

AUGUST 27 2024

Evaluating the consistency of lenition measures: Neural networks' posterior probability, intensity velocity, and duration

Kevin Tang ; Ratreë Wayland ; Fenqi Wang ; Sophia Vellozzi ; Rahul Sengupta 



J. Acoust. Soc. Am. 156, 1367–1379 (2024)

<https://doi.org/10.1121/10.0028299>



ASA

Advance your science and career as a member of the
Acoustical Society of America

[LEARN MORE](#)



ASA
ACOUSTICAL SOCIETY
OF AMERICA

Evaluating the consistency of lenition measures: Neural networks' posterior probability, intensity velocity, and duration

Kevin Tang,^{1,2,a),b)}  Ratree Wayland,^{2,c)}  Fenqi Wang,^{3,d)}  Sophia Vellozzi,^{4,e)}  and Rahul Sengupta^{4,f)} 

¹Department of English Language and Linguistics, Institute of English and American Studies, Faculty of Arts and Humanities, Heinrich Heine University Düsseldorf, Düsseldorf, 40225, Germany

²Department of Linguistics, University of Florida, Gainesville, Florida, 32611-5454, USA

³Department of Linguistics, Simon Fraser University, Burnaby, British Columbia V5A-1S6, Canada

⁴Department of Computer and Information Science and Engineering, University of Florida, 32611-116120, USA

ABSTRACT:

Predictions of gradient degree of lenition of voiceless and voiced stops in a corpus of Argentine Spanish are evaluated using three acoustic measures (minimum and maximum intensity velocity and duration) and two recurrent neural network (Phonet) measures (posterior probabilities of sonorant and continuant phonological features). While mixed and inconsistent predictions were obtained across the acoustic metrics, sonorant and continuant probability values were consistently in the direction predicted by known factors of a stop's lenition with respect to its voicing, place of articulation, and surrounding contexts. The results suggest the effectiveness of Phonet as an additional or alternative method of lenition measurement. Furthermore, this study has enhanced the accessibility of Phonet by releasing the trained Spanish Phonet model used in this study and a pipeline with step-by-step instructions for training and inferring new models. © 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1121/10.0028299>

(Received 8 November 2023; revised 31 July 2024; accepted 1 August 2024; published online 27 August 2024)

[Editor: Jody Kreiman]

Pages: 1367–1379

I. INTRODUCTION

Lenition is a collection of phonological processes involving reduction of the size and duration of consonantal constriction gestures (Kirchner, 2004). The goal of this study is to compare an acoustic approach to a computational method known as Phonet to quantify degree of lenition. In the acoustic approach, a value along an acoustic dimension is a direct measure of lenition. In particular, this work focuses on a well-established measure of lenition proposed by Kingston (2008), which captures the changes in a segment's intensity. In contrast, in the computational approach, lenition degree is indicated by posterior probabilities of relevant phonological features estimated by Phonet. The effectiveness of the two approaches is evaluated on Spanish stops in an Argentine Spanish corpus against known factors affecting varying degrees of lenition, including voicing (voiceless or voiced), preceding segment (vowel or nasal), stress (stressed or unstressed), and openness of flanking

vowels (close, mid, or open). To improve the accessibility of Phonet as an alternative measure of lenition, we released both the trained model used in this study and a pipeline with step-by-step instructions for training and inferring new models (see the Data Availability section.).

According to Kirchner (2004), lenition is an articulatory-reduction strategy driven by the grammatical constraints called LAZY in Optimality Theoretic accounts, which stipulate that pronunciation of any given sound should be achieved with as little effort as possible. On the other hand, Kingston (2008) views lenition as a perceptual strategy: a more open articulation increases the affected consonants intensity and reduces the interruption of the stream of speech, hence conveying that the affected consonant is inside a prosodic constituent. Similarly, Katz (2016) and Harris *et al.* (2024) argued that lenition provides listeners with cues to prosodic and morphosyntactic parsing. Independent of the view adopted, it is generally accepted that in Spanish, lenited voiceless stops become partially or totally voiced (e.g., Lewis, 2001; Martínez Celdrán and Regueira, 2008), and lenited voiced stops become voiced fricatives [β, ð, ʝ] or approximants [β̞, ð̞, ʝ̞] in intervocalic and all other positions, except after a nasal,¹ a pause and, in the case of /d/, an /l/ (e.g., Martínez Celdrán, 2001). However, the complementary distribution of lenited vs non-lenited segments is undermined by studies revealing

^{a)}Email: kevin.tang@hhu.de

^{b)}Current address: Department of English Language and Linguistics, Institute of English and American Studies, Faculty of Arts and Humanities, Heinrich Heine University Düsseldorf, Düsseldorf, 40225, Germany.

^{c)}Email: ratree@ufl.edu

^{d)}Email: fenqiw@sfu.ca

^{e)}Email: s.vellozzi@ufl.edu

^{f)}Email: rahulseng@ufl.edu

gradient phonetic effects of lenition (e.g., Eddington, 2011; Lewis, 2001; Tetzloff, 2020).

A. Lenition conditioning factors

The most frequently cited factor of lenition is speaking rate. A strong positive correlation between length and degree of oral constriction was found in lenited Spanish voiced stops by Soler and Romero (1999). Kirchner (2001) stated that, “if a consonant lenites in some context, at a given rate or register of speech, it also lenites in that context at all faster rates or more casual registers of speech” (pp. 217–218). Cohen Priva and Gleason (2020) argued that reduced duration is the cause of lenition processes, at least for American English.

Place of articulation also affects lenition. The hierarchy of places of articulation ordered by their likelihood of undergoing lenition is velar > bilabial > alveolar (Foley, 1977). However, evidence in support of this hierarchy has been mixed. For example, with respect to closure duration, intervocalic voiceless dentals are more lenited than labials and velars, but voice-onset-time measures indicate that voiceless labials weaken more than dentals and velars (Lewis, 2001). In contrast, intervocalic voiceless labials are the most prone to voicing during stop closure (i.e., become lenited), whereas intervocalic velars demonstrate the greatest degree of resistance to closure voicing (Lewis, 2001). For voiced stops, Kingston (2008) found velars to be more likely to lenite than dentals and bilabials, whereas Colantoni and Marinescu (2010) found dentals to be more likely to lenite than labials and velars.

Quality of flanking vowels is another variable investigated for its influence on lenition. The articulatory-effort view of lenition predicts more lenition when the target consonants are surrounded by more open vowels since the distance that the articulators must travel to achieve a high degree of constriction from a more open vowel is greater than from a close vowel. However, no asymmetry in vowel openness on lenition is expected according to the perceptual view of lenition since, unlike consonants, difference in openness among vowels is negligible (Kingston, 2008). Empirical evidence available points to a complex interaction between vowel openness and place of articulation in the lenition process. For example, Spanish /g/ was less lenited between low vowels than between high vowels (Cole *et al.*, 1999; Ortega-Llebaria, 2004), while no effects of vowel height were found for /b/ (Ortega-Llebaria, 2003, 2004). Simonet *et al.* (2012) showed that /d/ is more lenited after lower vowels than after high vowels. In contrast, Kingston (2008) reported a higher degree of lenition of Spanish voiced stops next to higher as opposed to lower vowels. However, the account by Kingston (2008) does not expect vowel height to have an effect, whereas the effort-based account of Kirchner (2004) does.

Position of the target segment in a prosodic domain is another factor relevant to lenition. Escure (1977) observed that initial lenition is dispreferred at the syllable, word, and

utterance level. Evidence in support of dispreference for domain-initial lenition is largely borne out cross-linguistically (compare Ségéral and Scheer, 2008). Kingston (2008) reported that lenition of Andean Spanish voiced stops is categorically more likely inside a prosodic constituent than at its edge.

Finally, stress is a known conditioning factor of lenition. Cole *et al.* (1999) reported a higher degree of lenition when Castilian Spanish /g/ occurs after a stressed syllable compared to an unstressed one. Similar results were reported for Caribbean Spanish /b/ and /g/ in Ortega-Llebaria (2004). Carrasco and Hualde (2009) observed a higher degree of constriction when the voiced stops occur before a stressed syllable. Eddington (2011) found that a following unstressed syllable promotes lenition in telephone conversations of eight native Spanish speakers from seven countries. We expect a consonant in an unstressed syllable will be more lenited than one in a stressed syllable.

B. Acoustic metrics of lenition

Several acoustic metrics have been used to measure lenition, but the view that lenition impacts the intensity properties of the target consonants largely prevails. For example, intensity difference (preceding segments maximum intensity minus minimum intensity of the target consonant) was used to quantify degree of lenition by Martínez Celdrán and Rgueira (2008); Figueroa and Evans (2015), and Broś *et al.* (2021). On the other hand, Hualde *et al.* (2012) employed the difference between the maximum intensity value during the vowel following the target consonant and the minimum value during the target consonant portion as an acoustic reflex of degree of lenition. The smaller the difference, the more open the constriction of the target consonant (i.e., the more lenited the target consonant). Ortega-Llebaria (2003) computed the root mean square intensity ratio between the intervocalic target segment and the VCV portion of a CVCV word (in which the second C is the target segment): a ratio closer to 1 indicates a more vowel-like (i.e., higher degree of lenition), while a ratio closer to 0 indicates a more stop-like (i.e., lesser degree of lenition) pronunciation. As a second measure of lenition, Ortega-Llebaria (2003) quantified the speed with which the intervocalic target consonant is released into the following vowel by taking the difference between the maximum and the minimum intensity of the consonants as a function of time: the faster the release, the more stop-like the production, thus a lesser degree of lenition. Hualde *et al.* (2012) used maximum rising velocity from the midpoint of the target consonant to the midpoint of the following vowel as a measure of lenition. The more lenited the target consonant is, the less abrupt the transition in intensity is expected to be and, thus, the smaller the maximum rising velocity value. Similarly, a change in intensity velocity from the preceding vowel to the intervocalic target segment as well as from the target segment to the following vowel were used by Kingston (2008) to quantify lenition. Most recently, Harris *et al.* (2024) presented a new metric,

“Edge.”² Edge measures the fluctuation of acoustic energy across a VCV frame. Edge is a standard-deviation metric which differs from methods which depend on the consecutive order of values (t minus $t-1$) for computing C-to-V velocity, such as that by [Kingston \(2008\)](#) and [Hualde et al. \(2012\)](#). A high Edge value indicates a drop in energy during the target consonant.

