languages

MDPI

Article

Quantitative Acoustic versus Deep Learning Metrics of Lenition

Ratree Wayland [1,*,†], Kevin Tang [2,†], Fenqi Wang [1,†], Sophia Vellozzi [3] and Rahul Sengupta [3]

[1] Department of Linguistics, University of Florida, Gainesville, FL 32611-5454, USA
[2] Department of English Language and Linguistics, Institute of English and American Studies, Heinrich-Heine-University, 40225 Düsseldorf, Germany
[3] Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611-5454, USA
* Correspondence: ratree@ufl.edu
† These authors contributed equally to this work.

Abstract: Spanish voiced stops /b, d, g/ surfaced as fricatives [β, ð, ɣ] in intervocalic position due to a phonological process known as spirantization or, more broadly, lenition. However, conditioned by various factors such as stress, place of articulation, flanking vowel quality, and speaking rate, phonetic studies reveal a great deal of variation and gradience of these surface forms, ranging from fricative-like to approximant-like [β̞, ð̞, ɣ̞]. Several acoustic measurements have been used to quantify the degree of lenition, but none is standard. In this study, the posterior probabilities of sonorant and continuant phonological features in a corpus of Argentinian Spanish estimated by a deep learning Phonet model as measures of lenition were compared to traditional acoustic measurements of intensity, duration, and periodicity. When evaluated against known lenition factors: stress, place of articulation, surrounding vowel quality, word status, and speaking rate, the results show that sonorant and continuant posterior probabilities predict lenition patterns that are similar to those predicted by relative acoustic intensity measures and are in the direction expected by the effort-based view of lenition and previous findings. These results suggest that Phonet is a reliable alternative or additional approach to investigate the degree of lenition.

Keywords: Spanish; lenition; spirantization; phonet; deep learning; neural networks

Citation: Wayland, Ratree, Kevin Tang, Fenqi Wang, Sophia Vellozzi, and Rahul Sengupta. 2023. Quantitative Acoustic versus Deep Learning Metrics of Lenition. *Languages* 8: 98. https://doi.org/10.3390/languages8020098

Academic Editor: Adrian P. Simpson

Received: 23 October 2022
Revised: 9 March 2023
Accepted: 13 March 2023
Published: 29 March 2023

1. Introduction

Consonants are produced differently in different phonetic environments. For example, /b, d, g/, in most if not all Spanish dialects, are produced as voiced stops [b, d, g] after a pause, a homorganic nasal, and in the case of /d/, after a lateral /l/ (Navarro Tomás 1977; Hammond 2001; Hualde 2005), but as voiced fricatives (i.e., voiced continuants) [β, ð, ɣ] in other contexts. These include intervocalic and postvocalic syllable-onset positions, both within and across word boundaries. For example, [bas] 'you go', [das] 'you give', and [gas] 'gas', but [aβas] 'beans', [aðas] 'fairies', and [aɣas] 'you do (subjunctive)' (González 2002). The weakening of an underlying stop consonant to a voiced continuant demonstrated by these examples is commonly referred to as spirantization, one of the most widely studied phonological phenomena of Spanish. In turn, spirantization belongs to a broader class of a phonological process known as lenition, which also includes degemination, [tt → t]; deaspiration, [t^h] → [t]; voicing, [t] → [d]; flapping, [t, d] → [ɾ]; debucalization, [t] → [ʔ, h]; gliding, [t] → [j]; and deletion or loss, [ʔ, h, j] → [ø] (Gurevich 2011).

Despite having been traditionally described as having a continuant and a non-continuant realization in complementary distribution, phonetic studies revealed a more varied and gradient distribution of the surface, lenited forms of Spanish voiced stops. For instance, continuant realizations, previously characterized as fricatives (i.e., produced with turbulent airflow) (e.g., Navarro Tomás 1977; Harris 1969; Lozano 1978; Mascaró and Aronoff 1984), have been shown to be phonetically closer to approximants [β̞, ð̞, ɣ̞] (i.e., produced without

turbulent airflow) (e.g., Celdrán 1984; Romero 1995). Phonetic investigations also revealed large phonetic variability and gradience among continuant realizations conditioned by various factors, including surrounding vowel quality, stress, and speaking rate (e.g., Cole et al. 1999; Ortega-Llebaria 2004) suggesting a continuum rather than a fixed degree of constrictions across environments. Furthermore, the degree of constriction shows an effect on the place of articulation and surrounding vowel height (Ortega-Llebaria 2004; Carrasco et al. 2012; Simonet et al. 2012). Besides voiced stops, voiceless stops in Spanish also undergo lenition (e.g., Broś et al. 2021). The goal of this study is to compare lenition quantification metrics of lenited Argentinian Spanish stops consonants in intervocalic positions.

1.1. Acoustic Correlates of Lenition

The degree of lenition has been acoustically quantified along several acoustic dimensions, including intensity, duration, spectral (e.g., spectral peak, mean, standard deviation, and kurtosis), and periodic acoustic measures (e.g., harmonic-to-noise ratio), of which intensity, calculated as a difference or as a ratio, is the most prevalent (Cole et al. 1999; Ortega-Llebaria 2004; Soler and Romero 1999; Hualde et al. 2011). For example, Martínez-Celdrán and Regueira (2008), Figueroa and Evans (2015), and Broś et al. (2021) used intensity difference (preceding segment's maximum intensity minus minimum intensity of the target consonant) as a lenition marker. Similarly, Hualde et al. (2011) used the difference between the maximum intensity value during the vowel following the target consonant and the minimum value during the target consonant portion to quantify the degree of lenition. The more open the constriction of the target consonant (i.e., the more lenited the target consonant), the smaller the difference is expected to be.

The second intensity measure previously used to quantify the degree of lenition is the maximum rising velocity from the midpoint of the target consonant to the midpoint of the following vowel (Hualde et al. 2011, 2012; Kingston 2008). The more lenited the consonant is, the less abrupt the transition in intensity is and, thus, the smaller the maximum rising velocity value. Lastly, the mean intensity of the target sounds could also indicate their degree of lenition: the higher the mean intensity, the more advanced their degree of lenition. However, since this measure could vary greatly with speaking volume, it may not be as reliable as relative intensity measures.

Besides relative intensity, the relative duration of the target consonant (target sound duration divided by the total duration of the preceding sound + target sound + following sound) correlates negatively with the degree of consonant weakening and has been used as a reliable lenition marker (e.g., Dalcher 2008). This measurement is usually used when the target consonant occurs intervocalically but can be adapted to other contexts. For example, an alternative relative duration ratio was calculated by dividing the target-sound duration by the total duration of the preceding sound + target sound in Broś et al. (2021) because the segment following the target sound was not always a vowel in their data. The more lenited the target consonant, the shorter its duration is expected to be, thus the smaller the duration ratio.

The harmonics-to-noise ratio (HNR) is another lenition marker first employed by Broś et al. (2021). HNR is a measure of the proportion of acoustic periodicity (harmonic) to aperiodicity (noise) of a given sound and is expressed in decibels (dB). A HNR of 0 dB indicates equal energy in harmonics (periodicity) and noise. A positive HNR value indicates higher harmonic energy relative to noise energy, while a negative HNR value indicates higher noise energy relative to harmonic energy. For example, a HNR of 3 dB means that the harmonic energy is twice the noise energy ($10 \times \log^{10} (2/1)$), but a HNR of -3 dB HNR ($10 \times \log^{10} (1/2)$) indicates the opposite. Broś et al. (2021) reasoned that the more lenited a segment, the more vowel-like it is, hence the higher the HNR.

An alternative method of measuring the degree of periodicity, called 'Noise', was proposed by Harris et al. (Forthcoming), which builds on Harris and Urua (2001). Noise is computed using a three-step algorithm that combines measurements of amplitude and

aperiodicity within a VCX analysis frame. First, the algorithm computes the aperiodicity of the signal using an autocorrelation measure. Second, it locates the target consonant within the frame using the minimum amplitude. Third, it searches forward within the frame until it finds the point with the maximum product of aperiodicity and normalized amplitude values, which yields a Noise score. The higher the Noise score, the greater the degree of aperiodic energy in the signal. One of its advantages over the HNR measure is that Noise can investigate a wider set of lenition phenomena, such as the release properties of final stops, which does not exist in Spanish.

The goal of this study is to compare a new approach to quantify the degree of lenition known as 'Phonet' to the traditional acoustic markers such as those described above. Phonet is a deep learning model. Unlike the quantitative acoustic approach, where values along different acoustic dimensions are directly used to estimate lenition, in the Phonet model, the degree of lenition is estimated from the posterior probabilities of sonorant and continuant phonological features computed directly from the input signals by bidirectional recurrent neural networks (RNNs). Specifically, the approach projects gradient surface acoustic parameters onto two phonological features, continuant and sonorant, to capture both categorical and gradient realizations of lenition. Additionally, it is largely automatic and can be customized for a specific language investigated. The basis for this approach is outlined in the following sections.

1.2. Lenition and Phonological Features

Phonemes are classified into broad categories or classes based on their common phonological features (Jakobson et al. 1951; Chomsky and Halle 1968). The broadest class is [consonantal]. In most, if not all, languages, phonemes are either [+consonantal] or [−consonantal]. [+consonantal] sounds are produced with varying degrees of constriction of articulators in the vocal tract, while [−consonantal] phonemes are produced with no oral constriction. Stops, fricatives, affricates, nasals, and liquids belong to the first category of phonemes, while the second category is restricted to vowel and glide phonemes in most languages. [syllabic] is another main phonological feature in languages. [+syllabic] phonemes are the most sonorous segments of a language and are permitted to occupy the nucleus position of a syllable, whereas [−syllabic] phonemes are not. Vowels and syllabic consonants (ɹ̩, l̩, n̩, m̩ etc.) are [+syllabic] while consonants including glides are [−syllabic].

The second major class is [sonorant] and its inverse [obstruent]. [+sonorant] includes phonemes produced with little to no constriction in the oral cavity, hence: relatively free airflow and the ability to sustain resonance. Nasals, liquids, glides, and vowels are [+sonorant], while stops, fricatives, and affricates which are produced with complete or substantial airflow obstruction are [−sonorant] or [+obstruent].

The third phonological class relevant to our study is [continuant]. This feature describes the sustenance of airflow through the oral cavity. [+continuant] phonemes are produced with continual airflow through an incomplete closure between articulators. For example, fricatives are classified as [+continuant] because they are articulated with only partial oral occlusion, and airflow is permitted to flow continuously during their articulation. Other [+continuant] phonemes are liquids, glides, and vowels. For nasals, they are classified by some as [−continuant] because of airflow blockage through the oral cavity during their production, but as [+continuant] because continuous airflow is allowed through the nasal cavity. In this study, nasals are specified as [−continuant]. See Hayes (2008) for an introduction to other phonological features.

The distribution of phonological features allows us to group phonemes into natural classes (groups of phonemes that share one or a set of phonological features). Phonemes belonging to the same natural class pattern together when undergoing various phonological processes. For example, in English, /p, t, k/ are [−syllabic, −voice, −continuant, −sonorant, −delayed release] and form a natural class. They all become aspirated when they occur as an onset of a stressed syllable. [−delayed release] is characterized by an abrupt release of occluded airflow, while [+delayed release] is characterized by a grad-

ual release of airflow following the opening of an oral closure. In Spanish, /b, d, ɡ/ are [−syllabic, +voice, −continuant, −sonorant, −delayed release] and form a natural class. As discussed above, they undergo lenition and are realized as fricatives [+continuant] or approximants [+sonorant] in an intervocalic position, for example. In other words, the lenition of Spanish-voiced stops can be simplistically described as involving categorical changes of the [continuant] and the [sonorant] features. However, to capture the highly varied and gradient degree of lenition, it is necessary that we look beyond categorical manifestations of lenition changes and beyond the binary nature of phonological features.

