 0.001$), lexical stress ($X^2(1) = 5.716$, $p < 0.05$), and tongue-tip aperture ($X^2(1) = 17.551$, $p < 0.001$). Factors that failed to reach significance included lexical frequency ($X^2(1) = 0.0076$, $p = 0.930$), lexical class ($X^2(1) = 1.731$, $p = 0.188$), and word position ($X^2(2) = 0.487$, $p < 0.784$).

Table 3. Coefficients for the mixed-effects linear regression model of /ð/ duration: estimates, standard error, t-values, and corresponding p-values (t-tests use Satterthwaite's method). Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '. See Appendix C for full model output.

	Estimate	Std. Error	t value	Pr(>|t|)	
(intercept)	60.7914	26.3153	2.31	0.04066	*
tongue-tip aperture	-2.0447	0.4881	-4.189	2.99E-05	***
word stress = stressed	4.6998	1.9658	2.391	0.01688	*
preceding context = high vowel	7.6873	3.3168	2.318	0.02054	*
preceding context = liquid	12.8591	2.5247	5.093	3.76E-07	***
preceding context = low vowel	13.5863	5.766	2.356	0.01861	*
preceding context = mid vowel	9.3205	3.2062	2.907	0.00368	**
preceding context = nasal	-9.5521	2.2718	-4.205	2.70E-05	***
preceding context = pause/silence	25.8353	2.1765	11.87	< 2e-16	***
preceding context = stop	-8.7611	3.132	-2.797	0.00519	**
following context = low vowel	7.3553	6.0537	1.215	0.22538	
following context = mid vowel	-8.3732	1.8985	-4.41	1.07E-05	***
log word frequency	-0.5878	6.7389	-0.087	0.93219	
part of speech = function word	20.2334	15.3771	1.316	0.20362	
word position = final	-16.092	24.2333	-0.664	0.51759	
word position = initial	-10.1334	16.8733	-0.601	0.55558	

Unsurprisingly, prosodic factors had a significant effect on duration, with higher duration in stressed words and at a phrase boundary. Indeed, a preceding pause/silence predicted the largest increase in duration. Duration also varied considerably as a function of segmental context (see

Figure 4). For preceding context, factors that predicted increased duration included high vowels, mid vowels, liquids, and low vowels (in increasing order of magnitude). Stops and nasals both predicted decreases in duration. For following context, mid vowels predicted decreased duration (low vowels were not significant). For ease of comparison, Table 4 summarizes the magnitude, direction, and significance of results for each predictor in the stopping likelihood model, TT aperture model, and duration model.

Table 4. Comparison of model estimates for /ð/ stopping likelihood, minimum tongue-tip aperture, and duration. Significance codes: 0 ‘***’, 0.001 ‘**’, 0.01 ‘*’, 0.05 ‘.’, 0.1 ‘ ’.

	Estimate		TT ap		Duration	
(intercept)	0.594922		3.1700	***	60.791	*
duration	--		-0.0033	***	--	
TT aperture	--		--		-2.045	***
word stress = stressed	-0.047646		0.0243		4.700	*
preceding context = high vowel	0.070663		0.0312		7.687	*
preceding context = liquid	-0.463699	***	-0.2000	*	12.859	***
preceding context = low vowel	-0.731714	*	-0.3790	*	13.586	*
preceding context = mid vowel	-0.337065	.	-0.2770	*	9.321	**
preceding context = nasal	-0.97437	***	-0.1160		-9.552	***
preceding context = pause/silence	1.018932	***	-0.0486		25.835	***
preceding context = stop	1.41857	***	-0.6690	***	-8.761	**
following context = low vowel	-0.335227	**	0.4940	**	7.355	
following context = mid vowel	-0.005822		0.3390	***	-8.373	***
log word frequency	-0.477944	***	0.1030		-0.588	
part of speech = function word	0.491116		0.0130		20.233	
word position = final	1.29729	*	-0.6320		-16.092	
word position = initial	0.738846	.	-0.6910	*	-10.133	

Finally, place of articulation was determined for stopped vs. fricative /θ ð/, and compared to /t d/. This was done by taking the maximum x-value of the tongue-tip pellet, which indicates distance from the reference pellet on the mandibular incisors. Using this metric /θ ð/ are shown to be more anterior than /t d/ by an average of 6.1 and 6.5 mm, respectively. This is consistent with expected differences between alveolar and (inter-)dental place of articulation, as is canonically reported for these sounds (Hillenbrand, 2003). By contrast, no differences of comparable magnitude were found between the constriction location in stopped vs. fricative /θ ð/ (stopped /θ/ was 0.41 mm backer than fricative /θ/, while stopped /ð/ was 0.43 mm more front than fricative /ð/). Figure 5 illustrates the range in place of articulation exhibited by /d/, stopped /ð/, and fricative /ð/.