C. Phonet

In contrast to the quantitative acoustic method, Phonet ([Vásquez-Correa et al., 2018, 2019](#)) can serve as a deep learning approach to lenition. As a bi-directional recurrent neural network model, Phonet is trained to recognize input phones as belonging to different phonological classes defined by phonological features (e.g., sonorant, continuant, anterior, strident) based on log-energy distributed across triangular Mel filters³ computed from 25-ms windowed frames of each 0.5 s chunk of the input signal (see [Vásquez-Correa et al., 2019](#), for details). Once trained, posterior probabilities for different phonological features of the target segments can be computed by the model. Posterior probabilities are the probabilities of phonological features inferred from the speech signal.⁴ In this study, we focus on the probability of the phonological features [continuant] and [sonorant] to capture the two categorical realizations (fricatives and approximants) of stop lenition in Spanish. A fricative-like realization would be associated with a relatively high [continuant] but a low [sonorant] probability, while a relatively high [continuant] and [sonorant] probability would correspond to an approximant-like production. In addition to being able to capture both categorical and gradient natures of lenition, Phonet can be customized to specific languages with different sets of phonological features and acoustic representations. It is semi-automatic and only requires a segmentally aligned acoustic corpus, which can be obtained using forced alignment. This approach utilizes the acoustics of segments that share acoustic properties with lenited and non-lenited segments. It is motivated by a long line of research that seeks to examine gradient variations by using surface segments that do not have to be realized from the underlying segments that are subjected to the variation of interest. The approaches used in this line of research differ mainly in their model architectures, using forced alignment models ([Bailey, 2016](#); [Kendall et al., 2021](#); [Magloughlin, 2018](#); [Pandey et al., 2020](#); [Yuan and Liberman, 2009, 2011](#)), classification models with machine-learning methods ([McLarty et al., 2019](#); [Villarreal et al., 2020](#)), or linear regressions ([Cohen Priva and Gleason, 2020](#)). A complete account of the motivations behind this approach can be found in [Tang et al. \(2023\)](#).

D. Consistency

Previous studies by [Wayland et al. \(2023a, 2023b\)](#) evaluated deep-learning lenition metrics derived from Phonet against acoustic metrics of three broad acoustic dimensions of lenition (duration, intensity, and periodicity) with their

ability to detect known effects of lenition conditioning factors. The acoustic metrics include harmonic-to-noise ratio [measure of the proportion of acoustic periodicity (harmonics) to aperiodicity (noise) of a given sound], relative duration (the relative duration of the target consonant and the total duration of the preceding sound + target sound + following sound), intensity difference (preceding segment’s maximum intensity – minimum intensity of the target consonant), and mean intensity of the target sound. Focusing on intervocalic voiced and voiceless stops, /b, d, g, p, t, k/ in Spanish, the studies found that, except for the mean intensity metric, expected lenition patterns predicted by known lenition factors are more consistently revealed by the deep-learning metrics than by most acoustic metrics. Among the acoustic metrics, mean intensity measure is the most consistent and in the expected direction, while harmonic-to-noise ratio is the least consistent and largely in the unexpected direction.

However, it remains unclear how intensity velocity would perform against the deep-learning metrics. In particular, the current study focuses on the implementation of intensity velocity that takes into account different frequency bands (see Sec. [IE](#) for details of the bands), for instance, [Kingston \(2008\)](#) and [Ennever et al. \(2017\)](#). The proposed approach by [Kingston \(2008\)](#) is especially noteworthy, since it was independently proposed by [Harris and Urua \(2001\)](#), and it was recently built upon by [Ennever et al. \(2017\)](#). [Harris and Urua \(2001\)](#) and [Ennever et al. \(2017\)](#) employed a similar implementation of intensity velocity, not to examine lenition patterns of Spanish, but of two different languages (Ibibio and Gurindji). Furthermore, our literature review suggests that existing work on lenition tends to only focus on one type of intensity-based measure. This runs the risk of failing to determine the presence, robustness, and nature of lenition in a given language. Finally, while the Phonet measures are trained on acoustic representations that are related to intensity (log-energy), it does not take into account the acoustic transition from flanking segments into the target consonant, unlike the intensity velocity measures. Therefore, this research gap calls for a comprehensive comparison of the two approaches using the same language.

E. Acoustic representations

As discussed in Sec. [IB](#), many metrics of lenition focus on its effect on the intensity properties of the target consonants. While the exact methods of computing the intensity difference vary, they largely operationalize intensity over the entire frequency spectrum, with the exception of [Kingston \(2008\)](#). Measures of intensity over the entire frequency spectrum might not be able to distinguish different lenition outcomes. Therefore, focusing on specific frequency bands enables researchers to examine lenition outcomes that involve different distinctive features of speech sounds. This approach builds on the effort of identifying acoustic landmarks ([Stevens, 1981, 2002](#)), since different landmarks are known to be found in different bands. The [Kingston \(2008\)](#)

study adopted the six frequency bands used in Liu (1996): band 1 (0.0–0.4 kHz), band 2 (0.8–1.5), band 3 (1.2–2.0), band 4 (2.0–3.5), band 5 (3.5–5.0), and band 6 (5.0–8.0). The six bands were designed to capture the distinctive features of consonants. Band 1 is designed to monitor the presence or absence of glottal vibration. Bands 2–5 correspond to the frequency ranges for the spectral prominences of sonorant consonants, in order to capture closures and releases for sonorant consonants. Specifically, bands 2 and 3 span the range of 0.8 to 2 kHz to capture the large spectral change that occurs with intervocalic sonorant consonantal segments. Bands 2–5 can also capture noise energy, such as the onsets and offsets of aspiration and frication noise associated with stops, fricatives, and affricates. Band 6 is primarily used for silence detection for stops.

II. METHODS

A. Materials

A total of 8893 tokens of voiced and voiceless stops from an Argentine Spanish Corpus were included in the study. It was built by Guevara-Rukoz *et al.* (2020) and contains crowd-sourced recordings from 31 female and 13 male native speakers of Argentine Spanish. The male sub-corpus contains 2.4 h of recording with 16 914 words (3342 unique words), while the female sub-corpus contains 5.6 h of recording with 35 360 words (4107 unique words).⁵ Following Kingston (2008), word tokens with /b, d, g, p, t, k/ as word-initial segments followed by vowels of different heights and preceded by a vowel or a nasal from the preceding word were selected. Table I specifies the number of word tokens and word types by conditions: voicing (voiced or voiceless), place of articulation (bilabial, dental, or velar), previous phone (vowel or nasal), and following vowel (open, mid or close).

B. Phonet training procedures

The forced alignment process was performed on the corpus using the Montreal Forced Aligner (version 2.0) (McAuliffe *et al.*, 2017). A phonemic pronunciation dictionary for the transcription of the corpus words generated based

on the Hualde (2013) grapheme-to-phoneme mapping in the International Phonetic Alphabet (IPA) was then used to train a new triphone acoustic model and align the text grids to the acoustic signals. The phone set parameter was set to IPA to allow extra decision tree modeling based on the specified phone set. Any other parameters were kept as the default.

Model training was performed on a NVIDIA GeForce RTX 3090 GPU. The corpus was randomly split into a train subset (80%) and a test subset (20%) using the Python (version 3.9) scikit-learn library (Pedregosa *et al.*, 2011). To avoid model contamination by ambiguous tokens, the targets /b, d, g/, but not /p, t, k/ (Colantoni and Marinescu, 2010), were excluded during training (i.e., silenced out) since they are expected to be ambiguous with respect to their realizations of the continuant and sonorant features. A total of 23 phonological classes, namely, syllabic, consonantal, sonorant, continuant, nasal, trill, flap, coronal, anterior, strident, lateral, dental, dorsal, diphthong, stress, voice, labial, round, close, open, front, back, and pause, were trained by 20 different Phonet models. Following Vásquez-Correa *et al.* (2019), one additional model was included to train phonemes. However, in addition to the 18 phonemes of Vásquez-Correa *et al.* (2019), 8 phonemes including stressed /'a, 'e, 'i, 'o, 'u/, /ɲ/, /θ/ and /spn/ for speech-like noise were added. The complete feature chart can be found in Tang *et al.* (2023) for all phonemes with respect to the segmental inventory of Argentine Spanish and their corresponding graphemes along with their phonological feature values.