1.3. Posterior Probability and Gradience

Computational approaches have been used in several studies of phonetic variations whose aim was to measure gradient variations. Many of these studies have relied on forced alignment systems to determine pronunciation variations (e.g., [dʒ]-[z] and [pʰ]-[f] variations in Hindi English code-mixed speech (Pandey et al. 2020), 'g'-dropping in English (Kendall et al. 2021; Yuan and Liberman 2011), 'th'-fronting, 't, d'-deletion, and 'h'-dropping in English (Bailey 2016). The forced alignment systems typically take word-level orthographic transcriptions as the input making reference to a pronunciation dictionary with phone-level transcription. Importantly, multiple pronunciations can be assigned to each word entry in the dictionary. For instance, to model 'th' fronting, two pronunciations, one with [θ] and one with [f], could be given to all word entries that may undergo 'th'-fronting. Based on each word token's acoustic properties, a trained forced aligner can automatically determine which of the two pronunciations has the highest probability. However, since a forced alignment model contains an acoustic model for each phone type defined in the pronunciation dictionary, the degree of variation could not be determined beyond the granularity of the phone set (e.g., as either [θ] or [f]).

An innovative method to obtain a more gradient measure of variations (e.g., degree of 'th'-fronting) as opposed to simply coding a token as [θ] or [f]) was proposed by Yuan and Liberman (2009) in their investigation of the degree of /l/-darkness in American English. In this study, instead of relying on phone labels outputted by the forced alignment procedure, probability scores extracted during the forced alignment procedure were used as a measure of variation. The probability score is defined as the log probability (log probability density) of the aligned segment to be a particular phone. More specifically, all /l/ tokens from a corpus of American English were forced aligned twice: first by a model trained on light /l/s (word-initial) and second by a model trained on dark /l/s (word-final and word-final consonant clusters), and degree of /l/-darkness was indicated by the difference between the log probability scores from the dark /l/ alignment and the light /l/ alignment. The method was extended to examine the finer variation of both types of /l/s by Yuan and Liberman (2011). Their results demonstrated the categorical distinction between dark (in syllable coda) and light /l/ (in syllable onset) while also revealing that intervocalic dark /l/ is less dark than canonical syllable-coda dark /l/, and its degree of darkness depends on the stress of the flanking vowels. Intervocalic light /l/ is always light and is lighter than canonical syllable-onset /l/. This method was also applied to investigate gradient variation of /t/-/d/ affrication in English by Magloughlin (2018). In this case, the degree of affrication is the log probability scores from the /tʃ, d/ alignment and the /tɹ, dɹ/ alignment using acoustic models of /tʃ/ and /dʒ/, and /t/ and /d/, respectively.

Besides acoustic models in a forced alignment system, the probability estimates from token classification can be obtained from other methods. For instance, in their investigation of the degree of r-lessness of postvocalic /r/ in English, McLarty et al. (2019) trained the Support Vector Machines (SVM) model to classify the canonical r-less tokens (oral vowels that are not preceding a liquid or a nasal) and the canonical r-full tokens (prevocalic /r/) using Mel-Frequency Cepstral Coefficients (MFCCs) as the acoustic representations. Once successfully trained (mean classification accuracy of 98.95%), the model was applied to ambiguous tokens (postvocalic /r/) to obtain a probability estimate of being r-less as opposed to r-full. A similar approach was used by Villarreal et al. (2020) in their

examination of two English sociophonetic variables (non-prevocalic /r/ and word-medial intervocalic /t/). However, instead of SVM, the random forest classification method was used to automate coding categorical manifestations of the two variables. Note that the classification method used by most of these studies is trained on surface segments that are not necessary surface realizations of the segment undergoing variation of interest. It simply relies on acoustic similarities between these surface segments and the possible canonical realizations of a variation. For instance, in the case of 'th'-fronting, the model was trained to classify tokens that are either canonically [θ] or canonically [f], and these canonical tokens themselves are not subjected to 'th'-fronting. However, their acoustic characteristics would capture the range of possible surface realizations of 'th'-fronting. In the case of /l/-darkening, canonical light /l/s, and dark /l/s are used in the training phase, and the trained model is then applied to /l/s that exhibit variable degrees of darkening.

The viability of this approach to estimate the categorical manifestation of lenition is demonstrated by the results of Cohen Priva and Gleason (2020). In this study, a range of processes commonly recognized as lenition was modeled using a spoken corpus of American English. Specifically, three types of modeling methods that differ in the underlying representation of the surface segments were examined. The first method compared the surface forms of two segment types (e.g., [t] and [d] for the lenition process /t/→ [d]) regardless of whether their underlying form was the segment in question (e.g., the [t] and [d] tokens do not need to share the underlying form /t/). The second method compared only the surface forms of two segment types that share the same underlying form (e.g., /t/ is the underlying form for both [t] and [d]). The third method compared only segments that surfaced unchanged (e.g., the [t] tokens realized from /t/ and the [d] tokens from /d/). Of significance is the finding that all three modeling approaches yielded the same results, suggesting that the various acoustic manifestations of a given lenition process (/t/→ [d] in this case) can be captured by comparing relevant pairs of surface segments, regardless of their underlying form.

The Phonet approach targets 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. Specifically, 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. A high [continuant] probability but a low [sonorant] probability would indicate a fricative-like realization, while a high [continuant] probability and a high [sonorant] probability would suggest an approximant-like realization of lenition. Unlike Yuan and Liberman (2009, 2011), where the degree of phonetic variation was estimated from the difference between the log probability scores of the two forced-alignment models (dark /l/ and light /l/), the degree of lenition is reflected in the probability of each phonological feature estimated from acoustic properties of the input signals.

1.4. Phonet

First 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). Inputs to Phonet are feature sequences based on log energy distributed across triangular Mel filters computed from 25 ms windowed frames of each 0.5 s chunk 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. Phonet has been found to be highly accurate in detecting phonemes 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).

In our study, twenty-three phonological classes of Spanish were trained by a bank of twenty-three 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 Equation (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 $i = \{1 \ldots C\}$ 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 use of MFCC-based acoustic features was motivated by how they are 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). In addition, MFCCs have been successfully used as acoustic representations in previous studies of phonetic variations (Kendall et al. 2021; Yuan and Liberman 2009, 2011; McLarty et al. 2019). A comparison between our approach using MFCCs and the quantitative acoustic approach using various acoustic dimensions would allow us to consider alternative acoustic representations to improve our model.

In sum, Phonet is a phonologically motivated, language-specific, and largely automatic model. It is trained to recognize input phones as belonging to different groups, defined by their phonological features. Once trained, posterior probabilities for different phonological features of the target segments are computed by the model. It relies on a phonological concept in phonological analyses of lenition and can capture both categorical and gradient surface manifestations of lenition. Input to the model is log energy distributed across triangular Mel filters computed from 25 ms windowed frames of each 0.5 s chunk of the input of the target language phones, thus using acoustic information for a given phonological feature of the target language. Phonological feature sets can be customized for a given target language with different assumptions of their underlying specifications (Lahiri and Reetz 2002; Lahiri and Reetz 2010) and physical correlates (Jakobson et al. 1951; Chomsky and Halle 1968; Backley 2011). Lastly, it only requires a phonological feature set and a segmentally-aligned acoustic corpus which can be obtained using forced alignment (see Ennever et al. 2017 for an automated segmentation method).

2. Methods

2.1. Materials

The Argentinian Spanish Corpus containing crowd-sourced recordings from 44 (31 female, 13 male) native speakers of Argentinian Spanish built by Guevara-Rukoz et al. (2020) was used in this study. 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). For the study, word tokens with voiced and voiceless stops, /b, d, g, p, t, k/, occurring between two vowels with different degrees of openness, were selected. Table 1 specifies the number of word tokens and word types by conditions—voicing (voiced or voiceless), place of articulation (bilabial, dental, and velar), and preceding and following vowels (open, mid, and close).

Table 1. Word distribution by conditions –voicing, place of articulation, preceding vowel, and following vowel. The number left and right of the slash in each cell represents the number of word tokens and word types respectively.

		Voiced			Voiceless		
		Following Vowel					
Place	**Preceding Vowel**	**Close**	**Mid**	**Open**	**Close**	**Mid**	**Open**
Bilabial	Close	19/1	3/0	8/1	7/1	0/0	6/1
	Mid	134/28	216/30	142/23	183/26	435/44	228/21
	Open	128/18	90/18	103/8	69/10	316/28	248/24
Dental	Close	0/0	20/3	0/0	0/0	26/1	38/2
	Mid	143/32	409/34	43/8	185/13	451/36	97/8
	Open	51/8	388/20	11/2	92/10	141/16	107/7
Velar	Close	0/0	0/0	0/0	4/0	40/4	21/1
	Mid	0/0	2/0	0/0	190/25	912/50	215/40
	Open	0/0	32/1	0/0	42/12	5776/40	307/32

2.2. Phonet Training Procedure

The Montreal Forced Aligner (version 2.0) (McAuliffe et al. 2017) was performed on the corpus. Based on Hualde (2013)'s grapheme-to-phoneme mapping in IPA, a phonemic pronunciation dictionary for the transcription of the corpus words was generated and used to train new acoustic models for the corpus and align the textgrids to the acoustic signals. A tri-phone acoustic model in which the left and the right contexts of the target phone are used to adjust its alignment during the alignment procedure was used. The phone set parameter was set to IPA, which enabled extra decision tree modeling based on the specified phone set. All parameters were kept as the default. The corpus was randomly split into a training subset (80%) and a test subset (20%) using the Python (Version 3.9) scikit-learn library (Pedregosa et al. 2011). Since the surface realizations of the targets /b, d, ɡ/, but not the targets /p, t, k/ (Colantoni and Marinescu 2010), were expected to be ambiguous realizations of the two features of interest: [continuant] and [sonorant], they were not included (i.e., silenced out) during training to avoid model contamination by the ambiguous tokens. In total, twenty-three 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 twenty-three different Phonet models. Like Vásquez-Correa et al. (2019), one addition model was included to train the phonemes. However, in addition to the 18 phonemes from Vásquez-Correa et al. (2019), 8 additional phonemes, including stressed /ˈa, ˈe, ˈi, ˈo, ˈu/, /ɲ/, /θ/ and /spn/ for speech-like noise were also included. As previously discussed, weakened realizations of Spanish /b, d, ɡ/ are either a fricative or an approximant (e.g., Simonet et al. 2012); therefore, [sonorant] and [continuant] are our features of interest. Model training was performed on the NVIDIA GeForce RTX 3090 GPU. The model was highly accurate in showing unweighted average recall (UAR) ranges from 94–98% across the different phonological classes. The sonorant and continuant features' UARS were 97% and 96%, respectively, suggesting a good model fit for both features. The model was then applied to our target word tokens with intervocalic voiced and voiceless stops, /b, d, ɡ, p, t, k/. The predictions were computed for 10 ms frames. The average of the middle frame(s) was used as the prediction for phone tokens containing multiple frames. Thus, a sonorant posterior probability and a continuant posterior probability were obtained for each target stop.