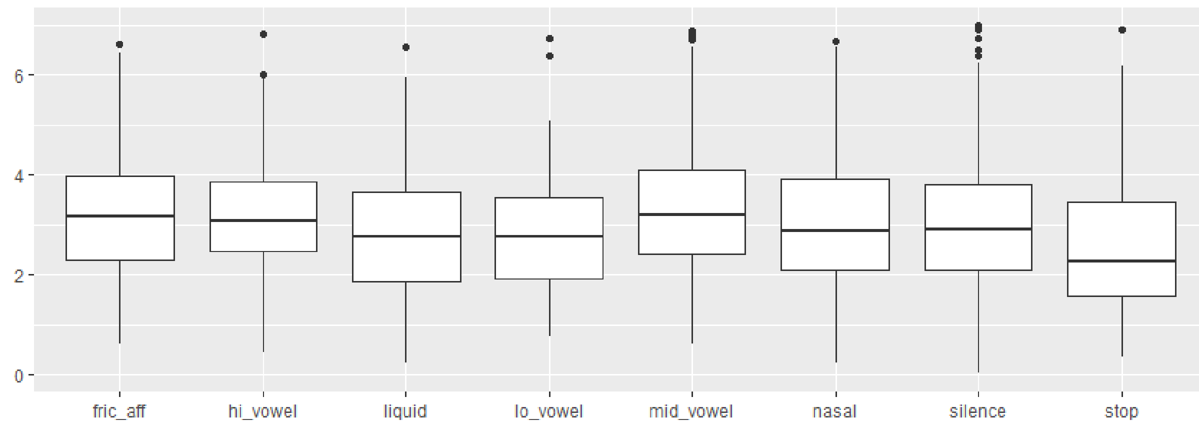

Figure 2. Minimum tongue-tip aperture for /ð/, by preceding context.

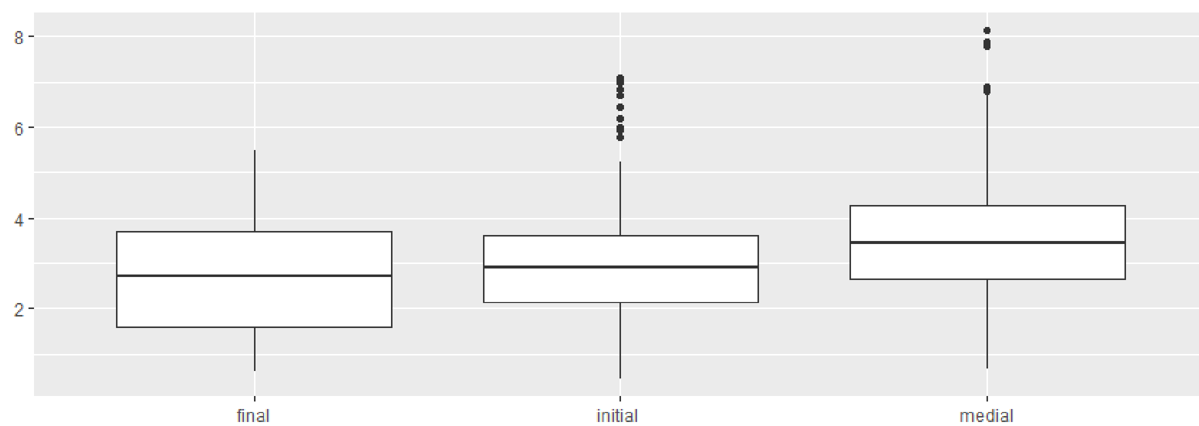

Figure 3. Tongue-tip aperture by word position, for non-stopped data (to control for segmental context, plot shows subset of data where targets are preceded by a vowel).