The model was highly accurate in detecting the different phonological classes, showing unweighted average recall ranges from 94% to 98%. Critically, the unweighted average recalls for the sonorant and continuant features are 97% and 96%, respectively. The model was then applied to our selected word tokens with /b, d, g, p, t, k/. The predictions were computed for 10-ms frames. For a token containing multiple frames, the average prediction from the middle third portion of all frames of a segment was used as its prediction. A sonorant posterior probability and a continuant posterior probability obtained for each target stop were then used for statistical analyses.⁶

C. Acoustic parameter measures

Following Kingston (2008), three acoustic parameters (minimum intensity velocity, maximum intensity velocity, and duration) were extracted from each target segment using a modified version of the DiCanio *et al.* (2020) Praat script using the following procedure. First, the signal was bandpass-filtered into six frequency bands: 0–400, 800–1500, 1200–2000, 2000–3500, 3500–5000, and 5000–8000 Hz, corresponding to those used by Liu (1996) in her study aiming at finding the acoustic landmarks for distinctive features of consonants. Second, the intensity extracted from each band was first differenced to exaggerate the magnitude of its change. Third, the minimum and the maximum values and time between the two were extracted.

TABLE I. Word distribution by conditions: stress, voicing, place of articulation, previous phone, and following vowel. The numbers left and right of the slash in each cell represent the number of word tokens and word types, respectively.

Place	Previous phone	Voiced			Voiceless		
		Following vowel height					
		Close	Mid	Open	Close	Mid	Open
Bilabial	Vowel	281/40	309/46	253/32	323/37	842/71	529/45
	Nasal	72/12	48/14	44/8	22/5	133/16	80/9
Dental	Vowel	195/39	816/55	54/9	321/22	711/55	274/16
	Nasal	39/9	333/10	6/2	71/4	102/9	15/4
Velar	Vowel	0/0	33/1	0/0	290/36	1683/95	577/71
	Nasal	0/0	0/0	0/0	69/9	259/14	109/16

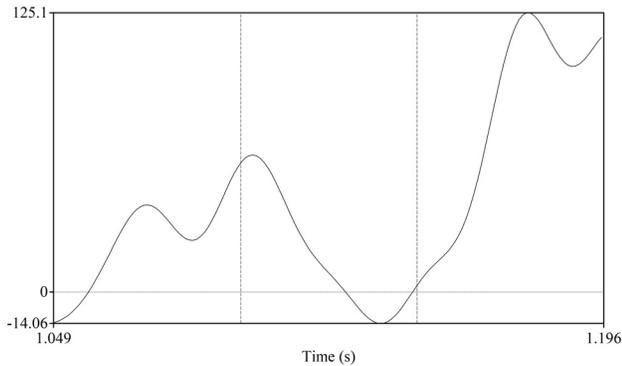


FIG. 1. The first-differenced intensity waveform for the interval filtered by the frequency band 0–400 Hz. This constriction interval includes the initial consonant in *vos* from *como vos*, which is the interval in between the vertical dashed lines, with ± 50 ms on each side.

As shown in Fig. 1, the minimum and the maximum values are located within an interval of ± 50 ms of the beginning and the end of the constriction, respectively. The vertical dashed lines are the sound edges which are generated by forced alignment. Specifically, for each frequency band, the minimum value is the lowest value within ± 50 ms of the left vertical dashed line, and the maximum value is the highest value within ± 50 ms of the right vertical dashed line. Duration is defined as the temporal distance between the minimum and maximum.⁷ These three acoustic values were then entered into the statistical analyses. For a detailed illustration of the intensity profiles extracted from six different bands, see Kingston (2008, p. 20).

D. Analyses

The three acoustic dimensions from six frequency bands and the sonorant and continuant posterior probabilities calculated by the Phonet model served as dependent variables in the linear mixed-effects regression models. Table II shows the descriptive statistics of the acoustic values, specifically, the first and third quartiles (25th and 75th) for /p t k b d g/ separately. /p t k/ have lower sonorant and continuant posterior probabilities, more negative minima, more positive maxima, and shorter duration than /b d g/.⁸ Five fixed factors — stress (stressed or unstressed), voicing (voiced or voiceless), place of articulation (bilabial, dental, or velar), previous phone (vowel or nasal), and following vowel (open, mid, or close) — were included in the models.⁹ Deviation coding was used for the variables stress, voicing, and previous phone, while forward difference coding was used for the variables place of articulation (bilabial > dental > velar) and following vowel (close > mid > open). The models were performed using the `lmer` function from the `lme4` package (Bates et al., 2015). After comparing multiple model structures with maximum likelihood, the best-fit model structure with difference interaction terms for each variable was identified. We then added by-speaker random slopes to the best-fit model structure for the variables place of articulation, stress, and voicing. These three slopes were added to capture potential individual speaker differences. The effect of stress, place of articulation, and voicing may be different for one person compared to another. This is particularly motivated for the variable voicing, since the voicing of Spanish /p t k/ is not a systematic process, and may be subject to coarticulatory effects as

TABLE II. Descriptive statistics of the acoustic measures. Each cell contains the first and third quartiles (25th and 75th) of the measurements as a range for each of the six stops (columns 2–7). The first column refers to each of the acoustic measures. Min, Max, and Dur denote minima, maxima, and duration, respectively; their following digit (1–6) refers to each of the six frequency bands. Min and Max are in dB/s; Dur is in seconds; Sonorant and Continuant are probabilities (0–1).

	/p/	/t/	/k/	/b/	/d/	/g/
Sonorant	0.082–0.570	0.031–0.254	0.084–0.713	0.959–0.995	0.931–0.987	0.867–0.973
Continuant	0.036–0.309	0.019–0.152	0.073–0.551	0.566–0.958	0.745–0.986	0.918–0.988
Min1	–781 to –545	–773 to –555	–786 to –560	–180 to –72	–173 to –68	–131 to –45
Min2	–1123 to –882	–1068 to –826	–1023 to –773	–439 to –150	–482 to –201	–360 to –201
Min3	–1229 to –972	–1198 to –951	–1113 to –838	–556 to –232	–487 to –204	–339 to –147
Min4	–1088 to –851	–1133 to –900	–1055 to –809	–562 to –267	–458 to –165	–243 to –118
Min5	–1121 to –828	–1172 to –886	–1117 to –823	–591 to –297	–596 to –214	–277 to –178
Min6	–924 to –584	–1040 to –675	–963 to –595	–578 to –316	–559 to –207	–273 to –162
Max1	864–1288	945–1318	851–1255	58–143	50–146	21–62
Max2	1159–1831	1040–1652	991–1633	137–485	52–501	131–250
Max3	1186–1839	1107–1778	1023–1739	190–691	89–597	129–235
Max4	1124–1709	1218–1804	929–1522	289–652	179–753	100–164
Max5	1038–1556	1310–1861	1024–1612	301–680	191–887	106–228
Max6	776–1213	1322–1804	842–1429	334–659	203–816	193–304
Dur1	0.027–0.057	0.026–0.057	0.034–0.059	0.032–0.097	0.028–0.087	0.053–0.141
Dur2	0.039–0.063	0.035–0.058	0.021–0.053	0.031–0.080	0.036–0.079	0.080–0.111
Dur3	0.040–0.063	0.036–0.057	0.021–0.054	0.032–0.079	0.029–0.077	0.066–0.116
Dur4	0.042–0.069	0.036–0.056	0.035–0.064	0.031–0.076	0.026–0.073	0.047–0.088
Dur5	0.045–0.073	0.036–0.058	0.031–0.057	0.036–0.078	0.033–0.074	0.051–0.104
Dur6	0.052–0.083	0.035–0.060	0.034–0.063	0.045–0.091	0.037–0.083	0.048–0.090

opposed to /b, d, g/ lenition, which is a stable allophonic process.¹⁰ The general structure of the models with three interaction terms¹¹ is as follows:

- [Dependent variables]
- ~ Stress + Voicing + Place of articulation
 - + Following vowel + Previous phone
 - + Place of articulation : Following vowel
 - + Place of articulation : Previous phone
 - + Following vowel : Previous phone
 - + (1 + Stress + Voicing
 - + Place of articulation |Speaker) + (1| Word)

Post hoc comparisons of the interaction terms were carried out using emmeans with Tukey’s honestly significant difference for *p*-value adjustment (Lenth *et al.*, 2021).

According to previous literature, lenition is expected to be more likely for voiced than voiceless stops and in an unstressed relative to a stressed syllable, thus showing less extreme minima, maxima, and duration values, and higher continuant and sonorant posterior probabilities. It is hypothesized that degree of lenition would follow the velar > bilabial > dental hierarchy (Foley, 1977). In addition, a more open vowel would promote lenition relative to

a less open vowel. Finally, lenition is predicted to be more likely after a vowel than after a nasal because of difficulty in precise articulatory coordination between a velum raising and an oral cavity opening gesture (Kingston, 2008; Ohala, 1981; Steriade, 1993).