2.3. Acoustic Parameters: HNR, Duration and Intensity

In order to compare our model to the quantitative acoustic approach, five common acoustic parameters covering three broad acoustic dimensions of lenitions were selected for comparisons. Harmonic-to-noise ratio (HNR), relative duration, intensity difference (two

types), and mean intensity were extracted from the target intervocalic voiced and voiceless stops, /b, d, g, p, t, k/.

HNR quantifies degree of harmonicity relative to noise in the sound. The higher the HNR, the more periodic or vowel-like the sound is. Therefore, the more lenited a target stop is, the higher the HNR value is expected. HNR was calculated as ten times the $\log^{10}$ ratio between the energy of harmonicity and noise. The mean HNR of the target segments was computed in Praat and was defined as (2). In the algorithm, t_1 and t_2 represent the starting point, and the ending point of the token, respectively, and $x(t)$ is the harmonicity (in dB) as a function of time.

$$1/(t_2 - t_1) \int_{t_1}^{t_2} dt\ x(t) \tag{2}$$

Relative duration of each target stop was obtained by taking the duration of a target stop and divided it by the total duration of the preceding vowel + target consonant + following vowel. The duration of the segmental tokens was generated during the forced alignment (Section 2.2). The more lenited the consonant is, the shorter the relative duration.

Two intensity difference values were calculated for each target stop by subtracting minimum intensity of the target segment from the maximum intensity of (a) the preceding vowel and (b) the following vowel. The assumption is that the smaller the intensity difference between the sound in question and the flanking vowels, the less constricted and hence the more lenited it is. The maximum intensity values of the preceding and following vowels and the minimum intensity value of the target segment were calculated using the parabolic interpolation method in Praat.

Finally, mean intensity captures average intensity of the target stops. The more lenited they are, the higher their mean intensity. The mean intensity values of the target segment were calculated in Praat with a definition as (3), where $x(t)$ is the intensity (in dB) as a function of time.

$$1/(t_2 - t_1) \int_{t_1}^{t_2} x(t)\ dt \tag{3}$$

2.4. Analyses

Values of the five acoustic parameters described above and the sonorant and continuant posterior probabilities generated by the Phonet model served as dependent variables in the linear mixed-effects regression models. The models' fixed variables were stress (stressed or unstressed), voicing (voiced or voiceless), place of articulation (bilabial, dental, and velar), preceding vowel height/openness (open, mid, and close), following vowel height (open, mid, and close), speaking rate [number of syllables in a word/word duration (in seconds)], and word status (content or function). Speaking rate and word status were included as they are known to influence lenition. Crucially, a higher degree of lenition is expected for a faster speaking rate relative to a slower speaking rate and for function words compared to content words (Broś et al. 2021; Soler and Romero 1999; Honeybone 2012). Similarly, a strong effect of stress on lenition has been reported, with a higher degree of lenition expected in unstressed syllables than in stressed syllables (Ortega-Llebaria 2004; Broś et al. 2021; Eddington 2011). On the contrary, the influence of place of articulation and flanking vowel openness has been inconsistent (Cole et al. 1999; Ortega-Llebaria 2004; Kingston 2008; Lewis 2001; Lavoie 2001). Overall, velar stops are expected to be weaker than labial and dental/alveolar stops, and the more open the flanking vowels, the greater the degree of lenition is expected. Regarding the effect of voicing, voiced stops are expected to be more lenited than voiceless stops (Broś et al. 2021; Colantoni and Marinescu 2010).

Deviation coding was used for the categorical variables stress, voicing, and word status, while forward difference coding was used for the variable's place of articulation (bilabial > dental > velar), preceding vowel (close > mid > open), and following vowel (close > mid > open). The models were performed using the lmer function from the lme4 package (Bates et al. 2015) in R Core Team (2022). After comparing multiple model structures with maximum likelihood, the best-fit model structure for each variable was identified.

Seven regression models were fitted with each of the five acoustic parameters and the two deep-learning-based features (the sonorant and the continuant phonological features) as the dependent variable. All models included different interaction terms but same random intercepts by speaker and word. The general formula of the model with three interaction terms is provided as follows:

DEPENDENT VARIABLES ~ Stress + Voicing + Place of articulation + Preceding vowel + Following vowel + Speaking rate + Word status + Place of articulation: Preceding vowel + Place of articulation: Following vowel + Preceding vowel: 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). Results of the best-fit model for each dependent variable are reported in the next section.

3. Results

3.1. Mean HNR

The best regression model for HNR yielded marginal R^2 and conditional R^2 values of 0.575 and 0.657, respectively, suggesting that the fixed factors in the model explained 57.5% of the total variance while 65.7% of the variance is explained when all factors are included.

The results reveal the significant role of voicing, place of articulation, flanking vowel height, and speaking rate. As shown in Table 2, the model yielded significant main effects of voicing: higher mean HNR for voiced stops than voiceless stops [$\beta = 10.433$, $t = 54.976$, $p < 0.001$]; place of articulation: bilabial < dental [$\beta = -1.232$, $t = -3.129$, $p = 0.002$], no significant difference between dental and velar; preceding vowel: close > mid > open [βs = 1.164, 1.004; *t*s = 3.347, 7.822; *p*s = 0.001, <0.001]; following vowel: close > mid > open [βs = 1.430, 0.595; *t*s = 3.312, 1.989; *p*s = 0.001, 0.047]; and speaking rate: the higher the speaking rate, the higher the mean HNR [$\beta = 0.124$, $t = 5.063$, $p < 0.001$]. These results suggest that, overall, voiced stops are more lenited than voiceless stops [e.g., /b/ in *de* ***b****o*ˈ*leto* vs. /p/ in *lo* ***p****o*ˈ*dés*], dental stops are more lenited than bilabial stops [e.g., /d/ in *de*ˈ***d****atos* vs. /b/ in *mi* ˈ***b****ase*], the less open the flanking vowels are, the more lenited the stops are (e.g., /b/ in *tu* ˈ*vida* vs. /b/ in *habla* ˈ*varios*], and the faster the speaking rate, the higher the degree of lenition based on HNRs.

Significant interactions are also found between the place of articulation and the following vowel; between the place of articulation and the preceding vowel; and between the preceding vowel and the following vowel. For the place of articulation x preceding vowel interaction (see Figure 1a), post hoc pair-wise comparisons using the Tukey method indicate no significant difference in HNR values of velar stops when preceded by different vowels. However, for bilabial and dental stops, their mean HNR values are significantly higher when they occur after close vowels than after open vowels [βs = 2.330, 3.340; *t*s = 3.257, 5.338; $p = 0.031$, <0.001, for bilabial and dental stops, respectively]. In addition, mean HNR values for bilabial stops are significantly lower than those of dental stops when they are preceded by mid vowels [$\beta = -1.014$, $t = -4.339$, $p = 0.0006$]. Interestingly, HNRs for both bilabial and dental stops are significantly lower than those of velar stops when preceded by open vowels [βs = −1.806, −0.970; *t*s = −6.535, −3.302; *p*s < 0.001, =0.027 for bilabial and dental stops, respectively]. These results indicate that HNRs are affected by different vowels for different stops: for bilabial and dental stops, HNRs increase after close vowels, while those for velar stops increase after low vowels.

Table 2. Summaries of Mean HNR: The fixed-effects in the linear mixed-effects model (a: upper), and the type-III-ANOVA analysis (b: lower).

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	10.951	0.364	30.115	**<0.001**
Stress (unstressed)	0.329	0.174	1.886	0.059
Voicing (voiced)	10.433	0.190	54.976	**<0.001**
Place (bilabial)	−1.232	0.394	−3.129	**0.002**
Place (dental)	0.280	0.327	0.854	0.393
Preceding vowel (close)	1.164	0.348	3.347	**0.001**
Preceding vowel (mid)	1.004	0.128	7.822	**<0.001**
Following vowel (close)	1.430	0.432	3.312	**0.001**
Following vowel (mid)	0.595	0.299	1.989	**0.047**
Speaking rate	0.124	0.024	5.063	**<0.001**
Word status (function)	0.411	0.279	1.476	0.140
Place (bilabial): Following vowel (close)	−0.011	0.446	−0.025	0.980
Place (dental): Following vowel (close)	−1.867	0.526	−3.546	**<0.001**
Place (bilabial): Following vowel (mid)	1.492	0.509	2.930	**0.003**
Place (dental): Following vowel (mid)	−1.655	0.518	−3.195	**0.001**
Preceding vowel (close): Following vowel (close)	2.687	1.185	2.268	**0.023**
Preceding vowel (mid): Following vowel (close)	1.118	0.313	3.567	**<0.001**
Preceding vowel (close): Following vowel (mid)	−0.897	0.737	−1.217	0.224
Preceding vowel (mid): Following vowel (mid)	−0.190	0.279	−0.681	0.496
Place (bilabial): Preceding vowel (close)	−0.831	1.067	−0.779	0.436
Place (dental): Preceding vowel (close)	1.256	0.729	1.724	0.085
Place (bilabial): Preceding vowel (mid)	−0.178	0.281	−0.635	0.526
Place (dental): Preceding vowel (mid)	1.247	0.266	4.692	**<0.001**
Effects	***SSE***	***MSE***	***F***	***p***
Stress	48	48	3.558	0.060
Voicing	40,685	40,685	3022.338	**<0.001**
Place	135	68	5.016	0.007
Preceding vowel	1150	575	42.704	**<0.001**
Following vowel	314	157	11.648	**<0.001**
Speaking rate	345	345	25.637	**<0.001**
Word status	29	29	2.179	0.142
Place: Following vowel	464	116	8.609	**<0.001**
Preceding vowel: Following vowel	265	66	4.921	**<0.001**
Place: Preceding vowel	442	111	8.218	**<0.001**

Note: significant *p*s < 0.05 are in bold.

For the place of articulation × following vowel interaction (see Figure 1b), post hoc analyses suggest that mean HNR values for bilabial stops are significantly higher when they are followed by close than by open vowels [β = 1.838, t = 3.969, p = 0.003]. For velar stops, HNR values are significantly higher when they are followed by close than by both mid and open vowels [βs = 2.677, 3.878; ts = 4.828, 6.935; ps < 0.001] for mid and open vowels, respectively]. On the other hand, no significant difference across different vowel types was found for dental stops. When HNR values of the three places of articulation were compared for each vowel context, it was found that HNR values for bilabial stops were significantly lower than those for velar stops [β = −2.258, t = −4.156, p = 0.001] when they are followed by close vowels. When followed by open vowels, HNRs for bilabial and velar stops are significantly lower than those of dental stops [βs = −2.223, 2.005; ts = −4.196, 4.062; ps = 0.001, 0.002 for bilabial and velar stops, respectively]. These results suggest that the following close vowels have a boosting effect on HNR values of all stops while the following open vowels depress HNRs, particularly for bilabial and velar stops. For the preceding vowel × following vowel interaction (see Figure 1c), results of the post hoc analyses suggested that when the following vowels are close, mean HNR values were significantly higher when the preceding vowels are also close or mid than when they are open [βs = 4.342, 1.686; ts = 4.711, 6.120; p = 0.0001, <0.001 for preceding close and mid vowels, respectively]. Similarly, when the following vowels are mid, HNRs were significantly higher when the preceding vowels are also mid than when they are open [β = 0.568, t = 3.969, p = 0.002]. When the following vowels are open, HNRs were significantly higher in the context of the preceding mid vowels than the preceding low

vowels [β = 0.758, t = 3.208, p = 0.037]. Similarly, when the preceding vowels are close, HNRs were significantly higher when the following vowels are also close than when they are open [β = 3.527, t = 3.180, p = 0.040]. When the preceding vowels are mid, HNRs were significantly higher when the following vowels are close than when they are mid or open [βs = 0.907, 1.738; ts = 3.839, 6.300; p = 0.004, <0.001 for mid and open vowels, respectively]. When the preceding vowels are open, HNRs were higher when the following vowels are mid than when they are open [β = 1.021, t = 3.943, p = 0.003]. Overall, these results suggested that HNRs increase when followed by relatively less open vowels than by relatively more open vowels. In addition, except for following open vowels, this boosting effect was more likely to occur when the flanking vowels are of the same or different in degree of openness by no more than one step from each other.