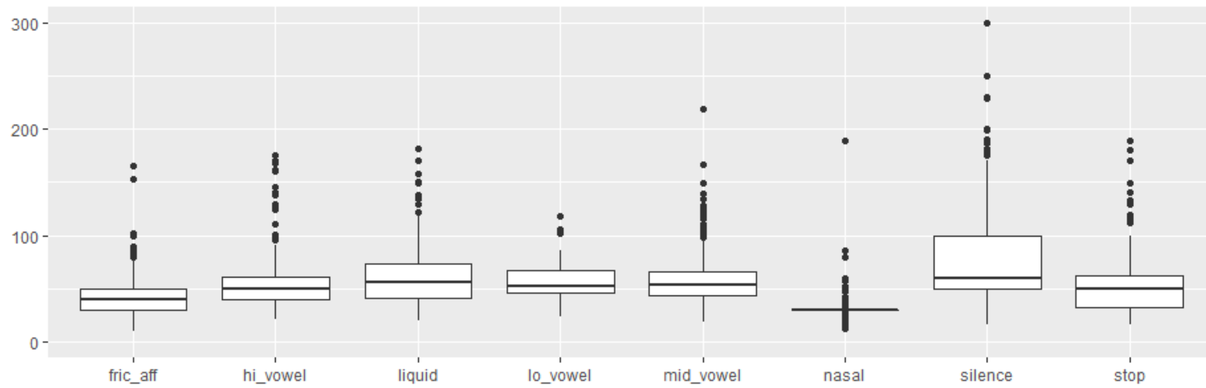

Figure 4. Duration of /ð/, by preceding context.

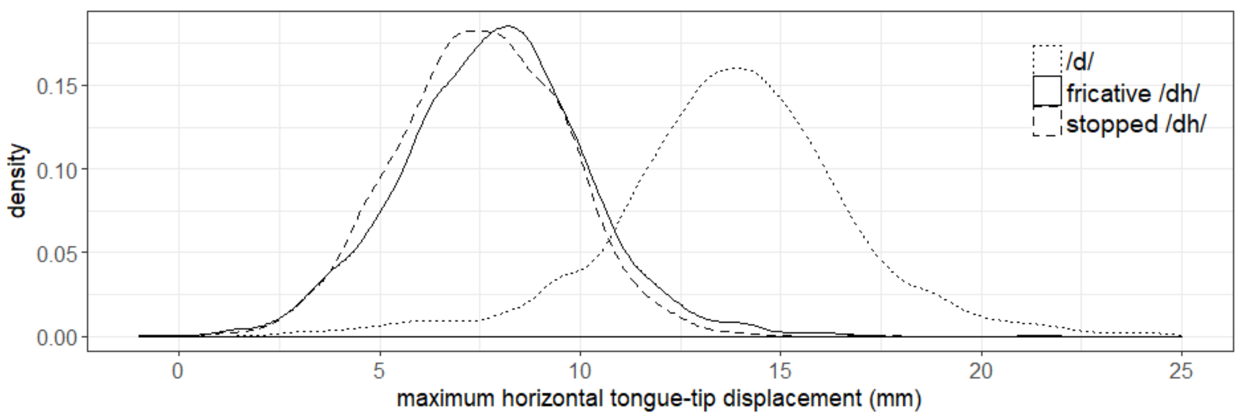

Figure 5. Density plots of maximum tongue-tip x-values for /d/ and stopped and fricative /ð/. Note that distances are relative to a landmark on the mandibular incisors; values closer to 0 indicate a more anterior place of articulation.

4. DISCUSSION

The pattern of results reported in this study provides mixed support for the hypothesis that stopping of dental fricatives in standard American English is the result of gradient strengthening (hyper-articulation) of these segments. On the one hand, consistent with predictions, dental fricatives showed evidence of prosodically-conditioned strengthening at both the phrase and word level, as illustrated by the fact that stopping was more likely domain-initially than domain-medially/-finally in both cases. Also consistent with predictions were the patterns of stopping by segmental context: fricatives showed less stopping after a vowel than after other segmental contexts, and more open vowels disfavored stopping more strongly than more closed ones. On the other hand, however, the present study fails to find a systematic relationship between likelihood of stopping and the phonetic characteristics of dental fricatives are not unambiguously realized as stops. Contexts that favored stopping did not consistently predict phonetic strengthening in the portion of the non-stopped data, nor did contexts which significantly disfavored the likelihood of stopping predict weakening (see Table 4). Despite exhibiting the predicted contextual distribution of stopping and preserving a dental place of articulation in stopped fricatives, the data thus do not fully support a picture of th-stopping in American English as being the endpoint of a gradient gestural strengthening process.