III. RESULTS

To facilitate comparison with the results of Kingston (2008), only the main effects of the models will be reported. Marginal R^2 values for the six frequency bands range from 0.329-0.725 (mean = 0.6185) for minima, 0.392-0.748 for maxima (mean = 0.570), and 0.076-0.174 (mean = 0.174) for duration. For all three dependent variables, the R^2 values decrease as the frequency band increases. These results suggest that better fits to the data were obtained for minima and maxima relative to duration and for a lower frequency band relative to a higher frequency band. In addition, a better model fit was obtained for the sonorant than the continuant probability (marginal $R^2 = 0.568$ vs 0.533). Figures 2–4 show β values of the five independent variables, namely, stress, voicing, place, following vowel height and previous phone, on the minima, the maxima, and the duration for each of the six frequency bands. Significant impact of the variable on each of the three acoustic parameters measured

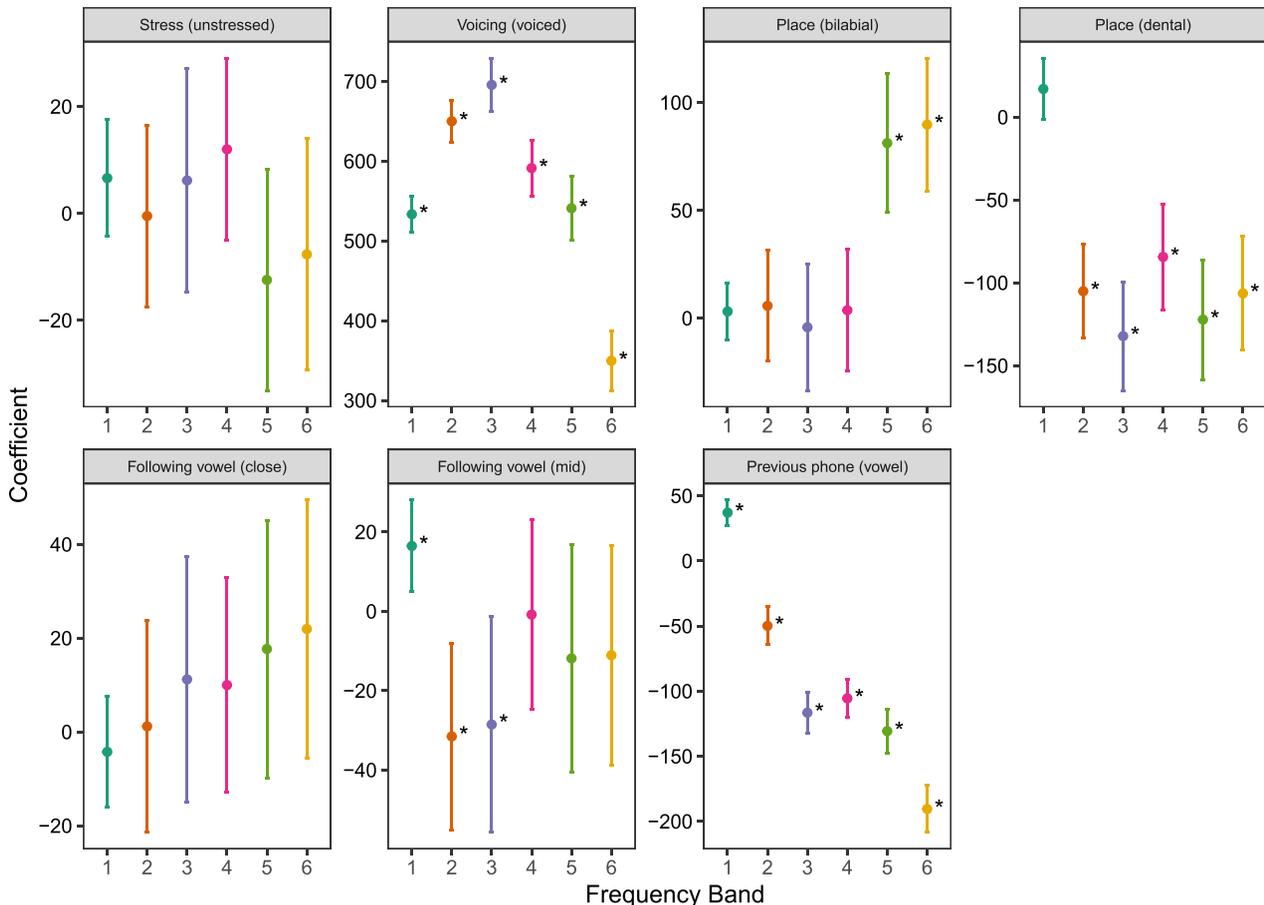


FIG. 2. (Color online) Effects of stress, voicing, place, and vowel on intensity minima at 6 frequency bands (1: 0–400; 2: 800–1500; 3: 1200–2000; 4: 2000–3500; 5: 3500–5000; and 6: 5000–8000 Hz). Positive coefficients indicate more lenition. **p* < 0.05.

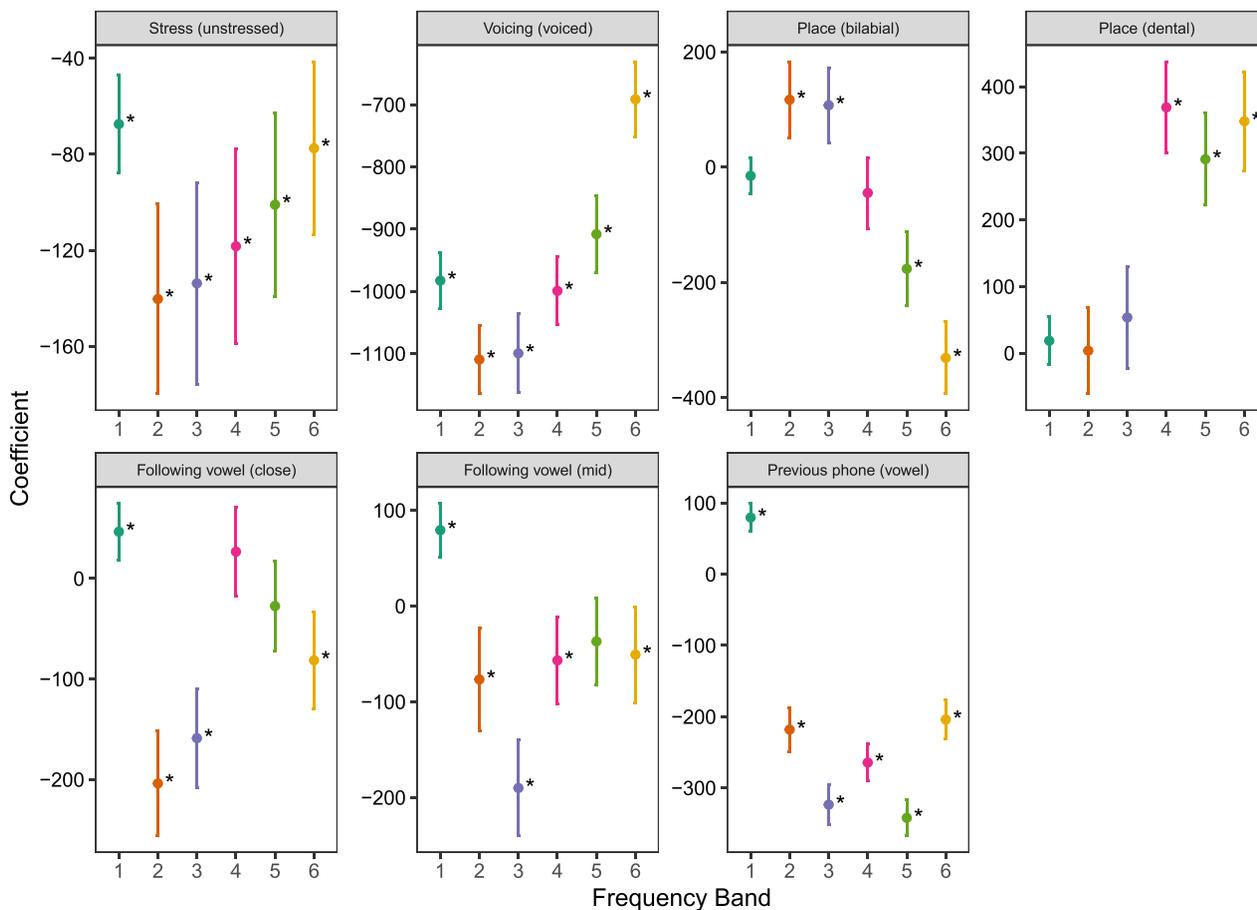


FIG. 3. (Color online) Effects of stress, voicing, place, and vowel on intensity maxima at 6 frequency bands (1: 0–400; 2: 800–1500; 3: 1200–2000; 4: 2000–3500; 5: 3500–5000; and 6: 5000–8000 Hz). Negative coefficients indicate more lenition. * $p < 0.05$.

at different frequency bands is indicated by an asterisk (*). Negative β values indicate more negative minima (i.e., more stop-like closure or less lenited) but lower maxima and shorter duration (i.e., more vowel-like closure or more lenited).

A. Minima

Voicing and preceding segment have significant effects on minima in all frequency bands, while no effect was observed for stress (Fig. 2). For voicing, a voiced stop exhibits less negative minima (positive β values), indicating that it is more lenited than a voiceless stop. Except for the first frequency band, a preceding vowel exaggerates the minima (negative β values), suggesting that less lenition is predicted when a stop consonant occurs after a vowel than after a nasal. The effects of place of articulation and openness of the following vowel are inconsistent. A bilabial stop is predicted to be more lenited than a dental stop in the two highest-frequency bands, while a dental stop is predicted to be less lenited than a velar stop in all frequency bands, except the first and the third bands. For following vowel height, more lenition is predicted when a stop occurs before a mid vowel relative to an open vowel for the first frequency

band, while the opposite is true for the second frequency band.

