







From sonority hierarchy to posterior probability as a measure of lenition: The case of Spanish stops

Kevin Tang,^{1,a)}  Ratree Wayland,²  Fenqi Wang,²  Sophia Vellozzi,³  Rahul Sengupta,³ 
and Lori Altmann⁴ 

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

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

³Department of Computer and Information Science and Engineering, University of Florida, Gainesville, Florida, 32611-116120, USA

⁴Department of Speech, Language, and Hearing Sciences, University of Florida, Gainesville, Florida, 32610, USA

ABSTRACT:

A deep learning Phonet model was evaluated as a method to measure lenition. Unlike quantitative acoustic methods, recurrent networks were trained to recognize the posterior probabilities of sonorant and continuant phonological features in a corpus of Argentinian Spanish. When applied to intervocalic and post-nasal voiced and voiceless stops, the approach yielded lenition patterns similar to those previously reported. Further, additional patterns also emerged. The results suggest the validity of the approach as an alternative or addition to quantitative acoustic measures of lenition.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0017247>

(Received 26 August 2022; revised 25 January 2023; accepted 27 January 2023; published online 15 February 2023)

[Editor: James F. Lynch]

Pages: 1191–1203

I. INTRODUCTION

Lenition is one of the most common phonological phenomena in the world’s languages. Broadly speaking, it refers to the “sound changes, whereby a sound becomes ‘weaker’ or where a ‘weaker’ sound bears an allophonic relation to a ‘stronger’ sound” (Kirchner, 1998). Processes commonly agreed to fall under the cover term lenition in the literature are degemination, [tt → t]; deaspiration, [t^h] → [t]; voicing, [t] → [d]; spirantization, [t, d] → [(θ) ð]; flapping, [t,d] → [r]; debuccalisation, [t] → [ʔ, h]; gliding, [t] → [j]; and deletion or loss, [ʔ, h, j] → [∅] (Gurevich, 2011). However, what constitutes “weakening” remains controversial (Bauer, 2008). In addition, conflicting hypotheses on the underlying cause of the lenition process has been proposed, and evaluation of these competing hypotheses is made difficult by a lack of consistent approaches to identify and quantify surface realizations of the target phonemes.

The goal of this study is to evaluate a new approach to quantify degrees of lenition. Unlike previous approaches where values along different acoustic dimensions are directly used to estimate lenition, in this approach, degrees of lenition were estimated from the posterior probabilities of sonorant and continuant phonological features computed directly from the speech signals by bidirectional recurrent neural networks (RNNs). Specifically, our approach projects gradient surface acoustic parameters onto two phonological features that capture the possible categorical manifestation

of Spanish stop lenition from stop (-continuant, -sonorant) to fricative (+continuant, -approximant) or to approximant (+continuant, +sonorant). In addition to being sensitive to language-specific acoustic parameters that are contrastive for the two phonological features, it is semi-automatic. Known factors affecting degrees of lenition of Spanish stops including preceding segments, following vowel height, voicing, and place of articulation of the target stop phonemes were tested to assess the validity of the approach.

A. Lenition: Actuation and cause

Phonologically, “A segment X is said to be weaker than a segment Y if Y goes through an X stage on its way to zero” (Hyman, 1975). Bauer (2008) raised two objections against this historical–phonological definition of lenition, namely, its exclusion of any change to be called lenition until its final zero stage and its assumption that the progression towards zero is monotonic. Phonetically, many definitions of lenition have been proposed. Lenition can be defined as a decrease in the amount of articulatory effort, ease, or undershoot (e.g., Bauer, 2008; Kirchner, 1998, 2013). Another definition views lenition as an increase in sonority which has been widely interpreted as an increase in intensity (e.g., Lavoie, 2001). Furthermore, decreasing resistance to airflow in the oral tract has been taken as the defining acoustic characteristic of lenition (e.g., Kingston, 2008; Lavoie, 2001). Finally, lenition has been defined as the extent to which a consonant modulates the carrier signal (Harris *et al.*, 2023). These many definitions differ, amongst other things, in their abilities to provide a unifying definition

^{a)}Also at: Department of Linguistics, University of Florida, Gainesville, FL, 32611-5454, USA. Electronic mail: kevin.tang@hhu.de

of the phonetic effect that lenition has on consonants. For instance, the articulatory-based account can unify processes that involve a loosening of articulatory stricture, such as spirantisation, vocalisation, and debuccalisation; however, it excludes obstruent voicing because voicing increases impedance to airflow which has the opposite aerodynamic effects of increasing the degree of articulatory aperture (Harris *et al.*, 2023). However, these approaches generally agree that lenition processes and outputs are found in similar environments across languages (e.g., in intervocalic, word-medial positions and in unstressed syllables), and that relative changes in duration and intensity, and degrees of oral constriction of the affected consonants are observed (Broś *et al.*, 2021).

Disagreement remains, however, on the ultimate motivation of the lenition process. Consistent with the articulatory effort-based approach, Kirchner (1998, 2013) proposed that lenition is driven by the grammatical constraint, called LAZY, in Optimality Theoretic account, which stipulates that pronunciation of any given sound should be exerted by as little effort as possible. On the other hand, Kingston (2008) argued that the purpose of lenition is to reduce the interruption of the stream of speech to convey that the affected consonant resides within a prosodic constituent. He hypothesizes that lenition is governed “not by how far articulators have to travel but instead by the difference in intensity the speaker wishes to create between the affected segment and its neighbors” (Kingston, 2008). That is, in Kingston’s view, lenition complements fortition and is governed by the position of the affected segment within a prosodic constituent: with greater intensity and less signal disruption, lenition signals continuation within prosodic constituent while fortition decreases signal intensity and increases signal disruption at the edges of prosodic constituents.

To empirically support his hypothesis, Kingston (2008) analyzed Spanish voiced and voiceless stops /b, d, g, p, t, k/ produced as consonant onsets of verbs by two female speakers—one from Peru and the other from Ecuador. These stops are followed by high (close), mid, or low (open) vowels, and the verbs are produced after a word ending in either the vowel [a] or the nasal [n] in four syntactically and semantically appropriate contexts. The acoustic parameters measured were duration, minimum, and maximum intensity velocity from six different frequency bands, ranging from 0 to 8000 Hz. If the purpose of lenition is to reduce articulatory effort, it should be more likely in the context of a lower than a higher vowel since the distance that the articulators have to travel to a lower vowel is reduced relative to that of a higher vowel. On the contrary, if what motivates lenition is the speakers’ desire to signal to the listeners how different the intensity of the target segment is from its neighbors, it should be more affected by the size of oral constriction of surrounding consonants than the height or openness of the surrounding vowels because differences in oral constriction among consonants are smaller compared to those among vowels. His results

showed that the target stops were less lenited after the nasal [n] than after vowels; when the preceding word was further away syntactically and semantically from the target verb; and voiced stops were more lenited than voiceless stops. Additionally, he mentioned that data from two additional speakers showed that lenition was more likely inside a prosodic constituent than at its edge. These results were interpreted as being consistent with his hypothesis that the purpose of lenition is to perceptually minimize separation between the affected consonant and its neighbors within a prosodic constituent.

However, consistent with the effort-based hypothesis, effects of vowel height on degrees of lenition have been reported. For example, Simonet *et al.* (2012) reported that /d/ is more lenited after lower vowels than after high vowels among Catalan–Spanish and Catalan-dominant bilinguals. In contrast, Cole *et al.* (1999) and Ortega-Llebaria (2004) found that Spanish /g/ was less lenited between low vowels than between high vowels while no effect of vowel height was found for /b/ (Ortega-Llebaria, 2003, 2004). In addition, though dispreferred compared to non-initial positions (Escure, 1977; Ségéral and Scheer, 2008), lenition at prosodically strong positions (i.e., prosodic domain-initial) has otherwise been attested. For example, /p/ and /k/ in Murrinh-Patha are lenited primarily in the onset of stressed and usually word-initial syllables (Mansfield, 2015). These results reveal the present but inconsistent effects of flanking vowels on lenition across places of articulation and suggest that consonant weakening may also occur at the edge of a prosodic constituent.