Crucially, what the results do demonstrate is that functional factors play important roles in conditioning "categorical" alternations between continuants and stops. This is shown by the fact that maximization of clarity and minimization of effort conditioned the distribution of stopping as we would expect if the latter were a form of clarity enhancement under H&H Theory. Here, prosodic prominence (domain-initial position) favored stopping, with utterance-initial position emerging as significant, and word-initial position closely approaching significance. This finding is consistent with literature showing the edges of prosodic domains are associated with gestural strengthening (Keating, 2006) and with fortition (Kiparsky, 1988; Kirchner, 1998). The only contradictory result here is the lack of an effect of word-stress, which has been associated with hyper-articulation (Shattuck-Hufnagel & Turk, 1996), and is known to favor fortition (Kiparsky, 1988; Kirchner, 1998). However, lack of a significant effect of this predictor is not completely surprising given the fact that high-frequency function words (which are generally unstressed) contributed to the majority of /ð/ tokens in this study, meaning there may not be statistical power to detect a potential difference between stressed and unstressed positions. Another piece of evidence for the hypothesis that stopping is the result of hyper-articulation is the fact that segmental context conditioned stopping in precisely the way we would expect from an effort-based perspective. Namely, data showed that a preceding vowel or liquid disfavored stopping compared to other preceding segments, and that fricatives in the context of low vowels (whether preceding or following) underwent significantly less stopping than those in the context of other vowels. This is precisely the mirror image of the kinds of aperture-conditioned lenition patterns reported in the literature (Giannelli & Savoia, 1978; Kirchner, 1998; Romero, 1996). This suggests that, all other factors being equal, more articulatorily open contexts will shift the speaker's articulations toward the hypo- end of the continuum, due to the increased effort required to achieve constriction in such contexts. This means that, on average, we get shallower articulations in more "effortful" contexts, and therefore less likelihood that speakers will achieve the degree of target constriction required to produce a stop. Evidence for stopping being a result of hyper-articulation is not limited to contextual factors: higher lexical frequency was a highly significant predictor that disfavored stopping, consistent with literature showing that high-frequency words tend to be reduced or hypo-articulated (Baker & Bradlow, 2009). Thus, it makes sense that that increased frequency would decrease the likelihood of stopping, if the latter is a form of hyper-articulation.[7] Finally, comparison of articulatory differences between fricative and stopped /θ ð/ and the stops /t d/ revealed no evidence for a categorical change in place of articulation between stopped and fricated realizations of /θ ð/. This corroborates the results of acoustic studies of the place of articulation of stopped /ð/ (Zhao, 2010), and is consistent with H&H theoretic predictions, where hyper-articulation may yield more "extreme" gestures but should not lead to categorical changes in articulatory target.

Crucially, however, because H&H theory seeks to model "on-line phonetic patterns," we also expected to see gradient weakening and strengthening effects in contexts that respectively disfavored and favored stopping, even when the target was not completely occluded. Results in this case are mixed. For instance, preceding stops predicted the largest decrease in tongue-tip aperture, followed by preceding low vowels, mid vowels, and liquids.[8] While the finding that the most constricted segments (stops) predicted increased constriction is as expected, the fact that the most open segments in the dataset (low vowels) also lead to increased constriction in the target (relative to fricatives) is clearly at odds with the pattern of segmental conditioning in the categorical data and deeply puzzling from an effort-based perspective. For following context, however, the results were as predicted: both low and mid vowels correlated with increased tongue-tip aperture relative to high vowels, and the magnitude of the effect was greater with a low vowel. The case with prosodic factors' gradient effect on tongue-tip aperture is similarly mixed. On the one hand, word position had the expected effect of decreasing aperture in word-

[7] Although, it should be noted that because the majority of words analyzed here were already relatively high-frequency due to being function words, we cannot conduct a suitable comparison with low-frequency (content) words.