B. Maxima

A rather different pattern of results is obtained for the maxima (Fig. 3). Stress, voicing, and previous phone show significant and largely consistent effects. Specifically, lenition is predicted to be more advanced (negative β values) in an unstressed relative to a stressed syllable and for a voiced stop compared to a voiceless stop. Similarly, except for the first frequency band, a stop is predicted to be more lenited after a vowel than after a nasal. The effects of place and following vowel height are mixed. A bilabial is predicted to be less lenited than a dental in the second and third frequency bands, but the opposite is true for the fourth and fifth frequency bands. However, as expected, a dental is predicted to be less lenited than a velar at all frequency bands except the second. For the effect of following vowel height, more lenition is predicted when a stop precedes a close vowel relative to a mid vowel in the second, third, and sixth frequency bands, but the opposite is true for the first band. In addition, more lenition is predicted before a mid vowel relative to an open vowel in the second and fourth frequency bands, but the opposite is predicted for the first band.

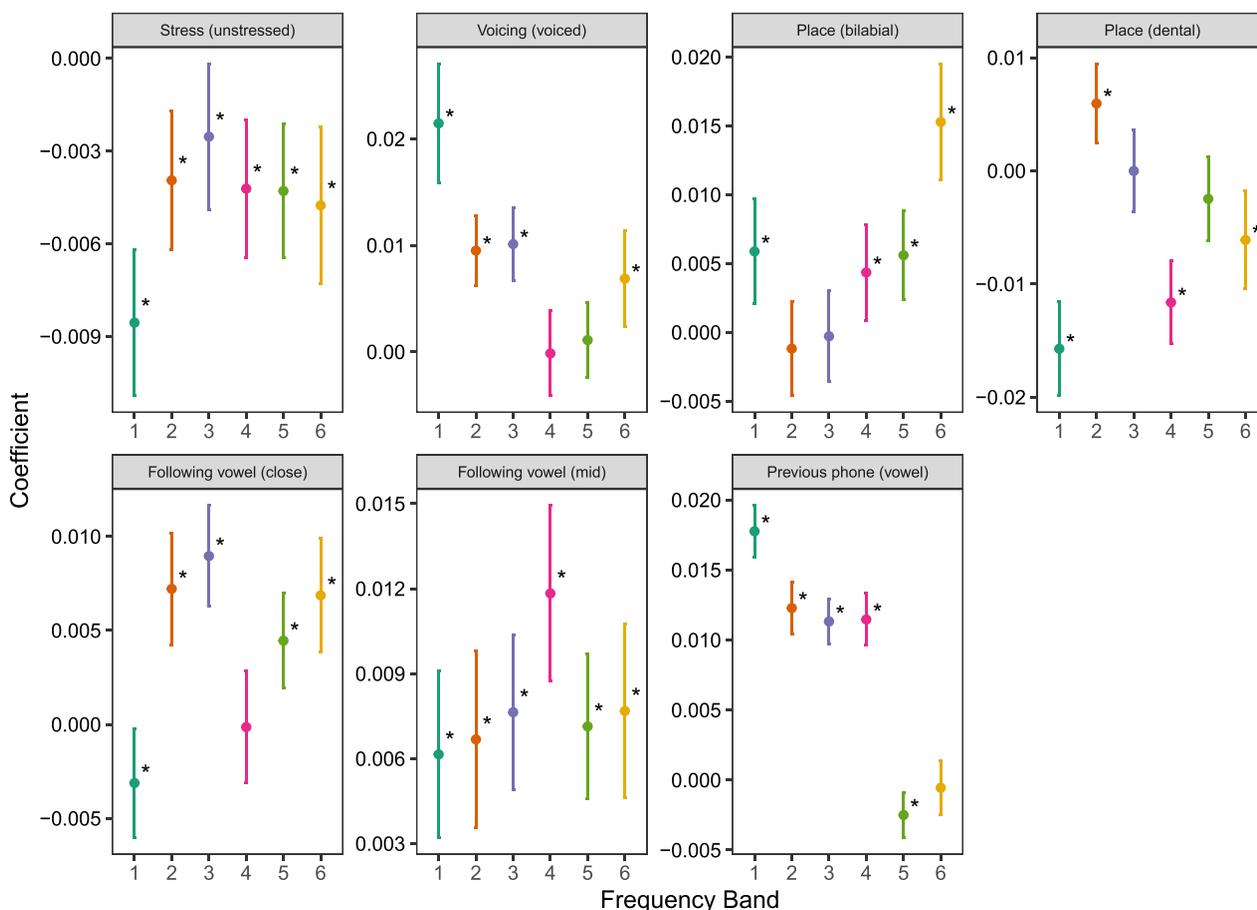


FIG. 4. (Color online) Effects of stress, voicing, place, and vowel on intensity duration at 6 frequency bands (1: 0–400; 2: 800–1500; 3: 1200–2000; 4: 2000–3500; 5: 3500–5000; and 6: 5000–8000 Hz). Negative coefficients indicate more lenition. * $p < 0.05$.

C. Duration

For duration (Fig. 4), the effects of stress and following vowel height are the most consistent. A more advanced degree of lenition (negative β values) is predicted for a stop in an unstressed relative to a stressed syllable in all frequency bands. However, less lenition (positive β values) is predicted when a stop is preceded by a close relative to a mid vowel in all but the fourth frequency band and when it is preceded by a mid vowel compared to an open vowel at all frequency bands. For place of articulation, a bilabial is predicted to be less lenited than a dental at all frequency bands, except the second and the third bands, while a dental is predicted to be more lenited than a velar in the first and the fourth bands, but the opposite is true for the second and the sixth bands. For previous phone, lenition is predicted to be less advanced after a vowel than after a nasal in the first four frequency bands, while the opposite is true for the fifth band.

D. Posterior probability

Stress, voicing, and previous phone exert significant and consistent effects on sonorant and continuant posterior probabilities (Fig. 5). More lenition (positive β values) is predicted in an unstressed relative to a stressed syllable, for

a voiced relative to a voiceless stop and when a stop occurs after a vowel rather than a consonant. The effects of place and following vowel height are less consistent. For the sonorant posterior probability, a bilabial is predicted to be more lenited than a dental while a dental is predicted to be less lenited than a velar. For continuant posterior probability, only a dental is predicted to be less lenited than a velar. For both sonorant and continuant posterior probabilities, less lenition is predicted when a stop occurs before a close vowel than a mid vowel, while no difference in the effect of a mid vowel relative to an open vowel on degree of lenition is predicted.

To facilitate comparisons of the effects of the fixed factors on the five dependent variables against the predictions based on previous literature, all main effects of the regression models are summarized in Table III.

Several generalizations can be made from these results. First, the effect of stress is in the expected direction (i.e., higher degree of lenition in unstressed than in stressed syllables) for maxima and duration in all six frequency bands as well as for the sonorant and continuant posterior probabilities. Second, among the three acoustic parameters, the effect of voicing is consistently in the expected direction for minima and maxima but not for duration. In contrast, this prediction is borne out for both the sonorant and the continuant

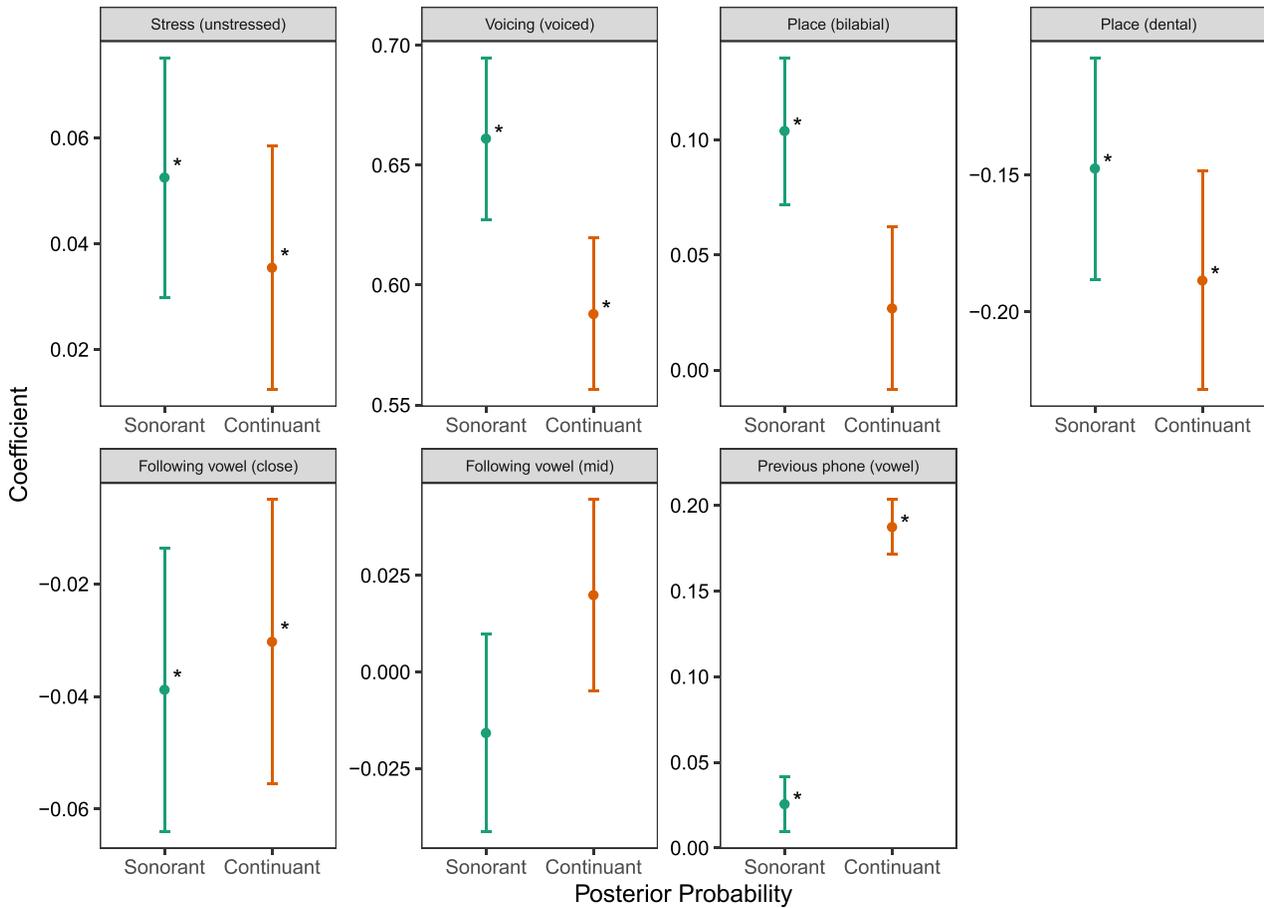


FIG. 5. (Color online) Effects of stress, voicing, place, and vowel on sonorant and continuant posterior probabilities. The x axis denotes the posterior probability of sonorant (left) and continuant phonological features. Positive coefficients indicate more lenition. * $p < 0.05$.