In addition to vowel context, place of articulation and position in a prosodic unit, lenition may also be affected by stress and segmental duration. For example, Cole *et al.* (1999), Ortega-Llebaria (2004), and Colantoni and Marinescu (2010) found that stress inhibits lenition. Soler and Romero (1999) found a significantly positive correlation between segmental duration and degree of constriction in the spirantization phenomenon of Spanish. Cohen Priva and Gleason (2020) argued that reduced duration is the cause of lenition processes, at least for American English.

However, testing hypotheses on the underlying cause of lenition and its surface realization is hindered in part by a lack of consistent and systematic approaches used to detect and quantify occurrences and degrees of lenition. Various methods from visual inspection of the waveforms and spectrograms to quantitative acoustic analysis have been used in the literature, making a comparison of results across studies difficult. In addition, it is unclear which acoustic correlates of lenition are language-general and which may be language-specific and how they should be properly measured (e.g., Bouavichith and Davidson, 2013; Cohen Priva and Gleason, 2020; Ennever *et al.*, 2017; Hualde *et al.*, 2012; Kingston, 2008; Warner and Tucker, 2011). Consequently, various acoustic dimensions including duration, intensity, rate of intensity change, percentage of voicing, harmonics-to-noise ratio, etc., absolutely or relatively

(to their flanking segments) measured, have been taken as acoustic correlates of lenition across studies. Moreover, different acoustic dimensions have been used to evaluate degrees of lenition. For example, in [Colantoni and Marinescu \(2010\)](#), duration, consonant-vowel (CV) intensity ratio, and percentage of voicing were used to evaluate degrees of lenition between voiced and voiceless consonants; however, only the first two parameters were used to investigate the effects of the type of consonant, the quality of the flanking vowels, and stress on lenition.

To address the inconsistent segmentation issue, [Ennever et al. \(2017\)](#) proposed an automated method to measure duration and lenition of stop consonants. The method, which was argued to correspond well to articulation, allows for a systematic, objective, and consistent demarcation of lenited segments from fully occluded to highly lenited types in the acoustic data. In addition to duration, quantitative measures of lenition based on the segment's rates of change in intensity profiles are also generated. When applied to the intervocalic voiceless stops /p, t, k/ of casual speech in Gurindji (Pama-Nyungan, Australia) produced by one female speaker, varying degrees of lenition were found. In addition, no significant effects of preceding and following vowels were found. Furthermore, no positive effect of word-medial position on lenition relative to word-initial position was found. However, adopting [Ennever et al. \(2017\)](#) method, [Katz and Pitzanti \(2019\)](#) revealed a number of prosodic and other contextual influences on the acoustics of lenition in Campidanese Sardinian consonants.

In this study, alternative to traditional quantitative acoustic methods, a new lenition quantification method based on posterior probabilities of the phonological features (i.e., continuant and sonorant) is introduced. The posterior probabilities are directly learned from the speech signal by a deep learning model known as "Phonet." Its performance was evaluated on voiced and voiceless stop lenition in a corpus of Argentinian Spanish. Our review of the existing approaches of quantifying lenition suggests that researchers face the challenge of (i) selecting the relevant acoustic correlates since the potential acoustic space is large and that lenition can be language-specific and (ii) measuring the correlates automatically. Building on these observations, our alternative approach aims to meet the following desiderata: it should be (i) motivated by phonology, (ii) customizable for a specific language, (iii) largely automatic, and (iv) able to measure categorical and gradient manifestations of lenition. In the sections below, we will outline the basis of our approach.

B. Sonority and lenition

One classic definition of lenition is an increase in the sonority of a consonant ([Lavoie, 2001](#)). Sonority is a fundamental notion in phonetics and phonology. According to a phone's sonority (consonants and vowels), all consonants and vowels can be ranked on a scale, called the sonority hierarchy. The sonority hierarchies build on systematic

observations of the attested cross-linguistic phonotactic patterns of different natural classes. It plays a central role in many descriptions of syllable structures and phonotactics.

The phonetic nature of sonority is not without criticism ([Harris, 2006](#); [Henke et al., 2012](#); [Kawasaki-Fukumori and Ohala, 1997](#)), with different phonetic correlates having been suggested, such as intensity or loudness ([Parker, 2002](#)) and pitch ([Albert and Nicenboim, 2022](#)). As pointed out by [Clements \(1990\)](#), the absence of a language-independent, consistent, physical characterizing of sonority makes it impossible to explain the nearly universal nature of sonority constraints across languages. Both physical and perceptual properties of sonority have been proposed. For instance, [Ladefoged \(1993\)](#) defined sonority in terms of the loudness of a sound, which is related to its acoustic energy relative to other sounds having the same length, stress, and pitch. On the other hand, [Clements \(1990\)](#) argued that sonority is related not to the sounds' loudness or audibility but to their relative perceived resonance and acoustically characterized by prominent, well-defined formant peaks. Possessing these characteristics of [+sonorant] to the highest degree, vowels stand at the top of the hierarchy while oral stops and fricatives (i.e., obstruents) stand at the bottom of the scale. While there are numerous versions of the hierarchies ([Parker, 2002](#)) for a review of more than 100 hierarchies, dating back to [Jespersen \(1899\)](#), [Sheldon \(1893\)](#), and [Whitney \(1865\)](#), the basic sonority scale from the most sonorous to the least sonorous sound is as follows: vowel > semivowel or glides > liquid > nasal > fricative > oral stop. The phonological feature [continuant] whose phonetic correlate is relatively less controversial can capture the contrast between oral stops and the remaining sounds on the sonority scale, particularly fricative. This feature denotes sounds articulated with air escaping through the oral cavity throughout their articulation. Articulated with oral airflow being completely obstructed, oral stops are specified as [-continuant] while fricative, liquid, glides, and vowels are [+continuant]. Due to the presence of an occlusion in the oral cavity, nasal consonants are classified as [-continuant] by some, but [+continuant] by others on the basis of continuous acoustic signal through the nasal cavity. In this study, nasals are specified as [-continuant]. Phonological analyses of lenition based on the sonority scale rely on symbolic representation of speech sounds and phonological features. The lenition of Spanish stops would therefore involve categorical feature changes from [-continuant] to [+continuant] and from [-sonorant] to [+sonorant]. However, to capture degrees of lenition, we must look beyond categorical manifestations of lenition changes.

1. From categories to gradience

Our approach is inspired by computational approaches used in existing studies of phonetic variations. The aim of these approaches is to measure gradient variations using the canonical realizations of the phenomenon (i.e., two ends of the variation spectrum). Many studies have relied on forced

alignment systems to determine pronunciation variations (e.g., [dʒ]-[z] and [p^h]-[f] variations in Hindi–English code-mixed speech (Pandey *et al.*, 2020), “g”-dropping in English (Kendall *et al.*, 2021; Yuan and Liberman, 2011a), “th”-fronting, “td”-deletion, and “h”-dropping in English (Bailey, 2016). This approach relies on the fact that forced alignment systems typically take word-level orthographic transcriptions as input referring to a pronunciation dictionary with phone-level transcription. Crucially, each word entry in the dictionary can be given multiple pronunciations. For instance, to model “th”-fronting, one could provide two pronunciations for all word entries that are potentially subjected to “th”-fronting, one with [θ] and one with [f]. A trained forced aligner can automatically determine which pronunciation has the highest probability given the acoustic signal of each word token.

