PHONETIC STRENGTHENING OR PHONOLOGICAL SUBSTITUTION? AN X-RAY MICROBEAM INVESTIGATION OF ACOUSTIC AND ARTICULATORY VARIABILITY IN THE PRODUCTION OF AMERICAN ENGLISH DENTAL FRICATIVES

Yevgeniy Vasilyevich Melguy

University of California, Berkeley
ymelguy@berkeley.edu

ABSTRACT

Research has shown that many phenomena traditionally treated as categorical (e.g., t/d-deletion) may actually be the outcome of gradient phonetic processes. The current study extends this approach to th-stopping, the realization of dental fricatives /θ ð/ as stops, to test whether it is better explained as the outcome of gradient phonetic strengthening, rather than a categorical phonological process, as generally assumed. Analysis of the acoustic and articulatory characteristics of /ð/ in a corpus of Standard American English speech shows that, as predicted, stopped /ð/ retains a dental place of articulation and is more likely to occur in ”strong” (prominent) prosodic contexts while being disfavored in ”weak” (articulatorily open) segmental contexts. However, comparison with non-categorical measures of consonant strength (duration and tongue-tip constriction) shows that the latter do not systematically align with predictors of stopping likelihood. These findings suggest that th-stopping is not the endpoint of gradient gestural strengthening.

Keywords: phonetic gradience, gestural strengthening, variability, dental frictives, corpus phonetics

1. INTRODUCTION

Empirical work over the past several decades has provided evidence that many cases of weakening/reduction that have traditionally been treated as categorical phonological alternations may in fact be the result of gradient phonetic processes. This has been argued, for instance, with English t/d-deletion [4, 23], alveolar flapping [11, 9], and schwa deletion [8].

However, these kinds of studies have generally been limited to examining weakening/lenition phenomena, which involve a decrease in gestural constriction or duration [5]. By contrast, strengthening/fortition processes, which involve an increase in consonantal constriction or duration have not attracted similar interest. A promising case study along such a line of inquiry is th-stopping, a well-documented process across dialects of English that involves the realization of dental fricatives /θ ð/ as stops [3]. Although th-stopping has generally been treated as a simple categorical alternation between stops and fricatives (e.g., [7]), there are two main bodies of evidence that support a more nuanced picture of this process.

The first comes from acoustic studies of dental fricatives in continuous speech, which suggests that their phonetic realization in fact spans a whole spectrum of variation, from deletion on the one extreme [15] to complete occlusion on the other [28], and with intermediate affricate realizations also attested [18, 10]. In addition, comparisons of stopped /ð/ with /d/ in Standard American English have shown that the two sounds have distinct acoustic profiles [29], suggesting that stopping preserves a dental place of articulation and involves no major change in articulatory target.

The second body of evidence comes from the distribution of th-stopping, which is conditioned by the same factors that have been shown to affect the gradient phonetic realization of consonants elsewhere. For example, th-stopping has been shown to be more likely at prosodic domain edges such as utterance- or phrase-initially, with the likelihood of stopping increasing as one advances up the prosodic hierarchy [28]. These same (domain-initial) prosodic contexts have been shown to predict gradient consonant strengthening (increased linguopalatal contact and longer segmental durations) across languages [16], displaying the same cumulative pattern. At the same time, segmental contexts that disfavor th-stopping have also been shown to predict gradient weakening of consonants elsewhere. For instance, stopping is strongly disfavored when the target is flanked by an open segment, such as a preceding vowel. Research has shown that the aperture of flanking vowels can affect consonant constriction, with the degree of weakening generally increasing in more articulatorily open contexts [17]. For example, Romero [25] showed that Spanish stops systematically exhibit less articulatory constriction in the context of a low vowel than a high vowel, presumably due to the increased distance the tongue must travel in the former case.[1]

If th-stopping is a gradient endpoint of gestural strengthening, then contextual factors that have been linked to such strengthening (domain-initial position) should predict not only higher likelihood of stopping, but also increased target duration/constriction, even when the target is not stopped. Analogously, factors associated with gradient weakening of consonants (open/low vowels) should predict both reduced likelihood of stopping and a reduction in duration/constriction. By contrast, if th-stopping is a phonological process, we may observe gradience in the phonetic realization of the target, but expect no correlation between likelihood of stopping and consonant strength. A change in place (e.g., to alveolar) would also support a phonological explanation, because a phonetic mechanism does not predict a change in articulatory target.