27 August 2024 12:14:11

TABLE III. Summary of the main effects of the regression models. The presence of either a positive sign (+) or a negative sign (-) indicates significant main effects of the fixed factors on the dependent variables, while nonsignificant effects are left blank. A positive sign (+) indicates the effects are in the predicted direction, while a negative sign (-) indicates the unpredicted direction. For consonant place: B, bilabial; D, dental; V, velar. For vowels: C, close; M, mid; O, open. For previous phone: V, vowel; N, nasal.

	Minima (frequency bands)						Maxima (frequency bands)						Duration (frequency bands)						
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
Stress							+	+	+	+	+	+	+	+	+	+	+	+	
Voicing	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-			-	
Place: B > D					+	+			-	-		+	+				-	-	-
D < V		+	+	+	+	+					+	+	+	-	+		-		+
Vowel height: C < M							+	-	-			+	-	+	+		+	+	
M < O	-	+	+				+	-	-	-			+	+	+	+	+	+	
Previous phone: V > N	+	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-	+		
							Sonorant posteriors						Continuant posteriors						
Stress					+													+	
Voicing					+													+	
Place: B > D					+														
D > V					+													+	
Vowel height: C vs M					+													+	
M vs O					+													+	
Previous phone: V > N					+													+	

probabilities. Third, the effect of the preceding phone is in the predicted direction (i.e., a higher degree of lenition after a vowel than after a nasal) only in the lowest frequency band for minima, the fourth and fifth bands for maxima and the fifth band for duration. In contrast, the prediction is borne out on both the sonorant and continuant posterior probabilities. Fourth, the effect of the following vowel height is largely in the expected direction for duration but not for minima and maxima. For sonorant and continuant posteriors, as expected, more lenition is predicted for a stop when it occurs after a mid vowel relative to a close vowel. The difference between a mid and an open vowel context was not predicted to be significant, however. Finally, for place of articulation, the prediction that a bilabial is more lenited than a dental is borne out for minima and maxima in the two highest-frequency bands and for sonorant but not for continuant probability. However, the prediction that a dental is less lenited than a velar is confirmed by both sonorant and continuant probabilities as well as by minima in most frequency bands.

IV. CONCLUSIONS

Lenition is a gradient phenomenon affecting pronunciations of different target consonants differently in different contexts. Prior literature on lenition suggests that voicing, an unstressed syllable, and a more open flanking segment (vowel or consonant) promote lenition. For place of articulation, the prediction has been mixed. While more lenition was hypothesized for a more posterior stop by Kingston (2008), a hierarchy of velar > bilabial > dental was proposed by Foley (1977). Following Foley (1977)'s implication hierarchy, we hypothesized that degree of lenition would be more advanced for a velar stop than for a bilabial and a dental stop. In turn, a bilabial stop will weaken to a greater degree than a dental stop.

The results obtained indicate that the effects of known lenition factors on the three acoustic metrics are mixed. For example, consistent and predicted effects of stress are observed for maxima and duration but not for minima. On the other hand, the effect of voicing is consistently realized as predicted on minima and maxima but not on duration. Furthermore, the prediction that lenition is more advanced following a vowel than a nasal is borne out only for maxima and only in some frequency bands. Similarly, the effects of place of articulation and height of a following vowel are inconsistent across the three acoustic metrics. Finally, the prediction that lenition is greater after a nasal than after a vowel by minima values in most frequency bands is totally unexpected.

In comparison, sonorant and continuant probability values calculated by the Phonet model are more consistent with the predicted direction of lenition than the three acoustic metrics. Specifically, sonorant and continuant probability values are in the direction predicted based on voicing, stress, and preceding segment type. The values of these two deep learning lenition metrics are also largely

consistent with the hierarchy of Foley (1977) with respect to the effect of place of articulation on lenition. Moreover, the effect of following vowel height on the two metrics is also in the direction predicted for a close vs a mid vowel context. Although not statistically significant, the effect of a mid vowel vs an open vowel is also in the predicted direction. These results appear to be consistent with the view of Kirchner (2004) of lenition. In addition, these results strongly suggest that Phonet is a reliable approach to lenition measurement. Our work can be extended in other directions by examining other known lenition measures, such as acoustic periodicity (harmonics) and aperiodicity (noise) (Broś *et al.*, 2021; Harris and Urua, 2001; Harris *et al.*, 2024; Wayland *et al.*, 2023a). Our findings suggest that the choice of lenition measures can influence researchers' ability to detect and examine lenition patterns.

Methodologically, the deep learning lenition metrics do not suffer from data loss which the approach of Kingston (2008) does. As noted in the Kingston (2008) study, around 9% of data tokens had to be discarded because the extracted minimum and maximum intensity-change values were temporally misordered (Kingston, 2008, p.20). While this can be improved by smoothing the time-varying profile of intensity and thresholding the intensity values [e.g., Liu (1996) and Ennever *et al.* (2017)], these steps introduce additional free parameters which increase the already high researcher degrees of freedom in phonetic research (Coretta *et al.*, 2023; Roettger, 2019).

The inconsistencies we found across the different frequency bands raise an important question. What does it mean when an effect of lenition is found within one band but not another? In principle, the different frequency bands are supposed to target distinct acoustic landmarks; therefore, one can interpret which acoustic cues are influenced by lenition or not. Furthermore, certain bands might be better suited for examining certain target sounds, depending on their acoustic landmarks. For instance, band 1 is designed to monitor the presence or absence of glottal vibration; therefore, it can tell us whether voicing is being affected and would be better suited for examining underlying voiceless consonants. However, the acoustic landmarks that each of the bands targets in fact overlap quite a lot. Bands 2–5 capture spectral prominences and closures and releases of sonorant consonants, as well as aspiration and frication noise associated with stops, fricatives, and affricates, covering four manners of articulation. Band 6 is associated with stops, but so are bands 2–5. Therefore, the lack of an effect within specific bands cannot easily indicate the nature of the lenited segments. This perhaps is why, in Kingston (2008), the inconsistencies found across bands were not discussed: "Each of these variables significantly affected these measures in a majority if not all of the frequency bands" (Kingston, 2008, p. 22). It is worth noting that the bands of Liu (1996) were based on English, and it is unclear if they are valid for all languages. In a lenition study of Ibibio stops, Harris *et al.* (2024) compared the bands of Liu (1996) with an alternative banding (100–2000, 1500–3500,

3000–5000, 100–5000 Hz) used in [Harris and Urua \(2001\)](#) and found that the two bandings are highly correlated. In another study on the lenition of Gurindji stops, [Ennever et al. \(2017\)](#) found that small changes to the precise band settings generally had little impact on their results. Together, these studies suggest that the choice between different frequency-band settings is unlikely to make much of a quantitative difference. While the results obtained from different frequency bands are not as consistent as the deep learning lenition metrics, it would be too hasty to conclude that there are no merits to frequency band divisions, especially because a systematic examination of the choice of different bands for computing intensity velocity has not been widely studied. As far as we know, only two studies ([Ennever et al., 2017](#); [Harris et al., 2024](#)) thus far have experimented with more than one set of bands.

Future work can modify the acoustic representation used by the deep learning model to take into account the different frequency bands and the acoustic landmarks they target. In addition, similar to [Ennever et al. \(2017\)](#), segmentation of the target lenited stops could be based on landmarks present at different frequency bands, allowing for a more nuanced measurement of lenition targeting different realizations of lenited variants.

As an additional contribution to our empirical findings, we have made a number of resources available as supplementary materials. First, we released the Phonet model trained on Spanish speech as used in this study, a detailed training instruction document which contains step-by-step instructions and the entire training pipeline. Second, we have preliminarily released a toolkit which is in-development, called Lenition Integrated Python System (LIPS) (<https://github.com/oliviadinicola/LIPS>). LIPS contains a Graphical User Interface which allows users to easily compute [Kingston \(2008\)](#)'s measures, harmonic to noise ratio ([Broś et al., 2021](#)), and the posterior probability values from a trained Phonet model. We hope our effort in releasing these resources would encourage future researchers to further evaluate and build on Phonet and compare with other lenition measures.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation (SenSE) Award No. 2037266. We thank Nele Marie Mastracchio for assistance in typesetting the manuscript; and Rachel Meyer for proofreading the manuscript.