This method can therefore determine the surface realization of phonological variations. However, this method is limited to the granularity of the phone set. A forced alignment model contains an acoustic model for each phone type defined in the pronunciation dictionary; therefore, this method could only determine the variation with predefined segments. How could one obtain a more gradient measure of variations (e.g., degrees of “th”-fronting as opposed to simply coding a token as [θ] or [f])? Yuan and Liberman (2009) proposed an innovative method of measuring the gradient variation of /l/-darkness in American English using the probability scores extracted during the forced alignment procedure. Probability score is defined as the log probability (log probability density) of the aligned segment to be a particular phone. In this method, all /l/ tokens from a corpus of American English were forced aligned twice: first by a model trained on light /l/ (word-initial) and second by a model trained on dark /l/ (word-final and word-final consonant clusters). The difference between the log probability scores from the dark /l/ alignment and the light /l/ alignment indicates degrees of /l/-darkness. The method was extended to examine finer variation of both types of /l/s by Yuan and Liberman (2011b). In addition to demonstrating the categorical distinction between dark (in syllable coda) and light /l/ (in syllable onset), their results also revealed that intervocalic dark /l/ is less dark than canonical syllable-coda dark /l/, and its degrees of darkness depend on the stress of the flanking vowels. Intervocalic light /l/ is always light and is lighter than canonical syllable-onset /l/. Similarly, Magloughlin (2018) applied this method to measuring the gradient variation of /t/-/d/ affrication in English by aligning /tɹ/ and /dɹ/ tokens twice, using acoustic models of /t/ and /d/, and models of /tʃ/ and /dʒ/. The degree of affrication is the log probability scores from the /tʃ,dʒ/ alignment and the /tɹ,dɹ/ alignment.

This use of probability estimates from token classification is not limited to using acoustic models in a forced alignment system. For instance, McLarty *et al.* (2019) examined the degree of r-lessness of postvocalic /r/ in English. Three kinds of segments were extracted: canonical r-less tokens

(oral vowels that are not preceding a liquid or nasal), canonical r-full tokens (prevocalic /r/), and ambiguous tokens (postvocalic /r/ which has variable r-lessness). All but the ambiguous tokens were divided into a training set and a test set. Support vector machines were trained to classify the canonical r-less tokens and the canonical r-full tokens using Mel-frequency cepstral coefficients (MFCCs) as the acoustic representations. The model achieved a mean accuracy of 98.95% on the test set. Degrees of r-lessness were measured for each of the ambiguous tokens (postvocalic /r/) by applying the trained model to yield a probability estimate of being r-less as opposed to r-full. In a similar study on two English sociophonetic variables (non-prevocalic /r/ and word-medial intervocalic /r/), Villarreal *et al.* (2020) employed a different classification method, random forest, to automate coding categorical manifestations of the two variables using a set of acoustic measures.

Importantly, the method used by most of these studies trains on surface segments that are *not* realized from the segments that are subject to the variation of interest. That is, the method relies on the fact that these surface segments have sufficiently similar acoustic characteristics to the possible canonical realizations of a variation. For instance, in the case of “th”-fronting, the model would be trained to classify tokens that are either canonically [θ] or canonically [f], and these canonical tokens are not subjected to “th”-fronting. Similarly, in the case of /l/-darkening, canonical light /l/s and dark /l/s would be used for the training phase, and the trained model would then be applied to /l/s that have variable degrees of darkening.

The suitability of this method to estimate the categorical manifestation of lenition using surface segments is suggested by the results of Cohen Priva and Gleason (2020). In that study, a spoken corpus of American English was used to model a range of processes commonly accepted as lenition processes. Using regression models, the authors examined the acoustic properties of lenition processes. Pairs of segments that are relevant to a lenition process were selected and subjected to a regression analysis, e.g., for the lenition process /t/ → [d], the two relevant surface segments would be [t] and [d]. The authors examined three types of modeling methods which differ in the underlying representation of the surface segments. Their first method compares the surface forms of two segment types, regardless of whether their underlying form was the segment in question, e.g., for /t/ → [d], the [t] and [d] tokens do not need to share the underlying form /t/. In contrast to the first method, their second method compares only the surface forms of two segment types that have the same underlying form, e.g., [t] and [d] have the underlying form /t/. Their third method compares only segments that surfaced unchanged, e.g., the [t] tokens realized from /t/ and the [d] tokens from /d/. Crucially, all three modeling approaches yielded the same findings, suggesting that the acoustic changes of a given lenition process can be captured by comparing the surface forms of two segments, regardless of whether their underlying form was the segment in question.

Our approach aims to tackle a whole class of lenition. Therefore, unlike [Cohen Priva and Gleason \(2020\)](#), we must go beyond classifying pairs of segments that are relevant to a lenition process, but rather two groups of segments that are categorized by a binary phonological feature. In this study, we focus on the probability of the phonological feature [continuant], which differentiates stops from non-stops (e.g., stops lenited as a fricative), and the phonological feature [sonorant], which differentiates stops and fricatives from non-stops and non-fricatives (e.g., stops lenited as an approximant), because they capture the two categorical realizations of stop lenition in Spanish. Other phonological features, such as [syllabic], could further capture other stages of lenition, such as a vowel-like realization, but they are not analyzed in the current study. For instance, in Cibaño Spanish, the coda liquids undergo vocalisation ([Harris, 1969](#)). A fricative-like realization would have a high [continuant] probability but a low [sonorant] probability, while an approximant-like realization would have a high [continuant] probability and a high [sonorant] probability. Two models were trained: one for each phonological feature. Unlike [Yuan and Liberman \(2009, 2011b\)](#) where degrees of phonetic variation were estimated from the difference between the log probability scores of the two forced alignment models (dark /l/ and light /l/), degrees of lenition are reflected in the probability of each phonological feature estimated from acoustic properties of the input signals by deep neural networks known as the “Phonet” model.

2. “Phonet”

Originally proposed by [Vásquez-Correa et al. \(2019\)](#), Phonet estimates posterior probabilities of phonological features using bi-directional recurrent neural networks (RNNs) with gated recurrent units (GRUs) and is highly accurate in detecting phoneme and phonological classes in Spanish ([Vásquez-Correa et al., 2019](#)) and modeling the speech impairments of patients diagnosed with Parkinson’s disease ([Vásquez-Correa et al., 2019](#)).

The architecture of Phonet is described in detail in [Vásquez-Correa et al. \(2019\)](#). Briefly, inputs to Phonet are feature sequences based on log-energy distributed across 33 triangular Mel filters computed from 25 ms windowed frames of each 0.5 s chunks of the input signal. These feature sequences are processed by two bidirectional GRU layers so information from the past (backward) and future (forward) states of the sequence are modeled simultaneously. The output sequences of the second bidirectional GRU layer are then passed through a time-distributed, fully connected hidden dense layer, producing an output sequence of the same length as the input. Finally, a phonological class associated with the feature sequence from the input is produced by the connected time-distributed output layer with a softmax activation function. In our study, 23 phonological classes of Spanish were trained by a bank of 23 Phonet networks and 26 phonemes by one network using an Adam optimizer ([Kingma and Ba, 2014](#)). Following [Vásquez-](#)

[Correa et al. \(2019\)](#), to avoid the unbalance of the classes in the training process, a weighted categorical cross-entropy loss function, defined according to Eq. (1) was used:

$$L = - \sum_{i=1}^C w_i p_i \log(\hat{p}_i). \quad (1)$$

The weight factors w_i for each class $i = \{1 \dots C\}$ are defined based on the percentage of samples from the training set that belong to each class. To improve the generalization of the networks, dropout and batch normalization layers were considered.