2. METHODOLOGY

2.1. Data

To test these questions, the present study examines the effects of prosodic, segmental, and lexical factors on the phonetic realization of 7486 tokens of /ð/ in the X-ray Microbeam Speech Production Database (XRMBDB) [26]. The XRMBDB utilizes a point-tracking system to monitor the position of small (~3mm) gold pellets attached to the active articulators 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. Many sentences found in the XRBMDB were selected directly from previous speech production corpora such as the TIMIT database [12]. 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. In addition, because of model convergence issues due to the relatively small number of /θ/ tokens (1229) in the data, only the results of the /ð/ analysis are reported here.

2.2. Data preparation

Acoustic recordings from the XRMBDB were automatically transcribed using the Penn Forced Aligner [27] and subsequently hand-corrected by research assistants trained in phonetic transcription. A burst detector [14] was then run on each /θ ð/ token in the data. This script detects burst-like transients in the speech signal by comparing successive 5 ms windows in the spectrum and waveform, searching for the areas of maximum difference, and assigns a burst-strength score using a linear discriminate function trained on stop bursts in the TIMIT corpus. For /ð/, this yielded burst-strength values ranging from 0 to 7.02 (M = 0.90, SD = 1.25). For reference, values for /d/ ranged from 0 to 7.50 (M = 1.03, SD = 1.82).

A threshold for burst strength was then selected and tokens were automatically classified as stopped or non-stopped, based on whether they met this cutoff. This was done by selecting a single subject and binning all tokens of /θ ð/ in their data for which a burst score was returned into 0.5 increment burst-score windows. Following previous studies that have used the presence of a burst as a marker of stopping [28], tokens in each window were manually hand-checked to ensure the presence of a visible burst in the spectrum and waveform as well as an audible burst in the sound file. Each token was classified as stopped only if it met all three criteria. The threshold was then set at the lowest value of burst strength above which 70% or more of tokens were classified as being stopped. A total of 136 tokens were thus analyzed, and the threshold was set at 1.0, corresponding to the 1.0-1.5 window. At the selected threshold, 36% of /ð/ tokens were classified as stopped (compared to 50% of /d/ tokens).

Tongue-tip aperture for each token of /ð/ was also determined by calculating the minimum distance between the (x,y) coordinates of the tongue-tip pellet (T1) and those of the palate (Pal), using an adapted Python script (Susan Lin, p.c.) with the equation in (1).

(1) $$TT\ \ ap = \sqrt{(T1_{\mathrm{x}} - Pal_{\mathrm{x}})^2 + (T1_{\mathrm{y}} - Pal_{\mathrm{y}})^2}$$

Finally, duration was calculated based on the hand-corrected segmental boundaries for each token of /ð/, and place of articulation was calculated by taking the maximum x-value of the tongue-tip pellet for each token of stopped /ð/, fricative /ð/, and /d/.

2.3. Analysis

Likelihood of stopping, tongue-tip (TT) aperture, and duration were analyzed with mixed-effects regression modeling using the lme4 package [1] in the statistical software R [22]. Logistic regression [13] was used to predict the likelihood of stopping, while the continuous measures (TT aperture and duration) were analyzed using linear models, fitted to the subset of tokens that were not stopped (since stopped tokens will necessarily be more constricted then non-stopped ones). All three models had near-identical effects structure (the linear models also included duration or TT aperture as predictors in order to test the relation between these variables, following previous studies [16]). All three models included random intercepts for speaker and word, and fixed effects for word position (word-initial, -medial, or -final), preceding context (low, mid, or high vowel; fricative/affricate, liquid, nasal, stop, or pause), following context (low, mid, or high vowel), lexical stress (stressed or unstressed, based on hand-corrected phonetic transcriptions), lexical class (function or content word) and lexical frequency (log10 frequency measures from the SUBTLEX corpus [6]). Although the lexical factors (word class, word stress, 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 [7, 10, 21, 2]. They were thus included as controls even if they failed to significantly improve model fit. Finally, place of articulation of stopped /ð/, fricative /ð/, and /d/ were analyzed using mixed-effects linear regression, with random intercepts for speaker and word and a single main effect of phone.