CRedit authorship contribution statement: We follow the CRediT taxonomy.¹² Conceptualization, R.W. and K.T.; methodology, R.W., K.T. and F.W.; software, R.S., F.W. and S.V.; validation, F.W. and S.V.; formal analysis, R.W. and K.T.; investigation, R.W., K.T. and F.W.; resources, R.W. and K.T.; data curation, R.S., F.W., and S.V.; writing—original draft preparation, K.T. and R.W.; writing—review and editing, K.T. and R.W.; visualization, F.W.; supervision, R.W. and K.T.; project administration, R.W. and K.T.; funding acquisition, R.W. and K.T.

AUTHOR DECLARATIONS

Conflict of Interest

There are no conflicts to disclose.

Ethics Approval

The use of the corpus ([Guevara-Rukoz et al., 2020](#)) for the current study does not require ethics approval. According to the University of Florida Institute Review Board, data obtained from another source (not directly from the patient or their records) that is either totally anonymous and unlinkable to the person from whom it was obtained, or is coded in such a way that the researcher obtaining the data does not know who it belongs to, AND a confidentiality agreement assures the researcher cannot learn the identity of the person from whom the data were obtained, does not meet the federal definition of a human subject and is therefore exempt from review.

DATA AVAILABILITY

The corpus we used ([Guevara-Rukoz et al., 2020](#)) is open-sourced under “Creative Commons Attribution-ShareAlike” (CC BY-SA 4.0) license (Creative Commons, 2019). The data that support the findings of this study are available in the Open Science Framework repository, <https://doi.org/10.17605/OSF.IO/384B5>. The repository contains two components: i) the model, code, and instructions for generating the acoustic parameter measures (<https://doi.org/10.17605/OSF.IO/YJA46>); and ii) the data and statistical analyses (<https://doi.org/10.17605/OSF.IO/H8YQA>).

¹A reviewer pointed out that this post-nasal blocking of spirantization is an outdated generalization. Different degrees of spirantization are applied depending on the preceding sound in most geographical regions, and in some places /b d g/ lenite only after vowels (e.g., in Central America, the Caribbean, Canary Islands, and Judeo-Spanish) [see [Hualde \(2005\)](#) for a review].

²[Harris et al. \(2024\)](#) also introduced another metric, called “Noise,” which measures the degree of periodicity, but it is not relevant to the present paper.

³For a detailed description of the triangular Mel filters, please see formula (1) in [Davis and Mermelstein \(1980\)](#) and Section 16.2.4 in [Jurafsky and Martin \(2024\)](#).

⁴For further details, see [Cernak et al. \(2017\)](#).

⁵It is also worth noting that the data under analysis are not excerpts of controlled read sentences or phrases like those examined by [Kingston \(2008\)](#). However, [Kingston \(2008, p. 25\)](#) stated that his method should still apply in less careful speech: “The analytical technique can, however, be applied just as readily to less careful and formal speech.”

⁶It is certainly possible that other Phonet feature models can be used, e.g., voice and syllabic, and pause for segmental deletion.

⁷In order to perform an evaluation of the approach of [Kingston \(2008\)](#), we chose to follow his parameters as closely as possible. This includes the ± 50 ms threshold to the left and right of the sound edges, and the choice of the frequency bands. For a detailed illustration of Kingston’s approach, see [Kingston \(2008, his Figs. 2 and 3\)](#). We acknowledge that these parameters are potentially arbitrary, and readers should experiment with other values, following the footsteps of [Ennever et al. \(2017\)](#).

⁸The full descriptive statistics tables (mean, standard deviation, first quartile, third quartile, IQR range) can be found in the OSF repository.

⁹While [Barr et al. \(2013\)](#) recommend fitting the most complex random effects structure justified by the data, we chose not to follow this recommendation. Instead, we chose to delimit our researchers’ degrees of freedom, and we specified our models’ structures (fixed and random) by focusing on the variables of greatest theoretical interest.

¹⁰We thank a reviewer for suggesting the addition of these three random slopes.

¹¹As suggested by a reviewer, there might be theoretical interests for examining the interaction between Voicing and Place of Articulation and Voicing and Preceding sound to better understand the nature of lenition with /p t k/ as it has not been reported in Argentine Spanish (see Colantoni and Marinescu, 2010). Given that the primary goal of this paper is not to further understand the lenition of stops in Argentine Spanish but rather to evaluate the consistency of lenition measures between the Phonet approach and Kingston's approach, we invite the readers to examine these interaction terms using the dataset and analysis scripts that we have made available under DATA AVAILABILITY.

¹²<https://credit.niso.org/>

- Bailey, G. (2016). "Automatic detection of sociolinguistic variation using forced alignment," in *University of Pennsylvania Working Papers in Linguistics: Selected Papers from New Ways of Analyzing Variation (NWAY 44)*, Vol. 22, pp. 10–20.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *J. Mem. Lang.* 68(3), 255–278.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.* 67(1), 1–48.
- Broś, K., Żygis, M., Sikorski, A., and Wołejko, J. (2021). "Phonological contrasts and gradient effects in ongoing lenition in the Spanish of Gran Canaria," *Phonology* 38(1), 1–40.
- Carrasco, P., and Hualde, J. I. (2009). "Spanish voiced obstruent allophony reconsidered," in *Phonetics and Phonology in Iberia (PaPI)*, Las Palmas de Gran Canaria, Spain (17–19 June 2009).
- Cernak, M., Orozco-Arroyave, J. R., Rudzicz, F., Christensen, H., Vásquez-Correa, J. C., and Nöth, E. (2017). "Characterisation of voice quality of Parkinsons disease using differential phonological posterior features," *Comput. Speech Lang.* 46, 196–208.
- Cohen Priva, U., and Gleason, E. (2020). "The causal structure of lenition: A case for the causal precedence of durational shortening," *Language* 96(2), 413–448.
- Colantoni, L., and Marinescu, I. (2010). "The scope of stop weakening in Argentine Spanish," in *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, Cascadilla Press, Austin, TX, pp. 100–114.
- Cole, J., Hualde, J. I., and Iskarous, K. (1999). "Effects of prosodic and segmental context on /g/-lenition in Spanish," in *Proceedings of the Fourth International Linguistics and Phonetics Conference*, Karolinum Press, Prague, Vol. 2, pp. 575–589.
- Coretta, S., Casillas, J. V., Roessig, S., Franke, M., Ahn, B., Al-Hoorie, A. H., Al-Tamimi, J., Alotaibi, N. E., AlShakhori, M. K., Altmiller, R. M., Arantes, P., Athanasopoulou, A., Baese-Berk, M. M., Bailey, G., Sangma, C. B. A., Beier, E. J., Benavides, G. M., Benker, N., BensonMeyer, E. P., Benway, N. R., Berry, G. M., Bing, L., Bjorndahl, C., Bolyanatz, M., Braver, A., Brown, V. A., Brown, A. M., Brugos, A., Buchanan, E. M., Butlin, T., Buxó-Lugo, A., Caillol, C., Cangemi, F., Carignan, C., Carraturo, S., Caudrelier, T., Chodroff, E., Cohn, M., Cronenberg, J., Crouzet, O., Dagar, E. L., Dawson, C., Diantoro, C. A., Dokovova, M., Drake, S., Du, F., Dubuis, M., Duême, F., Durward, M., Egurtzegi, A., Elsharif, M. M., Esser, J., Ferragne, E., Ferreira, F., Fink, L. K., Finley, S., Foster, K., Foulkes, P., Franzke, R., Frazer-McKee, G., Fromont, R., García, C., Geller, J., Grasso, C. L., Greca, P., Grice, M., Grose-Hodge, M. S., Gully, A. J., Halfacre, C., Hauser, I., Hay, J., Hayward, R., Hellmuth, S., Hilger, A. I., Holliday, N., Hoogland, D., Huang, Y., Hughes, V., Isasa, A. I., Ilchovska, Z. G., Jeon, H.-S., Jones, J., Junges, M. N., Kaefer, S., Kaland, C., Kelley, M. C., Kelly, N. E., Kettig, T., Khattab, G., Koolen, R., Krahmer, E., Krajewska, D., Krug, A., Kumar, A. A., Lander, A., Lentz, T. O., Li, W., Li, Y., Lialiou, M., Ronaldo, M., Lima, J., Lo, J. J. H., Otero, J. C. L., Mackay, B., MacLeod, B., Mallard, M., McConnellogue, C.-A. M., Moroz, G., Murali, M., Nalborczyk, L., Nenadić, F., Nieder, J., Nikolić, D., Nogueira, F. G. S., Offerman, H. M., Passoni, E., Péliissier, M., Perry, S. J., Pfiffner, A. M., Proctor, M., Rhodes, R., Rodríguez, N., Roepke, E., Röer, J. P., Sbacco, L., Scarborough, R., Schaeffler, F., Schlee, F., Schmitz, D., Shiryayev, A., Sósokuthy, M., Spaniol, M., Stanley, J. A., Strickler, A., Tavano, A., Tomaschek, F., Tucker, B. V., Turnbull, R., Ugwuanyi, K. O., Urrestarazu-Porta, I., van de Vijver, R., Engen, K. J. V., van Miltenburg, E., Wang, B. X., Warner, N., Wehrle, S., Westerbeek, H., Wiener, S., Winters, S., Wong, S. G.-J., Wood, A., Wottawa, J., Xu, C., Zárate-Sández, G., Zellou, G., Zhang, C., Zhu, J., and Roettger, T. B. (2023). "Multidimensional signals and analytic flexibility: Estimating degrees of freedom in human-speech analyses," *Adv. Methods Pract. Psychol. Sci.* 6(3), 25152459231162567.
- Davis, S., and Mermelstein, P. (1980). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust. Speech Signal Process.* 28(4), 357–366.
- DiCanio, C. T., Zhang, C., Whalen, D. H., and García, R. C. (2020). "Phonetic structure in Yoloxóchitl Mixtec consonants," *J. Int. Phonetic Assoc.* 50(3), 333–365.
- Eddington, D. (2011). "What are the contextual phonetic variants of in colloquial Spanish?," *Probus* 23(1), 1–19.
- Ennever, T., Meakins, F., and Round, E. R. (2017). "A replicable acoustic measure of lenition and the nature of variability in Gurindji stops," *Lab. Phonol.* 8(1), 20.
- Escure, G. (1977). "Hierarchies and phonological weakening," *Lingua* 43(1), 55–64.
- Figuroa, M., and Evans, B. G. (2015). "Evaluation of segmentation approaches and constriction degree correlates for spirant approximant consonants," in *ICPhS*, <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0718.pdf>.
- Foley, J. (1977). *Foundations of Theoretical Phonology* (Cambridge University Press, Cambridge).
- Guevara-Rukoz, A., Demirsahin, I., He, F., Chu, S.-H. C., Sarin, S., Pipatsrisawat, K., Gutkin, A., Butryna, A., and Kjartansson, O. (2020). "Crowdsourcing Latin American Spanish for low-resource text-to-speech," in *Proceedings of the 12th Language Resources and Evaluation Conference (LREC), European Language Resources Association (ELRA)*, Marseille, France, pp. 6504–6513.
- Harris, J., and Urua, E.-A. (2001). "Lenition degrades information: Consonant allophony in Ibibio," *Speech Hearing Lang. Work Progress* 13, 72–105.
- Harris, J., Urua, E.-A., and Tang, K. (2024). "A unified model of lenition as modulation reduction: Gauging consonant strength in Ibibio," [PsyArXiv](https://arxiv.org/abs/2408.14111).
- Hualde, J. I. (2005). *The Sounds of Spanish* (Cambridge University Press, Cambridge).
- Hualde, J. I. (2013). *Los Sonidos Del Español: Spanish Language Edition* (Cambridge University Press, Cambridge).
- Hualde, J. I., Simonet, M., and Nadeu, M. (2011). "Consonant lenition and phonological recategorization," *Lab. Phonol.* 2(1), 301–329.
- Jurafsky, D., and Martin, J. H. (2024). *Speech and Language Processing*, 3rd ed., <https://web.stanford.edu/jurafsky/slp3>, draft edition.
- Katz, J. (2016). "Lenition, perception and neutralisation," *Phonology* 33(1), 43–85.
- Kendall, T., Vaughn, C., Farrington, C., Gunter, K., McLean, J., Tacata, C., and Arnson, S. (2021). "Considering performance in the automated and manual coding of sociolinguistic variables: Lessons from variable (ING)," *Front. Artif. Intell.* 4, 648543.
- Kingston, J. (2008). "Lenition," in *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, Cascadilla Press, Somerville, MA, pp. 1–31.
- Kirchner, R. (2004). *Consonant Lenition*, Chap. 10 (Cambridge University Press, Cambridge), pp. 313–345.
- Kirchner, R. M. (2001). "Outstanding dissertations in linguistics," in *An Effort Based Approach to Consonant Lenition* (Routledge, New York).
- Lenth, R. V., Buernker, P., Herve, M., Love, J., Riebl, H., and Singman, H. (2021). "emmeans: Estimated marginal means, aka least-squares means [R package]," cran.r-project.org/web/packages/emmeans/index.html.
- Lewis, A. M. (2001). *Weakening of Intervocalic /p, t, k/ in Two Spanish Dialects: Toward the Quantification of Lenition Processes* (University of Illinois at Urbana-Champaign, Champaign, IL).
- Liu, S. A. (1996). "Landmark detection for distinctive feature-based speech recognition," *J. Acoust. Soc. Am.* 100(5), 3417–3430.
- Magloughlin, L. (2018). /t/ and /d/ in North American English: Phonologization of a coarticulatory effect, Ph.D. thesis, Université D'Ottawa/University of Ottawa, Ottawa.
- Martínez Celdrán, E. (2001). "Cuestiones polémicas en los fonemas sonantes del español," *LEA: Lingüística Española Actual* 23(2), 159–172.