[8] Here, as in the analysis of the burst data, the reference level for preceding environment was a fricative/affricate.

initial position, compared to word-medially. On the other hand, phrase-initial position, which significantly favored stopping, had no significant effect on constriction in the non-stopped data, nor was there a significant effect of word stress. Finally, while the positive relationship between duration and constriction degree was consistent with hypotheses, increased duration did not consistently line up with increased constriction degree in the same context. Phrase-initial position, for instance, did not significantly affect tongue-tip aperture but it did predict longer durations. Similarly, word position significantly affected constriction degree but not duration, while word stress did not significantly affect tongue-tip aperture but did predict longer segmental durations. The effects of segmental context are similarly mixed, with no clear evidence of a systematic relationship between context and observed tongue-tip aperture/duration.

In summary, the results of the present study show that factors which favor hypo-articulation as well as factors that have been correlated with hyper-articulation both predict the likelihood of strengthening of dental fricatives to stops, in opposite directions. Consistent with a functionalist explanation, th-stopping in Standard American English appears to be conditioned by both articulatory economy and the need to maintain articulatory contrast in prosodically prominent positions. However, results fail to show the predicted gradient differences in phonetic realization as a function of segmental and prosodic context, as we would predict under H&H Theory. The data thus do not fully support the hypothesis that stopping is an endpoint of a gradient gestural strengthening process

Rather, the alternation between stopped and non-stopped tokens in th-stopping looks more like a variable phonological process. In a functionalist account of synchronic and diachronic phonological patterns, Blevins (2004) suggests that lenition and fortition are simply phonologized cases of general phonetic processes of gestural reduction and gestural strengthening. This may be what we are seeing in the case of th-stopping, and a possible explanation for why we did not observe the expected patterns of gradience in the production of /θ ð/. Such results mirror the those of Bürki, Ernestus, Gendrot, Fougeron, & Frauenfelder (2011), who conclude from their corpus analysis of French schwa that the deletion of this segment is not due to gradient shortening, as predictors of deletion do not reliably also predict duration in the expected direction. Despite showing contextually conditioned reduction of schwa, the study ultimately fail to find evidence that gradient mechanisms are ultimately responsible for the observed categorical outcomes.

Unfortunately, an account of when and how the kinds of phonetic variation predicted by H&H theory may become part of a speaker's mental grammar lies outside the scope of the present study. What we can conclude, however, is that both minimization of effort and maintenance of perceptual contrast are crucial predictors of variation in spoken language: "Both tendencies are real, both are functional, and both are necessary parts of an understanding of phonology" (Donegan & Stampe, 1979, p. 129-130).

5. DIRECTIONS FOR FUTURE RESEARCH

While this study provides compelling evidence for the importance of functional factors in conditioning the realization of dental fricatives, it leaves a number of potentially interesting questions unexplored. One of these is speech rate: faster speech has been associated with more hypo-articulation (Laan, 1997; Van Son & Pols, 1996) and increased likelihood and degree of consonantal undershoot (Giannelli & Savoia, 1978), suggesting it may decrease the likelihood of Th-stopping. Another factor worth investigating is speech style. Clear/listener-oriented speech has been associated with more hyper-articulation (Bradlow, 2002; Cho et al., 2011; Picheny et al., 1986), which may result in increased rates of Th-stopping. Finally, the present study makes crucial assumptions about articulatory effort and segmental context that are supported by the literature but would benefit from direct investigation. For instance, it assumes that vowel height maps on straightforwardly to the height of the tongue tip, which may not always be the case. For instance, while a high front vowel like /i/ may consistently involve a raised tongue tip, for a high back vowel like /u/, the position of tongue tip may vary much more freely. Therefore, directly measuring the degree of tongue-tip displacement involved during the transition into the target consonant may allow for a more accurate approximation of articulatory effort, and may explain some of the puzzling results obtained in the present study.

References

Baayen, R. H. (2008). *Analyzing linguistic data : a practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Baker, R. E., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, *52*(4), 391–413.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv:1406.5823 [Stat]*. Retrieved from http://arxiv.org/abs/1406.5823

Bell, A., & Gibson, A. (2008). Stopping and Fronting in New Zealand Pasifika English. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 7.

Blevins, J. (2004). *Evolutionary phonology. [electronic resource] : the emergence of sound patterns.* Cambridge, UK ; New York : Cambridge University Press, 2004.