3. RESULTS

Results showed that prosodic, segmental, and lexical factors were significant predictors of stopping (Table 1). A likelihood ratio test revealed that both preceding and following context were significant ($\chi^2(7)$ = 559.25, p < 0.001 and $\chi^2(2)$ = 9.85, p < 0.01, respectively), as were word position ($\chi^2(2)$ = 6.39, p < 0.05) and word frequency ($\chi^2(1)$ = 35.34, p < 0.001). Predictors that favored stopping included a preceding stop, phrase-initial position (preceding pause), and word-final position and word-intitial position (although the latter only approached significance). Factors that disfavored stopping included high lexical frequency, a preceding liquid, or nasal, and, as predicted, a flanking low vowel (whether preceding or following).

Tongue-tip aperture of /ð/ (Table 2) was conditioned by preceding context ($\chi^2(7)$ = 54.90, p < 0.001), following context ($\chi^2(2)$ = 29.50, p < 0.001), and duration ($\chi^2(1)$ = 22.91, p < 0.001). Word position approached significance ($\chi^2(2)$ = 4.34, p = 0.114). Factors predicting decreased TT aperture (greater constriction) included increased duration, word-initial position, and a preceding stop, low vowel, mid vowel, or liquid. The reverse pattern was observed for following vowels: in both cases, as predicted, low and mid vowels predicted increased tongue-tip aperture relative to high vowels, with a greater effect observed for low vowels.

Duration of /ð/ (Table 2) was significantly affected by preceding context ($\chi^2(7)$ = 348.04, p < 0.001),

Table 1: Model estimates for log likelihood of stopped /ð/.

	Est.	SE
(intercept)	0.60	0.39
following low vowel	-0.34**	0.12
following mid vowel	-0.006	0.07
preceding low vowel	-0.73*	0.30
preceding mid vowel	-0.34·	0.18
preceding high vowel	0.071	0.17
preceding liquid	-0.46***	0.13
preceding nasal	-0.97***	0.15
preceding stop	1.42***	0.11
preceding pause	1.02***	0.10
target word-final	1.30*	0.52
target word-initial	0.74 ·	0.39
stressed word	-0.05	0.09
function word	0.49	0.39
log word frequency	-0.48***	0.080

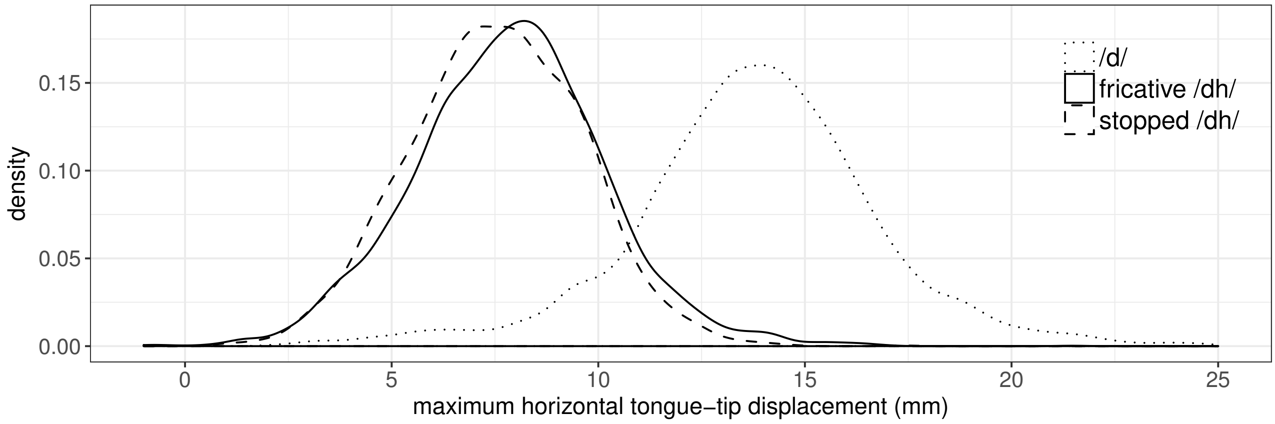
*** p < 0.001, ** p < 0.01, * p < 0.05, · p < 0.1