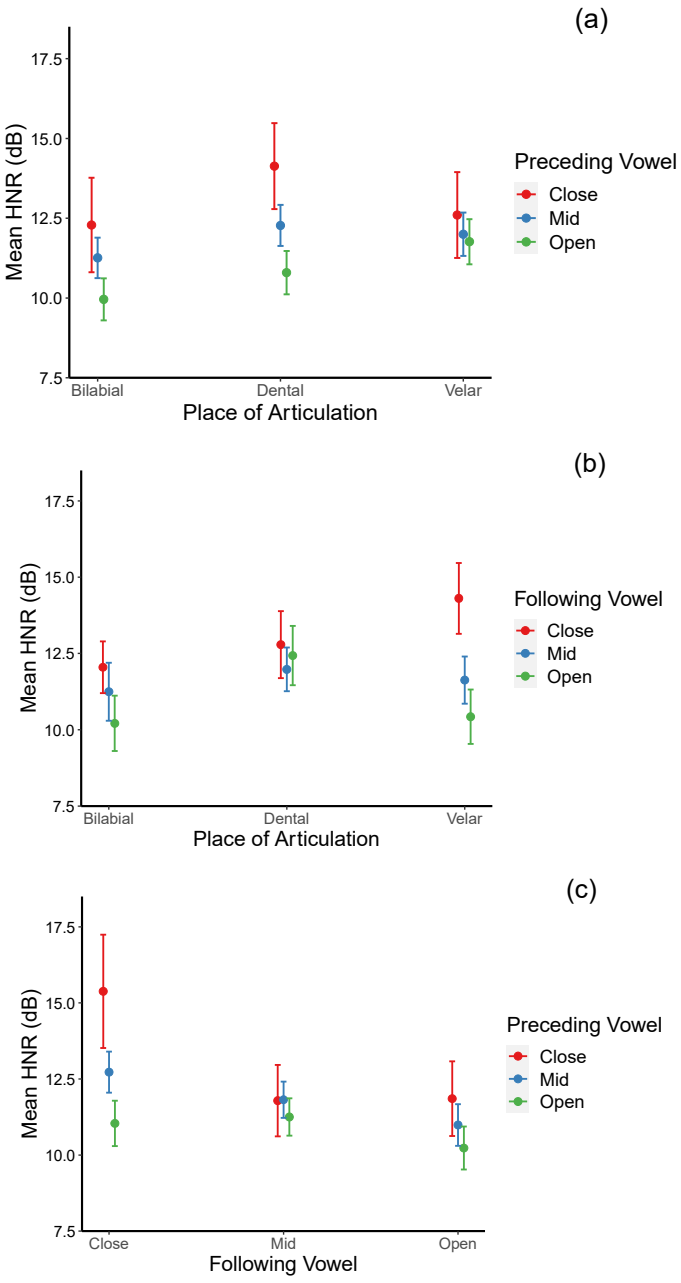

Figure 1. Estimated marginal means of mean HNR by place of articulation and preceding vowel (**a**, upper), by place of articulation and following vowel (**b**, middle), and by preceding and following vowel (**c**, lower). The dots represent the estimated marginal means, and the interval lines display 95% confidence interval.

3.2. Relative Duration

The best regression model for the relative duration (duration of the target sound/sum of the duration of the preceding sound, the target sound, and the following sound) showed that only 8.7% of the total variance was explained by the fixed factors (marginal $R^2 = 0.087$), but approximately 21% of the variance was additionally explained by the full mode (Conditional $R^2 = 0.301$).

Like HNR, the results reveal the significant role of voicing, place of articulation, flanking vowel height, and speaking rate. As shown in Table 3, the model revealed significant main effects of voicing: higher duration ratio for voiceless stops [$\beta = -0.045$, $t = -9.131$, $p < 0.001$]; place of articulation: bilabial > dental [$\beta = 0.039$, $t = 4.314$, $p < 0.001$], dental < velar [$\beta = -0.020$, $t = -2.471$, $p = 0.013$]; preceding vowel: close < mid [$\beta = -0.039$, $t = -5.30$, $p < 0.001$], no significant difference between mid and open vowels; following vowel: close < mid [$\beta = -0.030$, $t = -3.134$, $p = 0.002$], mid > open [$\beta = 0.027$, $t = 3.787$, $p < 0.001$]; and speaking rate: the higher the speaking rate, the lower the duration ratio [$\beta = -0.004$, $t = 7.381$, $p < 0.001$]. These results indicated that voiced stops are more lenited (shorter relative duration) than voiceless stops; dental stops are more lenited than bilabial and velar stops; a higher degree of lenition occurs when the preceding vowels are close relative to mid and when the following vowels are close or open relative to mid. In addition, lenition degree positively correlates with speaking rate. Significant interactions between the place of articulation and the following vowel, between the place of articulation and the preceding vowel, and between the preceding vowel and the following vowel were also found.

Table 3. Summaries of relative duration: The fixed-effects in the linear mixed-effects model (a: upper), and the type-III-ANOVA analysis (b: lower).

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	0.403	0.007	58.745	**<0.001**
Stress (unstressed)	−0.001	0.005	−0.143	0.886
Voicing (voiced)	−0.045	0.005	−9.131	**<0.001**
Place (bilabial)	0.039	0.009	4.314	**<0.001**
Place (dental)	−0.020	0.008	−2.471	**0.013**
Preceding vowel (close)	−0.039	0.007	−5.307	**<0.001**
Preceding vowel (mid)	0.004	0.003	1.293	0.196
Following vowel (close)	−0.030	0.010	−3.134	**0.002**
Following vowel (mid)	0.027	0.007	3.787	**<0.001**
Speaking rate	−0.004	0.001	−7.381	**<0.001**
Word status (function)	0.004	0.008	0.489	0.625
Place (bilabial): Following vowel (close)	0.044	0.012	3.773	**<0.001**
Place (dental): Following vowel (close)	−0.041	0.014	−2.999	**0.003**
Place (bilabial): Following vowel (mid)	−0.039	0.013	−2.902	**0.004**
Place (dental): Following vowel (mid)	0.027	0.014	2.015	**0.044**
Preceding vowel (close): Following vowel (close)	−0.054	0.025	−2.155	**0.031**
Preceding vowel (mid): Following vowel (close)	−0.004	0.007	−0.594	0.553
Preceding vowel (close): Following vowel (mid)	−0.001	0.016	−0.058	0.954
Preceding vowel (mid): Following vowel (mid)	−0.017	0.006	−2.746	**0.006**
Place (bilabial): Preceding vowel (close)	0.029	0.023	1.282	0.200
Place (dental): Preceding vowel (close)	−0.028	0.015	−1.814	0.070
Place (bilabial): Preceding vowel (mid)	0.003	0.006	0.561	0.575
Place (dental): Preceding vowel (mid)	0.021	0.006	3.789	**<0.001**
Effects	***SSE***	***MSE***	***F***	***p***
Stress	0.000	0.000	0.020	0.886
Voicing	0.458	0.458	83.376	**<0.001**
Place	0.104	0.052	9.507	**<0.001**
Preceding vowel	0.155	0.078	14.109	**<0.001**
Following vowel	0.098	0.049	8.948	**<0.001**
Speaking rate	0.299	0.299	54.484	**<0.001**
Word status	0.001	0.001	0.239	0.625
Place: Following vowel	0.107	0.027	4.849	**<0.001**
Preceding vowel: Following vowel	0.094	0.023	4.271	**0.002**
Place: Preceding vowel	0.129	0.032	5.869	**<0.001**

Note: significant *p*s < 0.05 are in bold.

For the place of articulation × preceding vowel interaction (see Figure 2a), relative durations of dental stops were found to be significantly shorter (more lenited) when preceded by close than by mid and open vowels [βs = −0.058, 0.049; ts = −4.342, −3.678; ps = 0.0005, 0.007, respectively). Similarly, relative durations of velar stops were significantly shorter when the preceding vowels were close than when they were open [β = −0.042, t = −3.287, p = 0.028]. Across place of articulation comparison for each vowel context revealed that when preceded by close and open vowels, relative durations of bilabial stops are significantly higher (less lenited) than dental stops [βs = 0.055, 0.050; ts = 4.678, 3.814; ps = 0.0001, 0.005 for close and open vowels, respectively]. Dental stops' relative durations are also shorter than those of velar stops when preceded by open vowels [β = −0.025, t = −3.374, p = 0.022]. These results suggested that dental stops are more lenited than bilabial and velar stops, particularly when they occur after open vowels.

For the place of articulation × following vowel interaction (see Figure 2b), post hoc pair-wise comparison indicated that relative durations of dental stops were significantly shorter when the following vowels are close vowels than when they are mid vowels [β = −0.058, t = −4.870, p < 0.0001), suggesting that they are more lenited in the following close vowel than in the following mid vowel context. No effects of vowel openness for bilabial and velar stops were indicated. When the three places of articulation were compared for each vowel context, the results suggested that relative durations of the dental stops were significantly shorter than those of bilabial stops in the close and open vowel contexts, suggesting that dental stops are, more lenited than bilabial stops when they are followed by these vowels [βs = −0.055, 0.050; ts = −4.678, −3.814, ps = 0.0001, 0.005 for close and open following vowel context, respectively]. The differences between dental and velar stops did not reach significance in any of the following vowel contexts.

For the preceding × following vowel interaction (see Figure 2c), post hoc analyses indicated that when followed by close vowels, relative durations are significantly shorter when the preceding vowels are close than when they are mid or open [βs = −0.076, −0.081; ts = −3.918, −4.100; ps = 0.003, 0.002 for preceding mid and open vowels, respectively]. However, the effects of preceding vowel openness were not significant when the following vowels were either mid or open. Regarding the effects of the following vowels on the preceding vowels, it was found that when the mid vowels precede, relative durations are significantly shorter when the following vowels are open than when they are mid [β = −0.022, t = −3.658, p = 0.008]. However, when preceded by open vowels, relative durations are significantly shorter when the following vowels are also open rather than close or mid [βs = −0.029, −0.038; ts = −3.622, −5.994; ps = 0.009, <0.001 for following close and open vowels, respectively]. No effects of the following vowels were found when the preceding vowels were close. Together, these results suggest that lenition is the most advanced when flanking vowels are close and that the lenition degree is stronger with either preceding or following open vowels relative to mid vowels.