Bradlow, A. R. (2002). Confluent talker-and listener-oriented forces in clear speech production. *Laboratory Phonology*, *7*.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3–4), 155–180.

Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., & Frauenfelder, U. H. (2011). What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *The Journal of the Acoustical Society of America*, *130*(6), 3980–3991. https://doi.org/10.1121/1.3658386

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics*, *24*(2), 263–282.

Cao, Y. A. (2002). *Analysis of acoustic cues for identifying consonant /h/ in continuous speech* (PhD Thesis). Massachusetts Institute of Technology.

Childs, B., De Decker, P., Deal, R., Kendall, T., Thorburn, J., Williamson, M., & Van Herk, G. (2010). Stop signs: The intersection of interdental fricatives and identity in Newfoundland. *University of Pennsylvania Working Papers in Linguistics*, *16*(2), 5.

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*(2), 155–190. https://doi.org/10.1006/jpho.2001.0131

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, *39*(3), 344–361. https://doi.org/10.1016/j.wocn.2011.02.005

Davidson, L. (2006). Schwa Elision in Fast Speech: Segmental Deletion or Gestural Overlap? *Phonetica*, *63*(2–3), 79–112. https://doi.org/10.1159/000095304

De Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, *26*(3), 283–310.

Donegan, P. J., & Stampe, D. (1979). The study of natural phonology. *Current Approaches to Phonological Theory*, *126173*.

Donegan, P. J., & Stampe, D. (2009). Hypotheses of Natural Phonology. *Poznań Studies in Contemporary Linguistics*, *45*(1), 1–31. https://doi.org/10.2478/v10010-009-0002-x

Dubois, S., & Horvath, B. (1998). Let’s tink about dat: Interdental fricatives in Cajun English. *Language Variation and Change*, *10*, 245–261.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, *101*(6), 3728–3740. https://doi.org/10.1121/1.418332

Fukaya, T., & Byrd, D. (2005). An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the International Phonetic Association*, *35*(1), 45–58. https://doi.org/10.1017/S0025100305001891

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., & Zue, V. (1993). Timit acoustic-phonetic continuous speech corpus. *Linguistic Data Consortium, Philadelphia*.

Giannelli, L., & Savoia, L. M. (1978). L'indebolimento consonantico in Toscana. *RID. Rivista Italiana Di Dialettologia Bologna*, *2*(1), 23–58.

Hillenbrand, J. M. (2003). American English: Southern Michigan. *Journal of the International Phonetic Association*, *33*(1), 121–126.

Hyman, L. M. (1975). *Phonology: theory and analysis*. Harcourt College Pub.

Johnson, K. (2004). Automatic burst detection in conversational speech. *Unpublished Manuscript*.

Jun, S.-A. (1996). *The phonetics and phonology of Korean prosody : intonational phonology and prosodic structure.* New York : Garland Pub., 1996.

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., & Raymond, W. D. (1998). Reduction of English function words in switchboard. In *ICSLP*.

Keating, P. A. (2006). Phonetic encoding of prosodic structure. *Speech Production: Models, Phonetic Processes, and Techniques*, 167–186.

Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2004). Domain-initial articulatory strengthening in four languages. *Phonetic Interpretation: Papers in Laboratory Phonology VI*, 143–161.

Kiparsky, P. (1988). Phonological change. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey, Vol. I, Linguistic Theory: Foundations* (pp. 363–415). Cambridge: Cambridge University Press.

Kirchner, R. (1998). *An effort based approach to consonant lenition* (Doctoral Dissertation). University of California, Los Angeles.

Krull, D. (1989). Consonant-vowel coarticulation in spontaneous speech and in reference words. *Papers in Linguistics from the University of Stockholm*, *10*, 101–105.

Laan, G. P. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, *22*(1), 43–65.

Levshina, N. (2015). *How to Do Linguistics with R : Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins Publishing Company.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, *35*(11), 1773–1781.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht: Springer.

Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *The Journal of the Acoustical Society of America*, *99*(3), 1683–1692. https://doi.org/10.1121/1.414691

Lindblom, B., Sussman, H. M., & Agwuele, A. (2009). A Duration-Dependent Account of Coarticulation for Hyper- and Hypoarticulation. *Phonetica*, *66*(3), 188–195. https://doi.org/10.1159/000235660

Mesthrie, R., & Bhatt, R. M. (2008). *World Englishes: The Study of New Linguistic Varieties*. Cambridge, MA: Cambridge University Press.