Table 2: Model estimates for /ð/ stopping likelihood, minimum tongue-tip aperture, and duration. Results consistent with hypotheses are in boldface.

	stop	TT ap	dur
(intercept)	0.59	31.70***	60.79*
foll low vow	**-0.34****	**4.94****	7.36
foll mid vow	-0.01	**3.39*****	**-8.37*****
prec low vow	**-0.73***	-3.79*	13.59*
prec mid vow	**-0.34·**	-2.77*	9.32**
prec high vow	0.07	0.31	7.69*
prec liquid	-0.46***	-2.00*	12.86***
prec nasal	-0.97***	-1.16	-9.55***
prec stop	1.42***	-6.69***	-8.76**
prec pause	**1.02*****	-4.94	**25.84*****
word-final	1.30*	-6.32	-16.09
word-initial	**0.74·**	**-6.91***	-10.13
stressed word	-0.05	00.24	4.70*
log word freq	-0.48***	1.03	-0.59
func word	0.49	0.13	20.23
TT aperture	–	–	-2.05***
duration	–	-0.03***	–

*** p < 0.001, ** p < 0.01, * p < 0.05, · p < 0.1

Figure 1: 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.



following context ($\chi^2(2) = 27.53$, $p < 0.001$), lexical stress ($\chi^2(1) = 5.71$, $p < 0.05$), and tongue-tip aperture ($\chi^2(1) = 17.55$, $p < 0.001$). Increased duration was observed with decreased TT aperture, in phrase-initial position, in stressed words, as well as with preceding high vowels, mid vowels, liquids, and low vowels (in increasing order of magnitude). Decreased duration was observed with preceding stops and nasals. Contra predictions, following mid vowels predicted increased duration compared to high vowels.

Finally, results showed significant place of articulation differences between fricative /ð/, stopped /ð/, and /d/ ($\chi^2(2) = 315.78$, $p < 0.001$), which had mean tongue-tip x-values of 7.87 mm, 7.54 mm, and 13.81 mm, respectively (Figure 1).

4. DISCUSSION

The pattern of results reported in this study provides mixed support for the hypothesis that stopping of /ð/ in Standard American English is the result of gradient gestural strengthening. On the one hand, consistent with hypotheses, prosodic contexts that have been shown in previous research to predict gradient strengthening of consonants also favored stopping of /ð/, which was more likely domain-initially than domain-medially. Moreover, factors that have been shown to result in gradient weakening of consonants (articulatorily open flanking segments) also decreased the likelihood of stopping, such that low vowels (whether preceding or following) disfavored stopping more strongly than non-low vowels. Finally, both stopped /ð/ and fricative /ð/ displayed a substantially more anterior tongue-tip than /d/, suggesting that stopping does not involve a change in articulatory target.

On the other hand, however, the present study fails to find a reliable relationship between likelihood of stopping and gradient measures of consonant strength (duration and tongue-tip constriction). Not a single significant predictor of stopping was able to also predict the patterns of both TT aperature and duration in the expected direction. Factors correctly predicted to favor stopping (domain-initial position) did not reliably predict increased constriction or duration in the portion of the non-stopped data, nor did factors correctly predicted to disfavor stopping (low vowels) reliably predict weakening. A mismatch between stopping likelihood of /ð/ and TT aperture/duration elsewhere may be due to assimilation to preceding segments, which has been attested with laterals [20], nasals [19], and stops [28]).

Thus, the data do not fully support a picture of /ð/-stopping in American English as being the extreme endpoint of a gradient gestural strengthening process. While the contextual distribution of stopping was as predicted, the disconnect between the gradient realization of the target and the likelihood of stopping is inconsistent with a phonetic mechanism. Rather, /ð/-stopping appears to be a variable phonological substitution process that is sensitive to fine-grained phonetic and prosodic factors.

5. ACKNOWLEDGEMENTS

I am grateful to Susan Lin and Brian Smith for their helpful comments on earlier versions of this paper, to Ronald Sprouse for assistance with data preparation, and to Keith Johnson for his suggestions on data analysis and presentation. This research was supported in part by an NSF GRFP fellowship.