The current study focused on using acoustic features that are based on MFCCs. This was motivated by how the use of MFCCs is standard in speech technology (such as automatic speech recognition) and known to provide a good overall representation of the acoustic signal, as they often provide a wider range of acoustic information than individual acoustic features ([Davis and Mermelstein, 1980](#); [Huang et al., 2001](#)). Furthermore, previous studies of phonetic variations have also used them as acoustic representations with success ([Kendall et al., 2021](#); [McLarty et al., 2019](#); [Yuan and Liberman, 2009, 2011b](#)). For lenition, a phenomenon without standard acoustic measures, we believe that our MFCC-based features can serve as useful acoustic features for illustrating our approach. We acknowledge that it would be beneficial for future work to examine alternative acoustic representations based on previously proposed acoustic measures of lenition.

In sum, our proposed approach is phonologically motivated, language-specific, largely automatic, and it can capture categorical and gradient manifestations of lenition. It is motivated by sonority, a phonological concept that is deeply rooted in phonological analyses of lenition. It is language-specific since it is trained on the acoustic data of the target language, which has the advantage of being able to make use of only contrastive acoustic information for a given phonological feature of the target language. Furthermore, phonological feature sets also can be customized for the target language, for instance, to under-specify specific phonological features ([Lahiri and Reetz, 2002, 2010](#)) or to use features that are motivated by articulatory ([Chomsky and Halle, 1968](#)), acoustic ([Jakobson et al., 1951](#)), or perceptual factors ([Backley, 2011](#)). It is largely automatic since it requires only a segmentally aligned acoustic corpus, which can be obtained using forced alignment and a phonological feature set. Finally, while probability estimates are by themselves gradient, categorical manifestations of lenition at the segmental or natural class level can still be captured by combining probability estimates of different phonological features.

II. METHODS

A. Materials

This study used the Argentinian Spanish Corpus built by [Guevara-Rukoz et al. \(2020\)](#). This open-sourced corpus includes the crowd-sourced recordings from 44 native

TABLE I. Word distribution by conditions: voicing, place of articulation, previous phone, and following vowel. The number left and right of the slash in each cell represents 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	818/55	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	34/1	0/0	290/36	1683/95	577/71
	Nasal	0/0	0/0	0/0	69/9	259/14	109/16

speakers of Argentinian Spanish (female: 31; male: 13). The corpus was divided into two subsets by gender. The male sub-corpus contains 2.4 h of recording with 16914 words (3342 unique words). The female sub-corpus contains 5.6 h of recording with 35360 words (4107 unique words). Based on Kingston (2008), we selected word tokens with /b, d, g, p, t, k/ as word-initial segments followed by vowels with different degrees of openness and preceded by vowels or nasals from the preceding words. Table I specifies the number of word tokens and word types by conditions: stress (stressed or unstressed), voicing (voiced or voiceless), place of articulation (bilabial, dental, and velar), previous phone (vowel and nasal), and following vowel (open, mid, and close).

B. Procedure

The forced alignment process was performed using the Montreal Forced Aligner (version: 2.0) (McAuliffe et al., 2017). A phonemic pronunciation dictionary for the transcription of the corpus words was generated based on grapheme-to-phoneme mapping in the International Phonetic Alphabet (IPA) by Hualde (2013), which was then used to train new acoustic models for the corpus and align the text-grids to the acoustic signals. The new acoustic model was a triphone model, meaning the training process looked at the preceding and succeeding phones for a target phone and made the necessary acoustic adjustments during alignment. The phone set parameter was set to IPA, which enabled extra decision tree modeling based on the specified phone set. Any other parameters maintained the default.

Model training was performed using the 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). The targets /b, d, g/ were not included (i.e., silenced out) since they are expected to be ambiguous in terms of their realizations in the two features of interest, continuant, and sonorant. That way, the resultant trained models would not have been contaminated by the ambiguous tokens. Altogether, 23 phonological classes including 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. Similar to Vásquez-Correa et al. (2019) one additional model was included to train phonemes. However, in addition to the 18 phonemes from Vásquez-Correa et al. (2019), seven additional phonemes including stressed /'a, 'e, 'i, 'o, 'u/, /ɲ/ and /spn/ for speech-like noise were also included. The phonemes /θ/ and /ɰ/ were excluded since they do not exist in Argentinian Spanish. The voiceless prepalatal fricative /j/ corresponds to the graphemes “ll” and “y.” Table II shows the complete feature chart for all phonemes in the corpus and their corresponding graphemes along with their phonological feature values for all 23 phonological classes. The feature value of “+” indicates that a phoneme is classified as that particular phonological feature while “-” indicates that it is not. Since weakened realizations of Spanish /b, d, g/ are either a fricative or an approximant (Hualde et al., 2012), of the 23 phonological features, sonorant and continuant are our features of interest.

The model was highly accurate in detecting the different phonological classes, showing unweighted average recall (UAR) ranges from 94%–98%. The UARs for the sonorant and continuant features are 97% and 96%, respectively. For individual phoneme detection, the model output a high degree of variation, showing detection accuracy ranges from 42% for /spn/ to 96% for /f/. Excluding /spn/, the range was 59% for /'e/ to 96% for /f/.

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. If a phone token contains multiple frames, then the average of the prediction of the middle frame(s) was used as the prediction of that phone.¹

C. Analysis

Similar to the model structure in Kingston (2008), five fixed factors were included in the linear mixed-effects regression models. These variables were stress (stressed or unstressed), voicing (voiced or voiceless), place of articulation (bilabial, dental, and velar), previous phone (vowel and nasal), and following vowel (open, mid, and close). There were two ways of coding the level contrasts for these categorical variables: deviation coding for the variables stress, voicing, and previous phone and forward difference coding for the variables’ place of articulation (bilabial > dental > velar) and following vowel (close > mid > open). The dependent variables in the two regression models were sonorant and continuant posterior probabilities, respectively, which were generated by the Phonet model. The two regression models included two different interactions terms but same random intercepts by speaker and word. The linear mixed-effects regression model was performed using the *lmer* function from the *lme4* package (Bates et al., 2015) in R (R Core Team, 2022). After comparing multiple model structures with maximum likelihood, we identified the best-fit model structure for sonorant and continuant posterior

probability, respectively, and their corresponding model formulae were provided as follows: Sonorant posterior probability \sim Stress + Voicing + Previous phone + Following vowel:Place + (1 | Speaker) + (1 | Word) and Continuant posterior probability \sim Stress + Voicing + Place + Previous phone:Following vowel + (1 | Speaker) + (1 | Word). *Post hoc* comparisons of the interaction terms were carried out using emmeans (with Tukey HSD for *p*-value adjustment) (Lenth *et al.*, 2021).

III. RESULTS

A. Sonorant posterior probability

Figure 1 presents the sonorant posterior probability of the word-initial sounds /b, d, g/ before different vowels. As shown in Fig. 1, regardless of the following vowel context, the density distribution of the sonorant posterior probability of the three sounds is left-skewed, suggesting a higher degree of lenition with a generally high sonorant posterior probability. In addition, /b/ had the highest mean sonorant posterior probability (before close vowels: $M=0.932$; before mid vowels: $M=0.966$; before open vowels: $M=0.982$), followed by /d/ (before close vowels: $M=0.904$; before mid vowels: $M=0.932$; before open vowels: $M=0.944$), and then /g/, which only appeared before the mid vowels (before mid vowels: $M=0.877$).

Table III summarizes the fixed-effects coefficients in the mixed-effects model (the upper table of Table III) and the type-III-ANOVA analysis (the lower table of Table III) with sonorant posterior probability as the dependent variable.