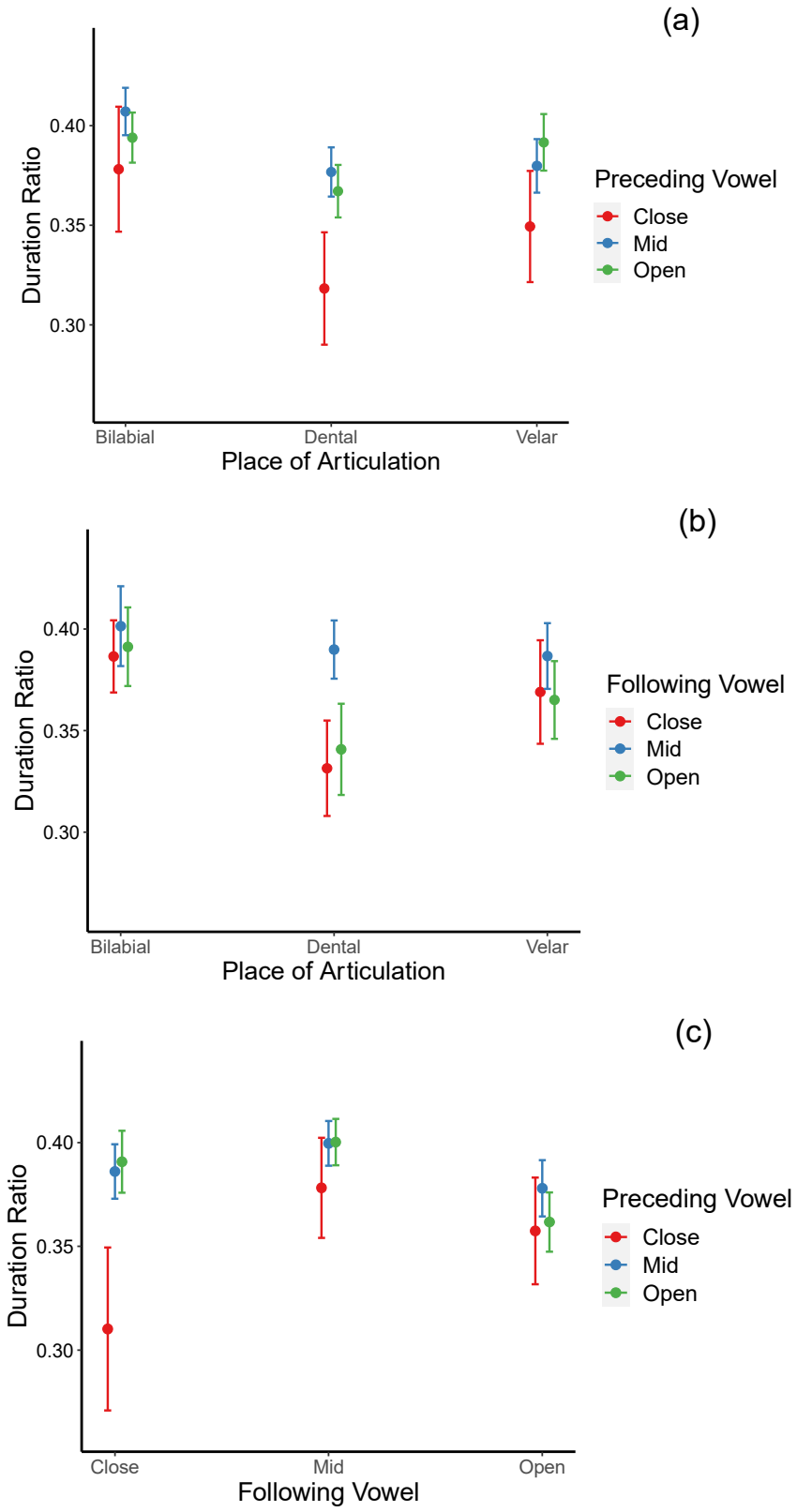

Figure 2. Estimated marginal means of relative duration by place of articulation and preceding vowel (**a**, upper), by place of articulation and following vowel (**b**, middle), and by preceding and following vowel (**c**, lower). The dots represent the estimated marginal means, and the interval lines display 95% confidence interval.

3.3. Intensity Difference Relative to the Preceding Vowel

The more lenited the consonants are, the lower the intensity difference between the target stops and the preceding vowel is expected. For this dependent variable, the best regression model yielded marginal R^2 and conditional R^2 values of 0.744 and 0.806, suggesting that 74.4% of the total variance of intensity difference (relative to the preceding vowel) was accounted for by the model's fixed factors and that 80.6% of the total variance was explained by the full model. Unlike HNR and relative duration, stress along with voicing, place of articulation, surrounding vowel height, and speaking rate emerged as significant predictors of lenition degree for this variable. As shown in Table 4, the model's significant main effects included the effects of stress: higher intensity difference for stressed syllables than unstressed syllables [$\beta = -2.231$, $t = -7.556$, $p < 0.001$]; voicing: higher intensity difference for voiceless stops than voiced stops [$\beta = -23.623$, $t = -72.992$, $p < 0.001$]; place: bilabial > dental > velar [βs = 1.762, 1.409; ts = 2.872, 2.664; ps = 0.004, 0.008]; preceding vowel: open > mid > close [βs = −2.092, −1.456; ts = −4.066, −7.475; ps < 0.001]; following vowel: no significant difference between close and mid, mid > open [$\beta = 1.900$, $t = 3.974$, $p < 0.001$]; and speaking rate: the higher the speaking rate, the lower the intensity difference [$\beta = -0.862$, $t = -23.499$, $p < 0.001$].

Table 4. Summaries of intensity difference relative to the preceding vowel: The fixed-effects in the linear mixed-effects model (a: upper), and the type-III-ANOVA analysis (b: lower).

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	20.788	0.532	39.085	**<0.001**
Stress (unstressed)	−2.231	0.295	−7.556	**<0.001**
Voicing (voiced)	−23.623	0.324	−72.992	**<0.001**
Place (bilabial)	1.762	0.613	2.872	**0.004**
Place (dental)	1.409	0.529	2.664	**0.008**
Preceding vowel (close)	−2.092	0.515	−4.066	**<0.001**
Preceding vowel (mid)	−1.456	0.195	−7.475	**<0.001**
Following vowel (close)	−0.295	0.661	−0.446	0.655
Following vowel (mid)	1.900	0.478	3.974	**<0.001**
Speaking rate	−0.862	0.037	−23.499	**<0.001**
Word status (function)	0.096	0.505	0.191	0.849
Place (bilabial): Following vowel (close)	0.941	0.757	1.243	0.214
Place (dental): Following vowel (close)	−2.708	0.891	−3.039	**0.002**
Place (bilabial): Following vowel (mid)	−1.053	0.875	−1.203	0.229
Place (dental): Following vowel (mid)	2.495	0.890	2.804	**0.005**
Preceding vowel (close): Following vowel (close)	−4.951	1.754	−2.822	**0.005**
Preceding vowel (mid): Following vowel (close)	−0.166	0.475	−0.349	0.727
Preceding vowel (close): Following vowel (mid)	0.787	1.089	0.722	0.470
Preceding vowel (mid): Following vowel (mid)	−0.513	0.423	−1.214	0.225
Place (bilabial): Preceding vowel (close)	2.235	1.593	1.403	0.161
Place (dental): Preceding vowel (close)	0.686	1.077	0.637	0.524
Place (bilabial): Preceding vowel (mid)	−0.371	0.423	−0.878	0.380
Place (dental): Preceding vowel (mid)	0.987	0.393	2.510	**0.012**
Effects	***SSE***	***MSE***	***F***	***p***
Stress	1558	1558	57.086	**<0.001**
Voicing	145,380	145,380	5327.831	**<0.001**
Place	739	369	13.541	**<0.001**
Preceding vowel	2398	1199	43.943	**<0.001**
Following vowel	464	232	8.498	**<0.001**
Speaking rate	15,069	15,069	552.225	**<0.001**
Word status	1	1	0.036	0.849
Place: Following vowel	375	94	3.438	**0.009**
Preceding vowel: Following vowel	284	71	2.603	**0.034**
Place: Preceding vowel	306	76	2.802	**0.024**

Note: significant ps < 0.05 are in bold.

These results suggested that stops weakened to a significantly higher degree when they are voiced than when they are voiceless, and when the syllables are unstressed than when they are stressed (e.g., /b/ in *que* ˈ***b***ueno vs. /b/ in *para* ***b***usˈcar). In addition, velar stops are lenited to a greater degree than dental stops, and, in turn, dental stops are more

lenited than bilabial stops. Additionally, both preceding and following vowels affected the degree of lenition. Specifically, lenition appears to be negatively correlated with the degree of openness of the preceding vowels: the less open the preceding vowels, the more weakened the stops are. However, the following vowel environments exert the opposite pattern of effects on stop lenition, and the difference reached significance only between the mid and the open vowels. More specifically, unlike preceding vowel contexts, the lenition degree is stronger when the following vowels are open than mid. Finally, as expected, the lenition degree increases when the speaking rate increases.

Significant interactions between the place of articulation and the preceding vowel (see Figure 3a), between the place of articulation and the following vowel (see Figure 3b), and between the preceding vowel and the following vowel (see Figure 3c) were also found.

For the place of articulation × preceding vowel interaction, smaller intensity differences are observed in the preceding close vowel context for stops at all three places of articulation. However, the effects are significant for dental and velar stops only. Specifically, post hoc pair-wise comparisons suggested that intensity difference was significantly smaller for dental consonants in the close vowel context relative to the open vowel context [$\beta = -3.612$, $t = -3.912$, $p = 0.003$], and for velar consonants in the close vowel than in both the mid and open vowel contexts [βs = −3.294, −5.285; ts = −3.630, −5.900; ps = 0.009, <0.0001, respectively]. For each preceding vowel context, the results revealed that velar consonants are more lenited than bilabial consonants in all three vowel contexts [βs = −5.323, −2.402, −1.786; ts = −3.440, −5.894, −3.958; ps = 0.017, <0.001, = 0.003 for the close, mid, and open vowel contexts, respectively) and dental consonants in the mid vowel context [$\beta = -1.509$, $t = -3.407$, $p = 0.020$]. Together, these results suggested that velar stops are more lenited than dental and bilabial stops and that preceding vowels with a lesser degree of openness induce a higher degree of weakening.

A different pattern emerges for the effects of the following vowels: relatively more open rather than more close vowels appear to trigger a higher degree of weakening, particularly for the velar stops. Specifically, post hoc pair-wise comparisons suggested that velar stops are more lenited than bilabial stops when followed by mid and open vowels [βs = −4.240, −2.798; ts = −6.012, 3.710; ps = 0.0001, 0.007, respectively]. However, no effects on the degree of openness of the following vowel within each place of articulation.

For the preceding × following vowel interaction, overall, regardless of the following vowels' height, a higher degree of weakening is observed when the preceding vowels are close. However, the effects of preceding vowel height reached significance only when the following vowels were also close. Specifically, post hoc pair-wise comparisons indicated that when the following vowel is close, the degree of lenition is highest when both the preceding vowels are also close relative to when they are mid or open [βs = −5.130, −6.868; ts = −3.814, −5.034; $p = 0.004$, <0.001]. No significant difference across the preceding vowels' height when the following vowels are either mid or open. These results suggest that preceding close + following close vowel context triggers the highest degree of stop weakening. In addition, when the preceding vowels are mid or low, the degree of weakening becomes greater (smaller intensity difference) as the following vowels change from close to mid and to open. For the preceding mid vowels, the difference reaches significance between the following close and the following mid vowels [$\beta = 1.300$, $t = 3.345$, $p = 0.024$] but not between the following mid and the following open vowels. For the preceding open vowels, the difference reaches significance between the following mid and the following open vowel context [$\beta = 1.979$, $t = 4.644$, $p = 0.0001$]. Together, these results suggest that the optimal lenition triggering environment is when both flanking vowels are close and that preceding mid and open vowels progressively increase the degree of lenition as the height of the following vowels increases.