6. REFERENCES

[1] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1).

[2] Bell, A., Gibson, A. 2008. Stopping and fronting in new zealand pasifika english. *University of Pennsylvania Working Papers in Linguistics* 14(2), A7.

[3] Bhatt, R. M., Mesthrie, R. 2008. *World Englishes: The study of new linguistic varieties*. New York: Cambridge University Press.

[4] Browman, C., Goldstein, L. 1990. Tiers in articulatory phonology, with some implications for casual speech. In: Beckman, M. E., Kingston, J., (eds), *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press 341–376.

[5] Browman, C., Goldstein, L. 1991. Gestural structures: Distinctiveness, phonological processes, and historical change. In: Mattingly, I. G., Studdert-Kennedy, M., (eds), *Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman*. Hillsdale, NJ: Lawrence Erlbaum Associates 313–338.

[6] Brysbaert, M., New, B., Keuleers, E. 2012. Adding part-of-speech information to the subtlex-us word frequencies. *Behavior research methods* 44(4), 991–997.

[7] 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), A5.

[8] Davidson, L. 2006. Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica* 63(2-3), 79–112.

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

[10] Dubois, S., Horvath, B. M. 1998. Let's tink about dat: Interdental fricatives in cajun english. *Language Variation and Change* 10(3), 245–261.

[11] 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.

[12] Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S. 1993. *TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1*. Philadelphia: Linguistic Data Consortium.

[13] Jaeger, T. F. 2008. Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language* 59(4), 434–446.

[14] Johnson, K. 2004. Automatic burst detection in conversational speech. Unpublished manuscript.

[15] Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., Raymond, W. Reduction of english function words in switchboard. *Fifth International Conference on Spoken Language Processing* Sydney.

[16] Keating, P., Cho, T., Fougeron, C., Hsu, C.-S. 2004. Domain-initial articulatory strengthening in four languages. In: Local, J., Ogden, R., Temple, R., (eds), *Phonetic interpretation: Papers in laboratory phonology VI*. Cambridge: Cambridge University Press 143–161.

[17] Kirchner, R. 1998. *An effort-based account of consonant lenition*. PhD thesis University of California, Los Angeles.

[18] Labov, W. 1966. *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.

[19] Manuel, S. Y. 1995. Speakers nasalize /ð/ after /n/, but listeners still hear /ð/. *Journal of Phonetics* 23(4), 453–476.

[20] Manuel, S. Y., Wyrick, G. C. 1999. Casual speech: A rich source of intriguing puzzles. *Proceedings of the XIV International Congress of Phonetic Sciences* San Francisco. 679–682.

[21] Newlin-Łukowicz, L. 2013. Th-stopping in new york city: Substrate effect turned ethnic marker? *University of Pennsylvania Working Papers in Linguistics* 19(2), A17.

[22] R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

[23] 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(1), 55–97.

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

[25] Romero, G. J. 1996. *Gestural organization in Spanish: An experimental study of spirantization and aspiration*. PhD thesis University of Connecticut.

[26] Westbury, J., Milenkovic, P., Weismer, G., Kent, R. 1990. X-ray microbeam speech production database. *The Journal of the Acoustical Society of America* 88(1), S56.

[27] Yuan, J., Liberman, M. 2009. Investigating /l/ variation in english through forced alignment. *Tenth Annual Conference of the International Speech Communication Association* Brighton. 2215–2218.

[28] Zhao, S. Y. 2007. *The stop-like modification of/ð/: a case study in the analysis and handling of speech variation*. PhD thesis Massachusetts Institute of Technology.

[29] Zhao, S. Y. 2010. Stop-like modification of the dental fricative /ð/: An acoustic analysis. *The Journal of the Acoustical Society of America* 128(4), 2009–2020.

[1] Such coarticulatory effects, as Kirchner [17] observes, may explain the different reduction patterns that have been observed in Spanish past-participle allomorphs, where /d/ is normally realized as a fricative or approximant after a high vowel (/-ido/ → [iðo] ~ [ið̞o]) but often lost entirely after a low vowel (/-ado/ → [ao̯]) [24].