- Martínez Celdrán, E., and Regueira, X. L. (2008). "Spirant approximants in Galician," *J. Intl. Phonetics Assoc.* **38**(1), 51–68.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). "Montreal Forced Aligner: Trainable text-speech alignment using Kaldi," in *Proceedings of Interspeech 2017*, in International Speech Community Association (ISCA), Stockholm, Sweden, pp. 498–502.
- McLarty, J., Jones, T., and Hall, C. (2019). "Corpus-based sociophonetic approaches to postvocalic R-lessness in African American language," *Am. Speech* **94**(1), 91–109.
- Ohala, J. J. (1981). "Speech timing as a tool in phonology," *Phonetica* **38**(1), 204–212.
- Ortega-Llebaria, M. (2003). "Effects of phonetic and inventory constraints in the spirantization of intervocalic voiced stops: Comparing two different measurements of energy change," in *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS-15)*, Barcelona, Spain, Vol. 7, pp. 2817–2820.
- Ortega-Llebaria, M. (2004). "Interplay between phonetic and inventory constraints in the degree of spirantization of voiced stops: Comparing intervocalic /b/ and intervocalic /g/," in *Laboratory Approaches to Spanish Phonology*, edited by T. L. Face (De Gruyter Mouton, Berlin), pp. 237–253.
- Pandey, A., Gogoi, P., and Tang, K. (2020). "Understanding forced alignment errors in Hindi-English code-mixed speech: A feature analysis," in *Proceedings of First Workshop on Speech Technologies for Code-Switching in Multilingual Communities 2020*, INCOMA Ltd., held online, pp. 13–17.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Courapeau, D., Brucher, M., Perrot, M., Duchesnay, E., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., and Passos, A. (2011). "Scikit-learn: Machine learning in Python," *J. Machine Learn. Res.* **12**, 2825–2830.
- Roettger, T. B. (2019). "Researcher degrees of freedom in phonetic research," *Lab. Phonol.* **10**(1), 1–27.
- Ségéral, P., and Scheer, T. (2008). *Positional Factors in Lenition and Fortition* (De Gruyter Mouton, New York), pp. 131–172.
- Simonet, M., Hualde, J. I., and Nadeu, M. (2012). "Lenition of /d/ in spontaneous Spanish and Catalan," in *Thirteenth Annual Conference of the International Speech Communication Association (Interspeech)*, Portland, OR, Vol. 2, pp. 1414–1417.
- Soler, A., and Romero, J. (1999). "The role of duration in stop lenition in Spanish," in *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS-14)*, San Francisco, Vol. 1, pp. 483–486.
- Steriade, D. (1993). "Closure, release, and nasal contours," in *Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow, Vol. 5 of *Phonetics and Phonology* (Academic Press, San Diego), pp. 401–470.
- Stevens, K. N. (1981). "Evidence for the role of acoustic boundaries in the perception of speech sounds," *J. Acoust. Soc. Am.* **69**(S1), S116.
- Stevens, K. N. (2002). "Toward a model for lexical access based on acoustic landmarks and distinctive features," *J. Acoust. Soc. Am.* **111**(4), 1872–1891.
- Tang, K., Wayland, R., Wang, F., Vellozzi, S., Sengupta, R., and Altmann, L. (2023). "From sonority hierarchy to posterior probability as a measure of lenition: The case of Spanish stops," *J. Acoust. Soc. Am.* **153**(2), 1191–1203.
- Tezloff, K. A. (2020). "On the gradient lenition of Spanish voiced obstruents: A look at onset clusters," *Stud. Hispanic Lusophone Linguistics* **13**(2), 419–449.
- Vásquez-Correa, J. C., Garcia-Ospina, N., Orozco-Arroyave, J. R., Cernak, M., and Nöth, E. (2018). "Phonological posteriors and GRU recurrent units to assess speech impairments of patients with Parkinson's disease," in *International Conference on Text, Speech, and Dialogue*, Springer, Cham, Switzerland, pp. 453–461.
- Vásquez-Correa, J., Klumpp, P., Orozco-Arroyave, J. R., and Nöth, E. (2019). "Phonet: A tool based on gated recurrent neural networks to extract phonological posteriors from speech," in *Proceedings of Interspeech 2019*, Graz, Austria, pp. 549–553.
- Villarreal, D., Clark, L., Hay, J., and Watson, K. (2020). "From categories to gradience: Auto-coding sociophonetic variation with random forests," *Lab. Phonol.* **11**(1), 1–31.
- Wayland, R., Tang, K., Wang, F., Vellozzi, S., and Sengupta, R. (2023a). "Lenition measures: Neural networks' posterior probability vs. acoustic cues," *Proc. Meetings Acoust.* **50**(1), 060002.
- Wayland, R., Tang, K., Wang, F., Vellozzi, S., and Sengupta, R. (2023b). "Quantitative acoustic versus deep learning metrics of lenition," *Languages* **8**(2), 98.
- Yuan, J., and Liberman, M. (2009). "Investigating /l/ variation in English through forced alignment," in *Proceedings of Interspeech 2009*, International Speech Community Association (ISCA), Brighton, UK, pp. 2215–2218.
- Yuan, J., and Liberman, M. (2021). "/l/ variation in American English: A corpus approach," *J. Speech Sci.* **1**(2), 35–46.