As shown in Table III (lower), there were significant main effects of all predictors including stress, voicing, place,

following vowel, and previous phone as well as a significant interaction between place of articulation and the following vowel. As shown in Table III (upper), the model results suggested that the target stops were more sonorant in an unstressed syllable than in a stressed syllable ($\beta = 0.051$, $t = 4.735$, $p < 0.001$). The voiced stops /b, d, g/ were predicted to be more sonorant than their voiceless counterparts ($\beta = 0.659$, $t = 52.803$, $p < 0.001$). Bilabial stops /p, b/ and velar stops /k, g/ were predicted to be more sonorant than dental stops /t, d/ (bilabial vs dental: $\beta = 0.101$, $t = 7.435$, $p < 0.001$; dental vs velar: $\beta = -0.152$, $t = -9.647$, $p < 0.001$). In addition, all target stops tended to be more sonorant when the following vowel was a mid vowel than when it was a close vowel ($\beta = -0.036$, $t = -2.766$, $p < 0.05$), but the difference between the following mid and open vowel contexts did not reach statistical significance ($\beta = -0.016$, $t = -1.237$, $p = 0.216$). Furthermore, all target sounds were predicted to be more sonorant when the previous phone was a vowel than when it was a nasal ($\beta = 0.028$, $t = 3.312$, $p = 0.001$). Finally, there was a significant interaction between place of articulation of the stop and the following vowel height.

To further investigate the significant place x following vowel interaction, a *post hoc* analysis using the Tukey method was performed. The results of pairwise mean comparisons indicated that there was a stronger effect of the following vowel on bilabial and velar stops such that bilabial stops had a significantly lower sonorant posterior probability before a mid vowel than before an open vowel ($\beta = -0.073$, $t = -3.764$, $p = 0.006$) and velar stops had a significantly lower sonorant posterior probability before a close vowel than before a mid vowel ($\beta = -0.122$, $t = -4.801$, $p = 0.006$). In addition, while bilabial and velar stops' posterior probabilities were significantly higher than dental stops in all three vowel contexts,

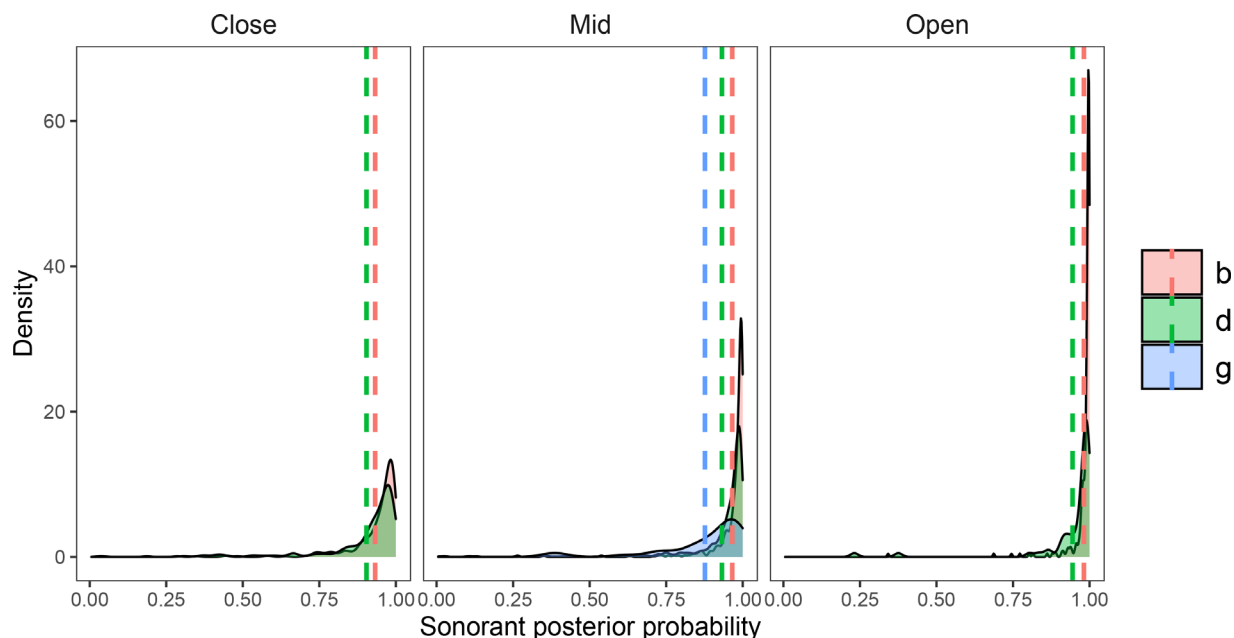


FIG. 1. (Color online) Sonorant posterior probability of /b, d, g/ by following vowel (close, mid, and open). Vertical dashed lines represent mean sonorant posterior probabilities by conditions.

TABLE III. Summaries of sonorant posterior probability: the fixed effects in the linear mixed-effects model (a: upper), and the type-III-ANOVA analysis (b: lower) (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$). The significant p -values are in bold.

Fixed effects: Sonorant posterior probability					
Predictor	β		t	p	
Intercept	0.629	0.014	46.036	<0.001	***
Stress (unstressed)	0.051	0.011	4.735	<0.001	***
Voicing (voiced)	0.659	0.012	52.803	<0.001	***
Place (bilabial)	0.101	0.014	7.435	<0.001	***
Place (dental)	-0.152	0.016	-9.647	<0.001	***
Following vowel (close)	-0.036	0.013	-2.766	0.006	**
Following vowel (mid)	-0.016	0.013	-1.237	0.216	
Previous phone (vowel)	0.028	0.008	3.312	0.001	***
Place (bilabial) \times Following vowel (close)	-0.024	0.029	-0.807	0.419	
Place (dental) \times Following vowel (close)	0.139	0.033	4.202	<0.001	***
Place (bilabial) \times Following vowel (mid)	-0.038	0.034	-1.109	0.268	
Place (dental) \times Following vowel (mid)	-0.095	0.035	-2.746	0.006	**

Type-III-ANOVA: sonorant posterior probability					
Effects	SSE	MSE	F	p	
Stress	1.306	1.306	22.421	<0.001	***
Voicing	162.377	162.377	2788.146	<0.001	***
Place	5.777	2.888	49.596	<0.001	***
Following vowel	0.757	0.379	6.500	<0.01	**
Previous phone	0.639	0.639	10.971	<0.01	***
Place \times Following vowel	1.923	0.481	8.256	<0.001	***

bilabial stops' posterior probabilities were significantly lower than those of velar stops in the mid vowel context only ($\beta = -0.133$, $t = -7.063$, $p < 0.001$) (Fig. 2).

B. Continuant posterior probability

Figure 3 presented the continuant posterior probability of the voiced /b, d, g/ by previous phone and following vowel. As shown in Fig. 3, regardless of the previous phone contexts, the three sounds had the highest mean continuant posterior probability before mid vowels (after nasals:

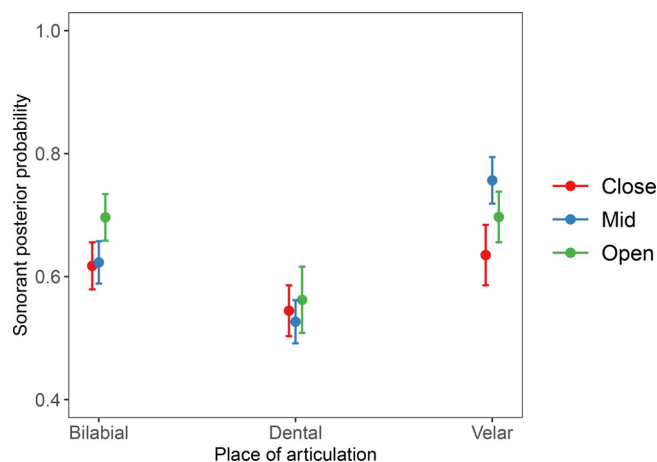


FIG. 2. (Color online) Estimated marginal means of sonorant posterior probability by place of articulation and following vowel. The dots represent the estimated marginal means and the interval lines display 95% confidence interval.