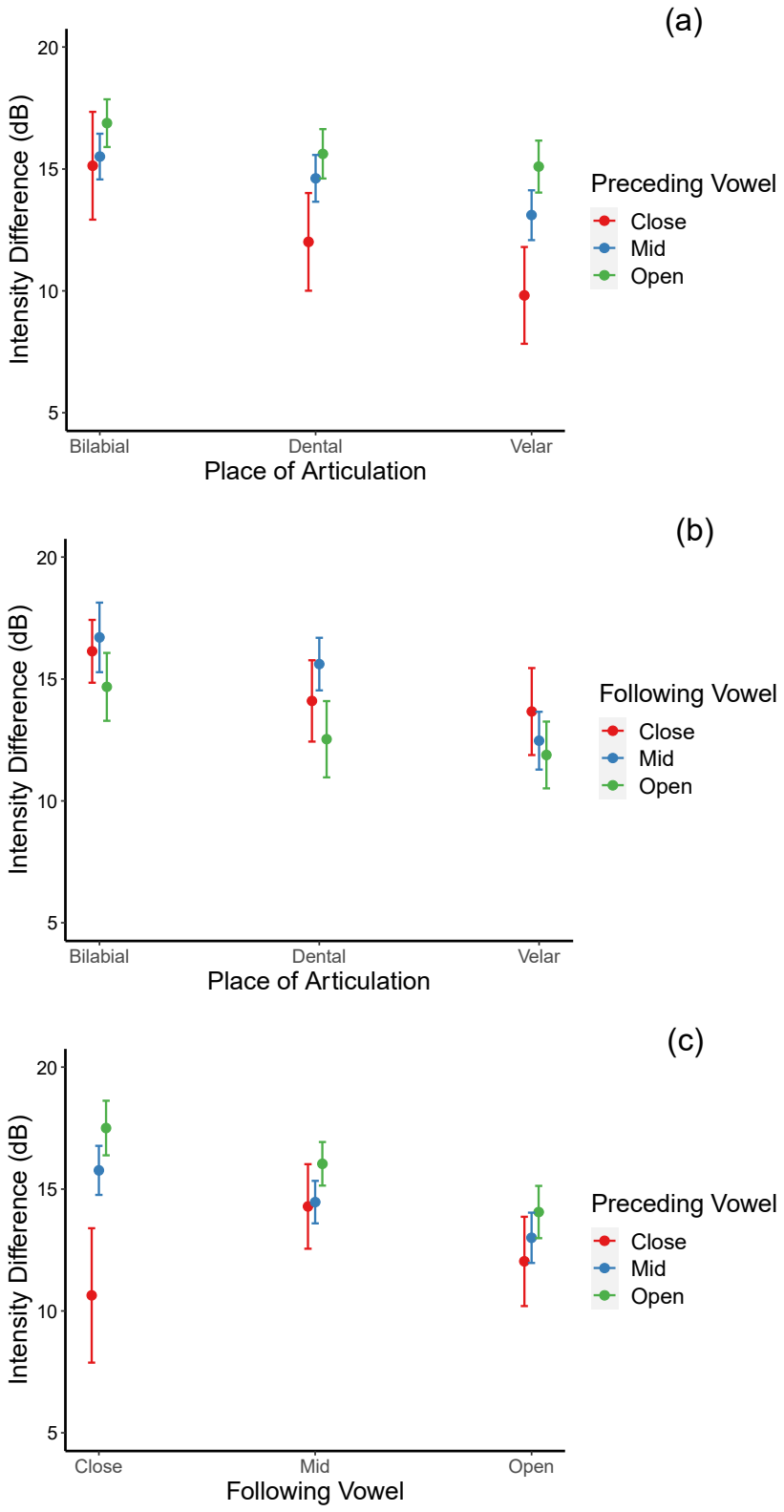

Figure 3. Estimated marginal means of intensity difference relative to the preceding vowel by place of articulation and preceding vowel (**a**, upper), by place of articulation and following vowel (**b**, middle), and by preceding and following vowel (**c**, lower). The dots represent the estimated marginal means, and the interval lines display 95% confidence interval.

3.4. Intensity Difference Relative to the Following Vowel

Similar to the intensity difference between the target stops and the preceding vowels, it is assumed that smaller intensity differences between the target stop and the following vowels would indicate a stronger degree of lenition. The best regression model for this dependent variable indicated that the model's fixed factors accounted for 72.9% of the total variance and that the full model explained 80.3% of the variance (marginal R^2 = 0.729, Conditional R^2 = 0.803). Like intensity difference relative to the preceding vowel, stress along with voicing, place of articulation, and speaking rate play a significant role in the lenition degree based on this dependent variable. However, unlike the intensity difference relative to the preceding vowel, the effect of the preceding vowel height is non-significant. Specifically, the model yielded significant main effects of stress: higher intensity difference for stressed syllables relative to unstressed syllables [$\beta = -3.741$, $t = -11.423$, $p < 0.001$]; voicing: higher intensity difference for voiceless stops than voiced stops [$\beta = -23.604$, $t = -65.454$, $p < 0.001$]; place of articulation: bilabial > dental > velar [βs = 1.838, 1.416; ts = 2.837, 2.594; ps = 0.005, 0.010]; following vowel: close < mid [$\beta = -1.563$, $t = -2.231$, $p = 0.026$]; and speaking rate: the higher the speaking rate, the lower the intensity difference [$\beta = -0.918$, $t = -23.732$, $p < 0.001$] (see Table 5). These results suggested that stops are more lenited in unstressed syllables than in stressed syllables, and when they are voiced than when they are voiceless. In addition, velar stops are more lenited than bilabial and dental stops. Furthermore, the degree of stop lenition increases when the following vowels are close and when the speaking rate increases.

Table 5. Summaries of intensity difference relative to the following vowel: The fixed-effects in the linear mixed-effects model (a: upper), and the type-III-ANOVA analysis (b: lower).

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	21.282	0.562	37.891	**<0.001**
Stress (unstressed)	−3.741	0.328	−11.423	**<0.001**
Voicing (voiced)	−23.604	0.361	−65.454	**<0.001**
Place (bilabial)	1.838	0.648	2.837	**0.005**
Place (dental)	1.416	0.546	2.594	**0.010**
Preceding vowel (close)	−0.097	0.539	−0.180	0.857
Preceding vowel (mid)	−0.327	0.207	−1.583	0.113
Following vowel (close)	−1.563	0.701	−2.231	**0.026**
Following vowel (mid)	0.494	0.510	0.969	0.333
Speaking rate	−0.918	0.039	−23.732	**<0.001**
Word status (function)	−0.453	0.576	−0.787	0.431
Preceding vowel (close): Following vowel (close)	−6.267	1.842	−3.403	**0.001**
Preceding vowel (mid): Following vowel (close)	−0.226	0.503	−0.449	0.653
Preceding vowel (close): Following vowel (mid)	1.337	1.143	1.170	0.242
Preceding vowel (mid): Following vowel (mid)	0.121	0.448	0.270	0.787
Place (bilabial): Preceding vowel (close)	3.072	1.677	1.831	0.067
Place (dental): Preceding vowel (close)	−0.161	1.128	−0.143	0.886
Place (bilabial): Preceding vowel (mid)	0.186	0.447	0.417	0.677
Place (dental): Preceding vowel (mid)	0.673	0.413	1.630	0.103
Effects	***SSE***	***MSE***	***F***	***p***
Stress	3823	3823	130.476	**<0.001**
Voicing	125,526	125,526	4284.163	**<0.001**
Place	745	372	12.709	**<0.001**
Preceding vowel	81	40	1.380	0.252
Following vowel	146	73	2.496	0.083
Speaking rate	16,502	16,502	563.200	**<0.001**
Word status	18	18	0.620	0.432
Preceding vowel: Following vowel	361	90	3.078	**0.015**
Place: Preceding vowel	262	65	2.231	0.063

Note: significant ps < 0.05 are in bold.

A significant interaction between the preceding vowel and the following vowel (see Figure 4) was also obtained. Post hoc analyses indicate that the significant interaction stemmed from the fact that intensity differences were smaller (higher degree of lenition)

when the preceding and the following vowels are close compared to when the following vowels are close, but the preceding vowels are mid [$\beta = -5.816$, $t = -3.148$, $p = 0.043$].

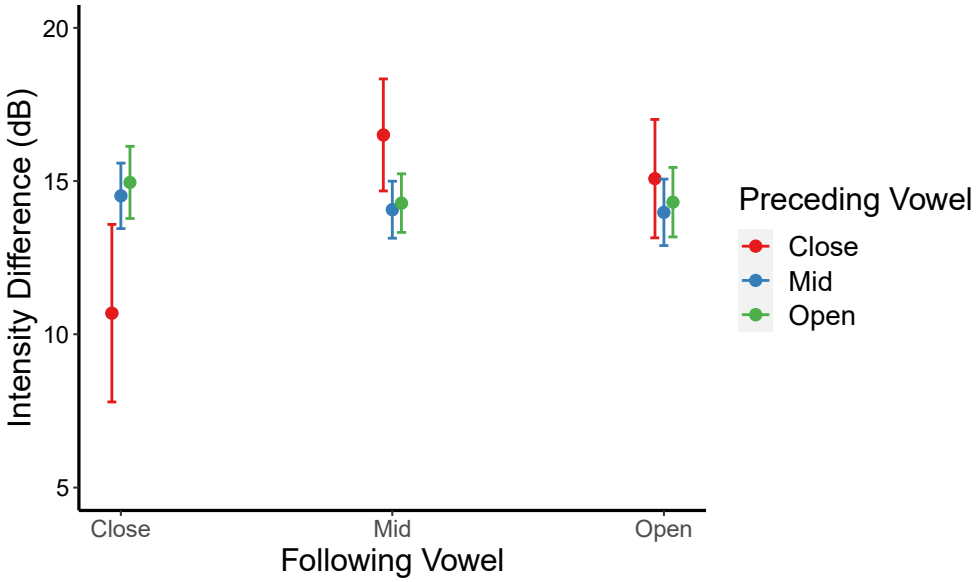

Figure 4. Estimated marginal means of intensity difference relative to the following vowel by preceding and following vowel. The dots represent the estimated marginal means and the interval lines display 95% confidence interval.

3.5. Mean Intensity

For the mean intensity of the target stops (in dB), the greater the mean intensity, the higher the degree of lenition. According to the best regression model, 65.9% (marginal $R^2 = 0.659$) of the total variance was explained by the fixed factors, and an additional 11.4% of the variance was accounted for by the random factors (condition $R^2 = 0.773$). Like relative intensity difference to the following vowel measure, stress, voicing, following vowel height, and speaking rate are strong predictors of lenition degree for this variable, while the role of place of articulation and preceding vowel height is minimal. Specifically, the model revealed significant main effects of stress: higher mean intensity for unstressed syllables [$\beta = 1.149$, $t = 5.449$, $p < 0.001$]; voicing: higher mean intensity for voiced stops [$\beta = 15.063$, $t = 65.316$, $p < 0.001$]; following vowel: close < mid < open [βs = −1.587, −1.277, ts = −6.271, −5.162, ps < 0.001]; and speaking rate: the higher the speaking rate, the higher the mean intensity [$\beta = 0.605$, $t = 22.507$, $p < 0.001$] (see Table 6). These results indicated that intervocalic stops are weakened to a significantly higher degree in unstressed syllables relative to the stressed syllables; when they are voiced than when they are voiceless; when the following vowels are relatively more open, and when the speaking rate increases.