Newlin-Lukowicz, L. (2013). TH-stopping in New York City: substrate effect turned ethnic marker? *University of Pennsylvania Working Papers in Linguistics, 19*(2), 17.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*(4), 434–446.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raymond, W. D., Dautricourt, R., & Hume, E. (2006). Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, *18*(01). https://doi.org/10.1017/S0954394506060042

Recasens, D., Pallares, M. D., & Fontdevila, J. (1995). Co-articulatory variability and articulatory-acoustic correlations for consonants. *International Journal of Language & Communication Disorders*, *30*(2), 203–212. https://doi.org/10.3109/13682829509082531

Resnick, M. C. (1975). *Phonological variants and dialect identification in Latin American Spanish.* The Hague : Mouton, 1975.

Romero, J. (1996). *Gestural Organization in Spanish: an Experimental Study of Spirantization and Aspiration* (Doctoral Dissertation). University of Connecticut.

Rose, M. (2006). *Language, place, and identity in later life* (Doctoral Dissertation). Stanford University.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247.

Sussman, H. M., Hoemeke, K. A., & Ahmed, F. S. (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *The Journal of the Acoustical Society of America*, *94*(3), 1256–1268. https://doi.org/10.1121/1.408178

Van Herk, G., Childs, B., & Thorburn, J. (2009). Identity marking and affiliation in an urbanizing Newfoundland community. *Papers from the 31st Annual Meeting of the Atlantic Provinces Linguistic Association*, 85–94.

Van Son, R. J., & Pols, L. C. (1996). An acoustic profile of consonant reduction. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 3, pp. 1529–1532). IEEE.

Vaux, B. (2000). Flapping in English. *Linguistic Society of America, Chicago, IL*.

Westbury, J., Milenkovic, P., Weismer, G., & Kent, R. (1990). X-ray microbeam speech production database. *The Journal of the Acoustical Society of America*, *88*(S1), S56.

Yuan, J., & Liberman, M. (2009). Investigating/l/variation in English through forced alignment. Presented at the Tenth Annual Conference of the International Speech Communication Association.

Zhao, S. (2007). *The stop-like modification of /ð/: a case study in the analysis and handling of speech variation* (Doctoral Dissertation). Massachusetts Institute of Technology.

Zhao, S. (2010). Stop-like modification of the dental fricative /ð/: An acoustic analysis. *The Journal of the Acoustical Society of America*, *128*(4), 2009–2020. https://doi.org/10.1121/1.3478856

Appendix A

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: burst ~ p_stress_following + p_manner_prev + p_manner_following +
    Lg10WF + part_speech + word_pos +      (1 | word) + (1 | speaker)
   Data: d_dh
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 10000))

     AIC      BIC   logLik deviance df.resid
  8380.8   8498.5  -4173.4   8346.8     7469

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.4742 -0.6472 -0.3853  0.8141  5.8242