$M = 0.618$; after vowels: $M = 0.881$), followed by close vowels (after nasals: $M = 0.420$; after vowels: $M = 0.818$), and open vowels (after nasals: $M = 0.348$; after vowels: $M = 0.774$), respectively. When the previous phone is a nasal, the density distribution of the continuant posterior probability of the three sounds was right-skewed before the closed and open vowels and left-skewed before the mid vowels. Yet all three vowel heights have a relatively flat distribution, suggesting a lower degree of lenition in this previous phone context. Conversely, when the previous phone is a vowel, the three sounds showed an extremely left-skewed density distribution of the continuant posterior probability, suggesting a relatively higher degree of lenition across the following vocalic contexts.

Table IV summarizes the fixed-effects coefficients in the mixed-effects model (Table IV, upper) and the type-III-ANOVA analysis (Table IV, lower) with continuant posterior probability as the dependent variable.

Similar to the results for sonorant posterior probabilities, we found significant main effects of all predictors. In addition, a significant interaction between previous phone and following vowel was also obtained (see Table IV, lower). The target stops were more continuant in an unstressed syllable than in a stressed syllable ($\beta = 0.031$, $t = 2.901$, $p < 0.01$). Compared to the voiceless stops /p/, /t/, and /k/, their voiced counterparts were predicted to be more continuant ($\beta = 0.587$, $t = 46.376$, $p < 0.001$). Dental stops were predicted to be less continuant than velar stops ($\beta = -0.154$, $t = -10.300$, $p < 0.001$). For vowel context, the target stops tended to be more

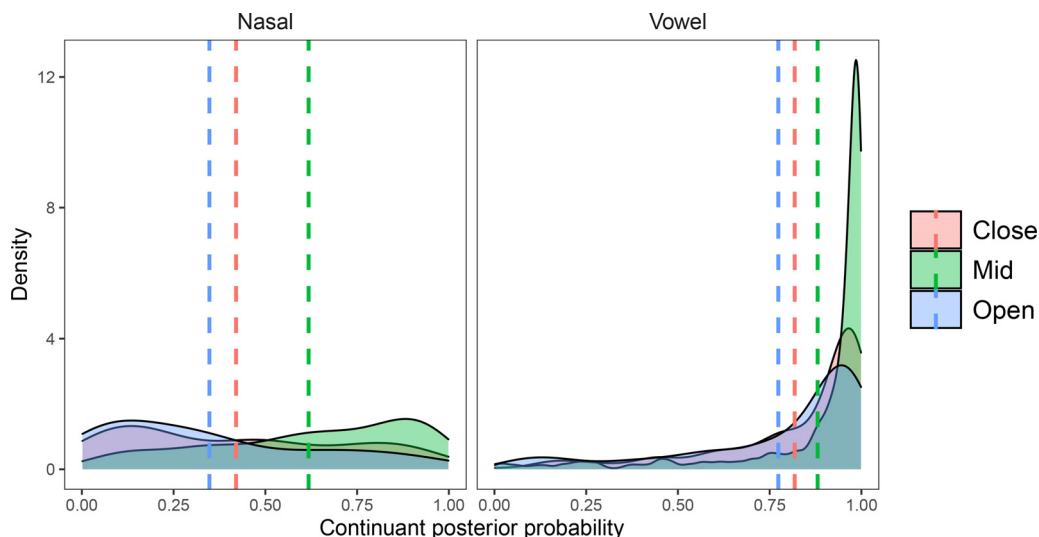


FIG. 3. (Color online) Continuant posterior probability of /b, d, g/ by previous phone (nasal and vowel) and following vowel (close, mid, and open). Vertical dashed lines represent mean continuant posterior probabilities by conditions.

continuant when the following vowel is a mid vowel than when it is a close vowel (close vs mid: $\beta = -0.037, t = -2.494, p < 0.05$). In addition, the target word-initial sounds were predicted to be more continuant when preceded by a vowel than when preceded by a nasal ($\beta = 0.192, t = 19.734, p < 0.001$).

Figure 4 showed the estimated marginal means of continuant posterior probability by previous phone and following vowel. The same *post hoc* analysis using Tukey method was performed to further examine the significant interaction between previous phone and following vowel in the regression model. The results suggested an effect of the following

vowels in the nasal context but not in the vowel context such that after nasals, continuant posterior probabilities were lower when the following vowels are close than when they are mid (after nasals: $\beta = -0.056, t = -2.539, p = 0.113$) or open ($\beta = -0.068, t = -2.541, p = 0.113$), although it is not statistically significant.

IV. DISCUSSION

A new approach to measure lenition was evaluated. In this approach, bidirectional recurrent neural networks (RNNs) were trained to classify Spanish phonemes and

TABLE IV. Summaries of continuant posterior probability: the fixed effects in the linear mixed-effects model (a: upper), and the type-III-ANOVA analysis (b: lower) (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$). The significant p -values are in bold.

Fixed effects: Continuant posterior probability					
Predictor	β	SE	t	p	
Intercept	0.466	0.011	42.015	<0.001	***
Stress (unstressed)	0.031	0.011	2.901	0.004	**
Voicing (voiced)	0.587	0.013	46.376	<0.001	***
Place (bilabial)	0.015	0.013	1.201	0.230	
Place (dental)	-0.154	0.015	-10.300	<0.001	***
Following vowel (close)	-0.037	0.015	-2.494	0.013	*
Following vowel (mid)	0.007	0.015	0.460	0.646	
Previous phone (vowel)	0.192	0.010	19.734	<0.001	***
Following vowel (close) × Previous phone (vowel)	0.038	0.021	1.764	0.078	
Following vowel (mid) × Previous phone (vowel)	0.038	0.022	1.718	0.086	
Type-III-ANOVA: continuant posterior probability					
Effects	SSE	MSE	F	p	
Stress	0.451	0.451	8.414	<0.01	**
Voicing	115.308	115.308	2150.763	<0.001	***
Place	7.021	3.511	65.480	<0.001	***
Following vowel	0.341	0.171	3.182	<0.05	*
Previous phone	20.879	20.879	389.447	<0.001	***
Following vowel × Previous phone	0.416	0.208	3.876	<0.05	*

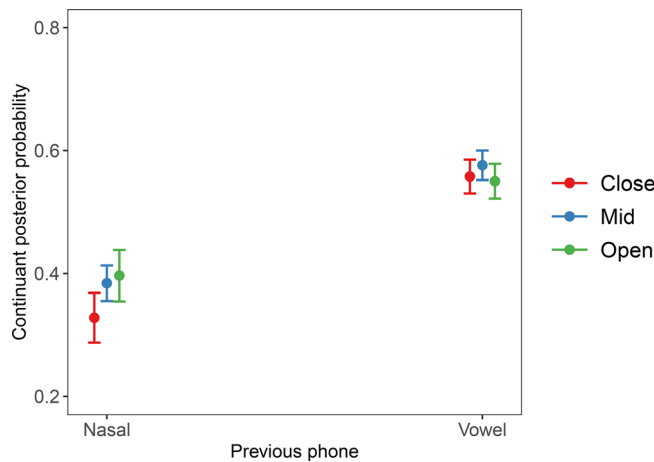


FIG. 4. (Color online) Estimated marginal means of continuant posterior probability by previous phone (nasal and vowel) and following vowel (close, mid, and open). The dots represent the estimated marginal means and the interval lines display 95% confidence interval.

phonological features from the acoustic signals. The networks were trained on an Argentinian Spanish Corpus with Mel-filtered log-energy from each 0.5 s chunk of the speech signal as input. The posterior probabilities of the continuant and sonorant features outputted by the networks were then used as estimates of lenition of voiced /b, d, g/ and voiceless /p, t, k/ stops. The performance of the model was evaluated by comparing lenition patterns as predicted by previous findings using quantitative acoustic methods. Variables known to affect lenition including voicing, stress, preceding segment, and following segment were included in the evaluation.