A significant interaction between the place of articulation and the following vowel (see Figure 5) was also found. Post hoc analysis confirmed that bilabials and velars were lenited when the following vowels were close than when they were mid and open [for bilabials: βs = −1.340, −2.588; ts = −3.592, −6.541; p = 0.01, <0.0001; for velars: βs = −2.300, −2.592; ts = −4.627, −4.964; ps = 0.0002, <0.001]. However, dental stops are more lenited only when close instead of open vowels followed (for dentals; $\beta = -3.411$, $t = -6.181$, $p < 0.001$]. These results suggest that the lenition degree of each stop negatively correlates with the degree of the following vowel's openness and that the effects are more pronounced for bilabial than for velar stops.

Table 6. Summaries of mean intensity: The fixed-effects in the linear mixed-effects model (a: upper), and the type-III-ANOVA analysis (b: lower).

Predictors	*β*	*SE*	*t*	*p*
(Intercept)	61.599	0.501	122.852	**<0.001**
Stress (unstressed)	1.149	0.211	5.449	**<0.001**
Voicing (voiced)	15.063	0.231	65.316	**<0.001**
Place (bilabial)	−0.432	0.252	−1.717	0.086
Place (dental)	−0.271	0.296	−0.917	0.359
Preceding vowel (close)	−0.443	0.342	−1.296	0.195
Preceding vowel (mid)	0.226	0.120	1.886	0.059
Following vowel (close)	−1.587	0.253	−6.271	**<0.001**
Following vowel (mid)	−1.277	0.247	−5.162	**<0.001**
Speaking rate	0.605	0.027	22.507	**<0.001**
Word status (function)	0.150	0.356	0.421	0.674
Place (bilabial): Following vowel (close)	−0.217	0.540	−0.403	0.687
Place (dental): Following vowel (close)	1.178	0.635	1.855	0.064
Place (bilabial): Following vowel (mid)	1.041	0.623	1.672	0.095
Place (dental): Following vowel (mid)	−1.997	0.633	−3.154	**0.002**
Effects	***SSE***	***MSE***	***F***	***p***
Stress	448	448	29.6935	**<0.001**
Voicing	64,410	64,410	4266.1671	**<0.001**
Place	117	59	3.8891	**0.021**
Preceding vowel	68	34	2.2589	0.105
Following vowel	1498	749	49.6162	**<0.001**
Speaking rate	7648	7648	506.5626	**<0.001**
Word status	3	3	0.1772	0.674
Place: Following vowel	172	43	2.8440	**0.024**

Note: significant *p*s < 0.05 are in bold.

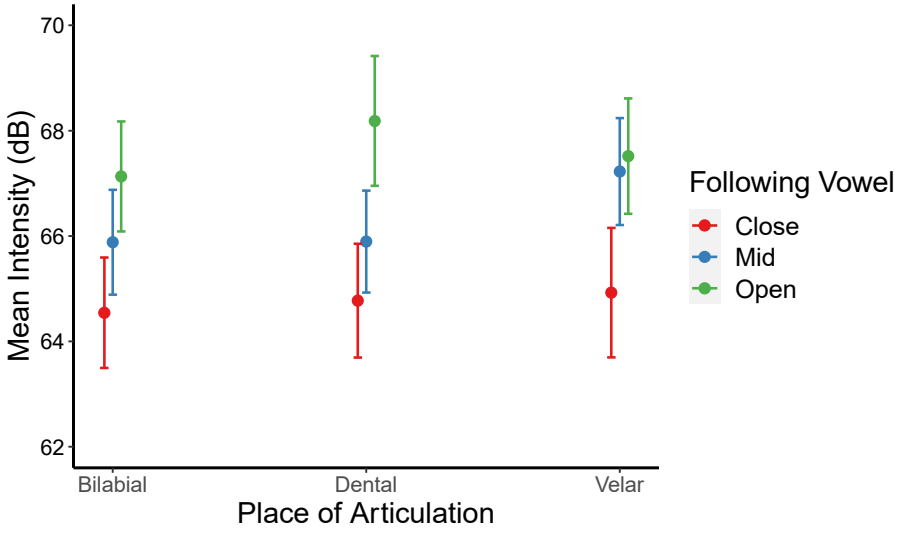

Figure 5. Estimated marginal means of mean intensity by place of articulation and following vowel. The dots represent the estimated marginal means and the interval lines display 95% confidence interval.

3.6. Sonorant Posterior Probability

Figure 6 shows the sonorant posteriors probabilities of intervocalic bilabial, dental, and velar voiced stops before (see Figure 6a) and after (see Figure 6b) close, mid, and open vowels in stressed and stressed syllables.

The best regression model for sonorant posterior probabilities showed that 52.4% of the total variance was explained by the fixed factors (marginal $R^2 = 0.524$), while the full model accounted for 60.9% of the variance (conditional $R^2 = 0.609$). The model revealed significant main effects of stress: unstressed > stressed syllables [$\beta = 0.039$, $t = 3.109$, $p = 0.002$; voicing: voiced stops > voiceless stops [$\beta = 0.660$, $t = 48.32$, $p < 0.001$]; place of articulation: bilabials >

dental [$\beta = 0.060$, $t = 2.237$, $p = 0.025$], but dental < velars [$\beta = -0.107$, $t = -4.668$, $p < 0.001$]; preceding vowel: close < mid [$\beta = -0.051$, $t = -2.196$, $p = 0.028$], mid < open [$\beta = -0.077$, $t = -8.881$, $p < 0.001$]; and speaking rate: the faster the speaking rate, the higher the sonorant posterior probabilities [$\beta = 0.013$, $t = 8.056$, $p < 0.001$] (see Table 7).

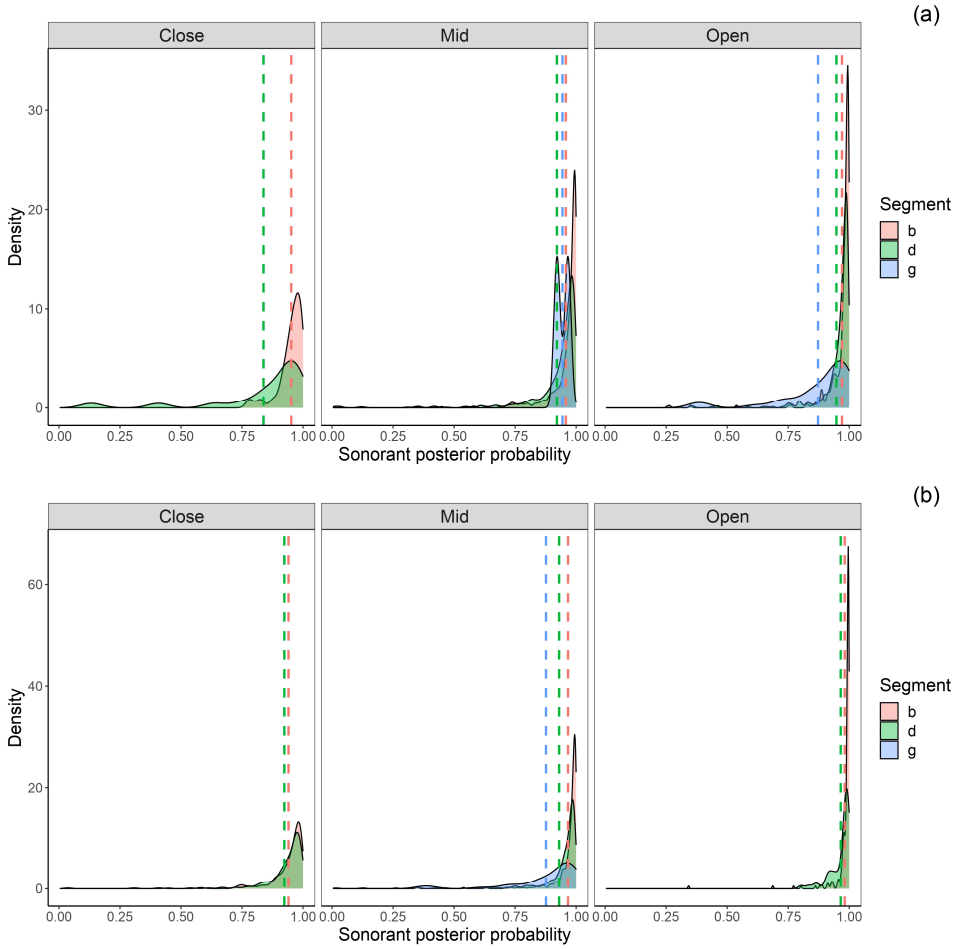

Figure 6. Sonorant posterior probability of /b, d, g/ before (**a**, upper) and after (**b**, lower) close, mid, and open vowels.

These results suggested that, like the quantitative acoustic metrics, sonorant posterior probabilities exhibit effects of known lenition factors, including stress, voicing, place of articulation, the openness of preceding vowels, and speaking rate. Specifically, sonorant posterior probabilities suggested that the degree of stop lenition is greater in unstressed syllables than in stressed syllables when the stops are voiced than when they are voiceless, when preceding vowels are lesser in the degree of openness, and when the speaking rate increases. Regarding the effects of place of articulation, the sonorant's posterior probabilities indicate that both bilabial and velar stops are more lenited than dental stops.

Interactions between place × preceding and place × following vowels, as well as between preceding × following vowels, were also significant (see Table 7). For the significant interaction between place and preceding vowel (see Figure 7a), post hoc analyses indicate that sonorant posterior probabilities were lowest when the preceding vowels are close relative to when they are mid and open, but the difference reached significance only for velar stops. Specifically, sonorant posteriors probabilities for velar stops following open vowels are significantly higher (more lenited) than when they occur after close vowels [$\beta = 0.172$, $t = 4.229$, $p = 0.0008$). In addition, when preceded by mid vowels, bilabial and velar stops are more lenited (higher sonorant posterior probabilities) than dental stops [βs = 0.075, 0.110; ts = 4.540, 5.862; ps = 0.0002, <0.001]. Similarly, when preceded by open

vowels, bilabial and velar stops were more lenited than dental stops [βs = 0.094, 0.159; ts = 4.936, 7.660; ps < 0.001, <0.001], but bilabial stops were less lenited than velar stops [β = −0.064, t = −3.311, p = 0.027].

Table 7. Summaries of sonorant posterior probability: The fixed-effects in the linear mixed-effects model (a: upper), and the type-III-ANOVA analysis (b: lower).

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	0.533	0.022	24.508	**<0.001**
Stress (unstressed)	0.039	0.013	3.109	**0.002**
Voicing (voiced)	0.660	0.014	48.323	**<0.001**
Place (bilabial)	0.060	0.027	2.237	**0.025**
Place (dental)	−0.107	0.023	−4.668	**<0.001**
Preceding vowel (close)	−0.051	0.023	−2.196	**0.028**
Preceding vowel (mid)	−0.077	0.009	−8.881	**<0.001**
Following vowel (close)	0.019	0.029	0.664	0.507
Following vowel (mid)	−0.024	0.021	−1.150	0.250
Speaking rate	0.013	0.002	8.056	**<0.001**
Word status (function)	−0.004	0.021	−0.190	0.849
Place (bilabial): Following vowel (close)	−0.054	0.032	−1.692	0.091
Place (dental): Following vowel (close)	0.134	0.038	3.566	**<0.001**
Place (bilabial): Following vowel (mid)	−0.015	0.037	−0.398	0.690
Place (dental): Following vowel (mid)	−0.117	0.037	−3.113	**0.002**
Preceding vowel (close): Following vowel (close)	0.216	0.079	2.718	**0.007**
Preceding vowel (mid): Following vowel (close)	−0.004	0.021	−0.173	0.863
Preceding vowel (close): Following vowel (mid)	−0.022	0.049	−0.449	0.653
Preceding vowel (mid): Following vowel (mid)	0.047	0.019	2.494	**0.013**
Place (bilabial): Preceding vowel (close)	−0.064	0.072	−0.885	0.376
Place (dental): Preceding vowel (close)	0.059	0.049	1.207	0.228
Place (bilabial): Preceding vowel (mid)	−0.019	0.019	−1.000	0.317
Place (dental): Preceding vowel (mid)	0.048	0.018	2.708	**0.007**
Effects	***SSE***	***MSE***	***F***	***p***
Stress	0.559	0.559	9.663	**0.002**
Voicing	135.093	135.093	2335.076	**<0.001**
Place	1.265	0.632	10.932	**<0.001**
Preceding vowel	5.489	2.744	47.435	**<0.001**
Following vowel	0.080	0.040	0.693	0.500
Speaking rate	3.754	3.754	64.895	**<0.001**
Word status	0.002	0.002	0.036	0.849
Place: Following vowel	1.595	0.399	6.893	**<0.001**
Preceding vowel: Following vowel	0.877	0.219	3.791	**0.004**
Place: Preceding vowel	0.626	0.156	2.703	**0.029**

Note: significant ps < 0.05 are in bold.