Random effects:
 Groups  Name        Variance Std.Dev.
 speaker (Intercept) 0.2671   0.5168
 word    (Intercept) 0.0000   0.0000
Number of obs: 7486, groups:  speaker,
41; word, 18
```

Fixed Effects	Estimate	Std. Error	z-value	Pr(>|z|)	
(intercept)	0.594922	0.390857	1.522	0.127985	
word stress = stressed	-0.047646	0.091614	-0.52	0.603016	
preceding context = high vowel	0.070663	0.168929	0.418	0.675728	
preceding context = liquid	-0.463699	0.131553	-3.525	0.000424	***
preceding context = low vowel	-0.731714	0.297157	-2.462	0.013802	*
preceding context = mid vowel	-0.337065	0.177772	-1.896	0.057953	.
preceding context = nasal	-0.97437	0.146301	-6.66	2.74E-11	***
preceding context = pause/silence	1.018932	0.095345	10.687	< 2e-16	***
preceding context = stop	1.41857	0.111802	12.688	< 2e-16	***
following context = low vowel	-0.335227	0.114634	-2.924	0.003452	**
following context = mid vowel	-0.005822	0.068581	-0.085	0.932347	
log word frequency	-0.477944	0.080422	-5.943	2.80E-09	***
part of speech = function word	0.491116	0.385412	1.274	0.202571	
word position = final	1.29729	0.519336	2.498	0.01249	*
word position = initial	0.738846	0.384881	1.92	0.054899	.

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix B

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: T1_min_ap ~ dur + p_stress_following + p_manner_prev + p_manner_following +
Lg10WF + part_speech +
relevel(word_pos, ref = "medial") + (1 |
word) + (1 | speaker)
Data: d_dh_fric_noburst

REML criterion at convergence: 8718.9

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.9120	-0.5917	-0.0888	0.4991	14.0653

Random effects:

Groups	Name	Variance	Std.Dev.
speaker	(Intercept)	0.50769	0.7125
word	(Intercept)	0.02112	0.1453
Residual		1.22142	1.1052

obs: 2810, groups: speaker, 41; word, 17

Fixed Effects:	Estimate	Std. Error	t-value	Pr(>|t|)	
(intercept)	3.17E+00	4.28E-01	7.402	3.70E-05	***
duration	-3.25E-03	6.78E-04	-4.787	1.80E-06	***
word stress = stressed	2.43E-02	6.85E-02	0.354	0.72316	
preceding context = high vowel	3.12E-02	1.19E-01	0.263	0.79238	
preceding context = liquid	-2.00E-01	8.98E-02	-2.225	0.02621	*
preceding context = low vowel	-3.79E-01	1.82E-01	-2.088	0.03992	*
preceding context = mid vowel	-2.77E-01	1.14E-01	-2.422	0.01551	*
preceding context = nasal	-1.16E-01	8.08E-02	-1.435	0.15134	
preceding context = pause/silence	-4.86E-02	7.92E-02	-0.613	0.54005	
preceding context = stop	-6.69E-01	1.07E-01	-6.243	9.62E-10	***
following context = low vowel	4.94E-01	1.53E-01	3.223	0.00834	**
following context = mid vowel	3.39E-01	6.51E-02	5.212	3.20E-07	***
log word frequency	1.03E-01	9.56E-02	1.076	0.3268	
part of speech = function word	1.30E-02	3.10E-01	0.042	0.96701	
word position = final	-6.32E-01	4.42E-01	-1.43	0.16888	
word position = initial	-6.91E-01	3.32E-01	-2.084	0.04863	*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix C

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: dur ~ T1_min_ap + p_stress_following + p_manner_prev + p_manner_following + Lg10WF + part_speech + relevel(word_pos, ref = "medial") + (1 | word) + (1 | speaker)
Data: d_dh_fric_noburst

REML criterion at convergence: 27197.5

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.4683	-0.4572	-0.0940	0.2688	30.5725

Random effects:

Groups	Name	Variance	Std.Dev.
speaker	(Intercept)	20.54	4.532
word	(Intercept)	237.92	15.425
Residual		938.42	30.634

obs: 2810, groups: speaker, 41; word, 17

Fixed effects:

	Estimate	Std. Error	t value	Pr(>|t|)	
(intercept)	60.7914	26.3153	2.31	0.04066	*
tongue-tip aperture	-2.0447	0.4881	-4.189	2.99E-05	***
word stress = stressed	4.6998	1.9658	2.391	0.01688	*
preceding context = high vowel	7.6873	3.3168	2.318	0.02054	*
preceding context = liquid	12.8591	2.5247	5.093	3.76E-07	***
preceding context = low vowel	13.5863	5.766	2.356	0.01861	*
preceding context = mid vowel	9.3205	3.2062	2.907	0.00368	**
preceding context = nasal	-9.5521	2.2718	-4.205	2.70E-05	***
preceding context = pause/silence	25.8353	2.1765	11.87	< 2e-16	***
preceding context = stop	-8.7611	3.132	-2.797	0.00519	**
following context = low vowel	7.3553	6.0537	1.215	0.22538	
following context = mid vowel	-8.3732	1.8985	-4.41	1.07E-05	***
log word frequency	-0.5878	6.7389	-0.087	0.93219	
part of speech = function word	20.2334	15.3771	1.316	0.20362	
word position = final	-16.092	24.2333	-0.664	0.51759	
word position = initial	-10.1334	16.8733	-0.601	0.55558	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1