All main effects were significant for both sonorant and continuant posterior probabilities, with larger effects for voicing, place of articulation, and preceding segment than for the following vowel or stress. Consistent with previous findings in the literature, our regression models predicted more lenition (i.e., higher sonorant and continuant posterior probabilities) for voiced /b, d, g/ relative to voiceless /p, t, k/. In addition, a greater degree of lenition was predicted in an unstressed syllable compared to a stressed syllable. These findings are consistent with both the articulatory effort-based and the perceptual-based approaches of lenition. In addition, the model's prediction that lenition is stronger when the target stops were preceded by a vowel than by a nasal /n/ is in agreement with findings by Kingston (2008). However, the model's prediction that lenition was greater following more open vowels, /e, o, a/, than close vowels, /i, u/, is inconsistent with the perceptual-based hypothesis. According to Kingston (2008) the difference in openness between close and open vowels is so small that more intervocalic consonant lenition in the context of open vowels than in the context of less open vowels is unlikely. Instead, consistent with the articulatory effort-based view of lenition, our results suggest that, overall, the difference in openness between close vs mid and open vowels, but not between mid and open vowels could lead to more lenition.

Nevertheless, significant interactions between place of articulation and openness of the following vowels for sonorant posterior probabilities and between preceding segment and following vowel openness for continuant posterior probabilities were also found. Difference in the interaction pattern between the two phonological features suggest that categorically different surface lenited forms (i.e., fricative and approximant-like) are gradiently varied in different contextual environments: following segments for approximant-like realization, but both preceding and following segments for fricative-like realization.

Follow-up tests revealed gradient effects of the following vowel height on place of articulation of the preceding stops for sonorant posterior probabilities and of following vowel height on preceding segment type (vowel vs nasal consonant) for continuant posterior probabilities. Specifically, for sonorant posterior probabilities, bilabial stops were more lenited (i.e., higher sonorant posterior probabilities) before an open vowel /a/ than before a mid vowel /e, o/, and velar stops were more lenited before a mid vowel relative to a close vowel /i, u/. However, no effect of the openness of the following vowel was found for dental consonants. These results are inconsistent with those of Ortega-Llebaria (2003, 2004) who reported no effects of vowel context on degrees of /b/ lenition but more lenition for /g/ between close vowels /i, u/ than between open vowels among native Caribbean Spanish speakers using a quantitative acoustic method. As for Spanish /d/, Simonet et al. (2012) reported more lenition of /d/ after low vowels than after close vowels in Spanish and Catalan. However, our finding that /g/ is more lenited in a mid vowel context is partially consistent with that of Cole et al. (1999), who found /g/ more lenited in unstressed syllables flanked by /o/ and /u/ vowels compared to /a/ and /i/ and /e/ in Castilian Spanish. Yet, the grouping of /o/ with /u/ and /i/ with /e/ in the study by Cole et al. (1999) renders the comparison less straight forward. Nonetheless, these results strongly suggested that output of a lenition process is gradient and in a language- and dialect-specific way.

The following hierarchy of lenition degree from least to most across places of articulation was revealed by the follow-up tests: dental < bilabial < velar. The hierarchy is the opposite of the prediction by Kingston (2008) prediction but is consistent with his findings. Specifically, since the more posterior the stops' constriction, the greater the intraoral air pressure build up (Ohala, 1974; Javkin, 1977), Kingston (2008) predicted less lenition for more posterior stops relative to more anterior stops. Yet, his results "hinted" that the opposite was true and he reasoned that this may be "because velar closures are more often incomplete" (Kingston, 2008, footnote 20, p. 21). More lenition for bilabial than dental stops predicted by our regression model may also suggest that bilabial closures are more incomplete than dental closures. This hypothesis awaits further research for confirmation, however.

For continuant posterior probabilities, follow-up tests revealed effects of the following vowel openness in the

preceding nasal context, but not in the preceding vowel context. The effect is illustrated in Fig. 4, where close vowels suppress weakening of the target stops to a greater extent than mid and open vowels when they occur after a nasal, but not after a vowel. Negative effects of preceding nasals on lenition are well documented. In several African languages surveyed by Kingston (2008) and Steriade (1993), stops never lenited to fricatives after nasals. In addition, prenasal fricatives become affricates in many Bantu languages (Steriade, 1993). Furthermore, stop intrusion between a nasal and a fricative is commonly observed in English as in *warm[p]th*, *ten[t]th*, and *leng[k]th* (Kingston, 2008). According to Steriade (1993), difficulty in simultaneously executing the velum raising and the release of the oral occlusion gesture of the nasal accounts for both post-nasal hardening and intrusive oral stop between a nasal and a fricative. However, while inhibitory effects of preceding nasals on lenition are well documented, to our knowledge, the gradient effect of the following vowel height in the preceding nasal context but not in the preceding vowel context has not been previously reported. This new result suggests that difficulty in coordinating simultaneous timing between the two gestures varies as a function of the following vowel: the higher the vowel, the greater the difficulty. This hypothesis is consistent with the well-documented finding that open vowels are generally more nasalized than close vowels (Chen, 1997) and that nasal vowels are produced with lower and more centralized tongue position than their oral counterparts due to a greater degree of coupling between the oral and the nasal cavities during open compared to close vowels (e.g., Arai, 2004; Carignan, 2017).

In sum, our approach yielded lenition patterns that are largely consistent with previous findings using quantitative acoustic methods as well as new finer-grained patterns not previously reported. However, the validity of the approach needs to be further tested against other acoustic dimensions, more sets of data from different languages, as well as on different lenition phenomena. Specifically, future work should look beyond Argentinian Spanish, since it is possible that the results that deviate from those reported in previous studies might be due to dialectal differences. Moreover, it should be tested against unifying accounts of lenition that seek to capture all types of lenition and do not rely on notions of sonority or articulatory aperture, such as the work by Harris and Urua (2001) and Harris *et al.* (2023), which builds on the model of speech as a modulated carrier signal and models lenition as modulation reduction. In addition, at least for intervocalic targets, the approach could be further improved by replacing forced alignment with the automated segmentation method proposed by Ennever *et al.* (2017).

ACKNOWLEDGMENTS

This research was supported by NSF-National Science Foundation (SenSE) Grant No. 2037266.

¹The selection of the middle frame(s) was dependent on the parity of the total number of frames of the phone token. If the total number of frames was odd and divisible by three, we used the division value to select the number of middle frames; if the total number of frames was odd and indivisible by 3, we increased it to the closest odd number that is divisible by three. The same protocol was applied to the phone tokens with the total number of frames that was even, except for tokens with only one or two frames in which all of the frames were used to compute the average.

- Albert, A., and Nicenboim, B. (2022). "Modeling sonority in terms of pitch intelligibility with the nucleus attraction principle," *Cognitive Sci.* **46**(7), e13161.
- Arai, T. (2004). "Formant shift in nasalization of vowels," *J. Acoust. Soc. Am.* **115**(5), 2541.
- Backley, P. (2011). *Introduction to Element Theory* (Edinburgh University Press, Edinburgh).
- 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)* (University of Pennsylvania Working Papers in Linguistics, Philadelphia, PA) Vol. 22, pp. 10–20.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Statist. Softw.* **67**(1), 1–48.
- Bauer, L. (2008). "Lenition revisited," *J. Linguist.* **44**(3), 605–624.
- Bouavichith, D., and Davidson, L. (2013). "Acoustic characteristics of intervocalic stop lenition in American English," *J. Acoust. Soc. Am.* **133**(5) 3565.
- Broś, K., Żygis, M., Sikorski, A., and Wołfejko, J. (2021). "Phonological contrasts and gradient effects in ongoing lenition in the Spanish of Gran Canaria," *Phonology* **38**(1), 1–40.
- Carignan, C. (2017). "Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels," *J. Phon.* **63**, 87–105.
- Chen, M. Y. (1997). "Acoustic correlates of English and French nasalized vowels," *J. Acoust. Soc. Am.* **102**(4), 2360–2370.
- Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English* (Harper & Row, New York).
- Clements, G. N. (1990). *The Role of the Sonority Cycle in Core Syllabification* (Cambridge University Press, Cambridge), Vol. 1, pp. 283–333.
- 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* (The Karolinum Press, Prague), Vol. 2, pp. 575–589.
- 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.
- 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.
- 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, Marseille, France), pp. 6504–6513.
- Gurevich, N. (2011). *Lenition* (John Wiley & Sons, Ltd, Chester, UK), Vol. 3, Chap. 66.
- Harris, J. (2006). "The phonology of being understood: Further arguments against sonority," *Lingua* **116**(10), 1483–1494.
- Harris, J., and Urua, E.-A. (2001). "Lenition degrades information: Consonant allophony in Ibibio," *Speech Hear. Lang.* **13**, 72–105.

- Harris, J., Urua, E.-A., and Tang, K. (2023). "A unified model of lenition as modulation reduction: Gauging consonant strength in Ibibio," *Phonology* (to be published) PsyArXiv.
- Harris, J. W. (1969). *Spanish Phonology* (MIT Press, Cambridge, MA).
- Henke, E., Kaisse, E. M., and Wright, R. (2012). "Is the sonority sequencing principle an epiphenomenon?," in *The Sonority Controversy*, edited by S. Parker (De Gruyter Mouton, Berlin, Boston), pp. 65–100.
- Hualde, J. I. (2013). *Los Sonidos Del Español (The Sounds of Spanish): Spanish Language Edition* (Cambridge University Press, Cambridge).
- Hualde, J. I., Simonet, M., and Nadeu, M. (2012). "Consonant lenition and phonological recategorization," *Lab. Phonol.* 2(1), 301–329.
- Huang, X., Acero, A., Hon, H.-W., and Reddy, R. (2001). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development* (Prentice Hall PTR, Upper Saddle River, NJ).
- Hyman, L. (1975). *Phonology: Theory and Analysis* (Holt, Rinehart and Winston, New York), Vol. 10.
- Jakobson, R., Fant, C. G., and Halle, M. (1951). *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates* (MIT Press, Cambridge, MA).
- Javkin, H. (1977). "Towards a phonetic explanation for universal preferences in implosives and ejectives," *Proc. Annu. Meeting Berkeley Linguistics Soc.* 3, 557–565.
- Jespersen, O. (1899). *Fonetik: En Systematisk Fremstilling af Læren om Sproglyd (Phonetics. A Systematic Presentation of the Doctrine of Language Sound)* (Det Schubotheske Forlag, Copenhagen).
- Katz, J., and Pitzanti, G. (2019). "The phonetics and phonology of lenition: A Campidanese Sardinian case study," *Lab. Phonol.* 10(1), 16.
- Kawasaki-Fukumori, H., and Ohala, J. (1997). "Alternatives to the sonority hierarchy for explaining segmental sequential constraints," in *Language and its Ecology: Essays in Memory of Einar Haugen*, edited by S. Eliasson and E. Jahr (De Gruyter Mouton, Berlin, New York), pp. 343–365.
- Kendall, T., Vaughn, C., Farrington, C., Gunter, K., McLean, J., Tacata, C., and Arnsen, S. (2021). "Considering performance in the automated and manual coding of sociolinguistic variables: Lessons from variable (ing)," *Front. Artif. Intell.* 4, 648543.
- Kingma, D. P., and Ba, J. (2014). "Adam: A method for stochastic optimization," arxiv.org/abs/1412.6980.
- Kingston, J. (2008). "Lenition," in *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, edited by L. Colantoni and J. Steele (Cascadilla Press, Somerville, MA), pp. 1–31.
- Kirchner, R. (1998). "An effort based approach to consonant lenition," Ph.D. thesis, University of California, Los Angeles.
- Kirchner, R. (2013). *An Effort Based Approach to Consonant Lenition* (Routledge, New York).
- Ladefoged, P. (1993). *A Course in Phonetics* (Harcourt Brace Jovanovich College, Fort Worth, TX).
- Lahiri, A., and Reetz, H. (2002). "Underspecified recognition," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Werner (Mouton de Gruyter, Berlin), pp. 637–676.
- Lahiri, A., and Reetz, H. (2010). "Distinctive features: Phonological underspecification in representation and processing," *J. Phon.* 38(1), 44–59.
- Lavoie, L. M. (2001). *Consonant Strength: Phonological Patterns and Phonetic Manifestations* (Garland, New York).
- Lenth, R. V., Buerkner, 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, R package version 1.8.1-1, <https://CRAN.R-project.org/package=emmeans>.
- Magloughlin, L. (2018). "/t/and/d/ in North American English: Phonologization of a coarticulatory effect," Ph.D. thesis, University of Ottawa, Ottawa, Canada.
- Mansfield, J. B. (2015). "Consonant lenition as a sociophonetic variable in Murrinh Patha (Australia)," *Lang. Var. Change* 27(2), 203–225.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). "Montreal forced aligner: Trainable text-speech alignment using Kaldi," in *Proc. Interspeech 2017*, 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. (1974). "A mathematical model of speech aerodynamics," in *Proceedings of the Speech Communication Seminar*, Stockholm, Vol. 2, 65–72.
- 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* (virtual), pp. 13–17. festvox.org/cedar/WSTCSMC2020.pdf.
- Parker, S. G. (2002). *Quantifying the Sonority Hierarchy* (University of Massachusetts Amherst, Amherst, MA).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.* 12, 2825–2830.
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria), www.r-project.org/ (Last viewed December 1, 2022).
- Ségéral, P., and Scheer, T. (2008). *Positional Factors in Lenition and Fortition* (De Gruyter Mouton, Berlin and New York), pp. 131–172.
- Sheldon, E. (1893). *Grundzüge Der Phonetik Zur Einführung in Das Studium Der Lautlehre Der Indogermanischen Sprachen (Fundamentals of Phonetics for an Introduction to the Study of the Phonetics of the Indo-European Languages)*, 4th ed. (Breitkopf & Härtel, Leipzig).
- 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, Oregon, 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 (CPhS-14)*, San Francisco, California, 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 (Academic Press, San Diego, CA), pp. 401–470.
- Vásquez-Correa, J., Klumpp, P., Orozco-Aroyave, J. R., and Nöth, E. (2019). "Phonet: A tool based on gated recurrent neural networks to extract phonological posteriors from speech," in *Proc. 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), 6.
- Warner, N., and Tucker, B. V. (2011). "Phonetic variability of stops and flaps in spontaneous and careful speech," *J. Acoust. Soc. Am.* 130(3), 1606–1617.
- Whitney, W. D. (1865). "The relation of vowel and consonant," *J. Am. Oriental Soc.* 8, 357–373.
- Yuan, J., and Liberman, M. (2009). "Investigating /l/ variation in English through forced alignment," in *Proceedings of Interspeech 2009*, Brighton, UK, pp. 2215–2218.
- Yuan, J., and Liberman, M. (2011a). "Automatic detection of 'g-dropping' in American English using forced alignment," in *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*, Waikoloa, Hawaii, pp. 490–493.
- Yuan, J., and Liberman, M. (2011b). "/l/ variation in American English: A corpus approach," *J. Speech Sci.* 1(2), 35–46.