For the significant interaction between the place and the following vowel (see Figure 7b), post hoc analyses revealed no effects of the following vowel's openness on any of the three types of stops. However, when followed by open vowels, bilabial and dental stops were less lenited (lower sonority posterior probabilities) than velar stops [βs = −0.117, −0.191; ts = −3.796, −7.535; ps = 0.005, <0.001].

For the significant interaction between preceding and following vowels (see Figure 7c), post hoc analyses suggested that when followed by mid vowels, posterior sonority probabilities were significantly higher when the preceding vowels are mid or open than when they are close [βs = 0.130, 0.191; ts = 3.655, 5.294, ps = 0.008, <0.001). Similarly, when followed by open vowels, posterior sonority probabilities were significantly higher when the preceding vowels were also open rather than close [β = 0.216, t = 5.933, p < 0.001). Interestingly, no effects of the preceding vowel's openness when the following vowels are close. With the exception of close vowels, these results suggested that, stronger lenition occurs when the following and the preceding vowels are equal or greater in openness.

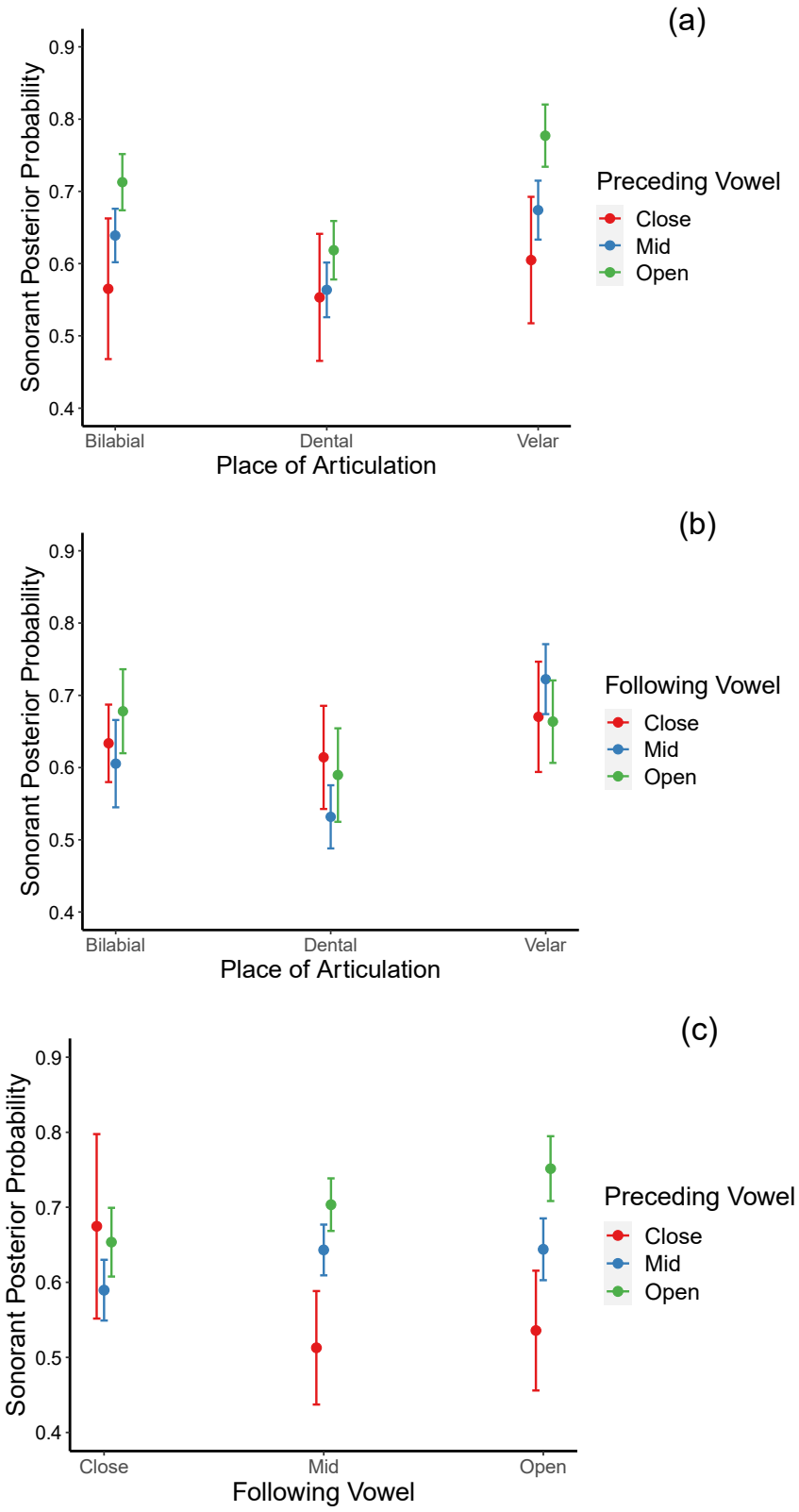

Figure 7. Estimated marginal means of sonorant posterior probability by place of articulation and preceding vowel (**a**, upper), by place of articulation and following vowel (**b**, middle), and by preceding and following vowel (**c**, lower). The dots represent the estimated marginal means and the interval lines display 95% confidence interval.

3.7. Continuant Posterior Probability

Figure 8 shows the sonorant posteriors probabilities of intervocalic bilabial, dental, and velar voiced stops before (see Figure 8a) and after (see Figure 8b) close, mid, and open vowels in stressed and stressed syllables.

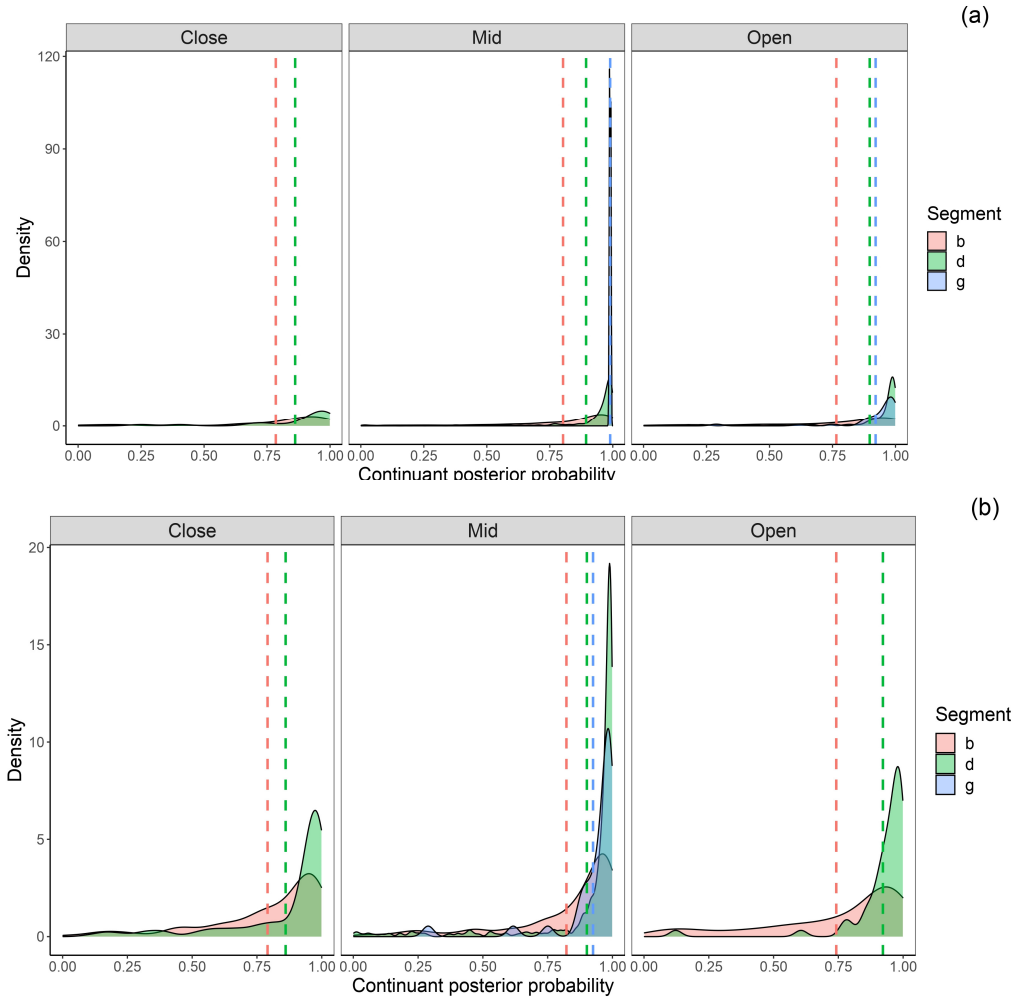

Figure 8. Continuant posterior probability of /b, d, g/ before (**a**, upper) and after (**b**, lower) close, mid, and open vowels.

52.5% of the total variance of continuant posterior probabilities was explained by the fixed factors in the regression model (marginal $R^2 = 0.525$), while the full model accounted for 61.3% of the variance (conditional $R^2 = 0.613$). The model yielded significant main effects for voicing: voiced > voiceless stops [$\beta = 0.637$, $t = 46.833$, $p < 0.001$); place of articulation: dental < velar [$\beta = -0.140$, $t = -6.591$, $p < 0.001$; preceding vowel: mid < open [$\beta = -0.030$, $t = -3.632$, $p < 0.001$], and speaking rate: the faster the speaking rate, the higher the continuant posterior probabilities [$\beta = 0.007$, $t = 4.468$, $p < 0.001$]. In addition, a significant interaction between place and preceding vowel and between preceding and following vowel contexts were also obtained (see Table 8).

Table 8. Summaries of sonorant posterior probability: The fixed-effects in the linear mixed-effects model (a: upper), and the type-III-ANOVA analysis (b: lower).

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	0.494	0.020	24.866	**<0.001**
Stress (unstressed)	0.021	0.012	1.694	0.090
Voicing (voiced)	0.637	0.014	46.833	**<0.001**
Place (bilabial)	−0.030	0.026	−1.169	0.243
Place (dental)	−0.140	0.021	−6.591	**<0.001**
Preceding vowel (close)	−0.042	0.022	−1.875	0.061
Preceding vowel (mid)	−0.030	0.008	−3.632	**<0.001**
Following vowel (close)	0.027	0.028	0.977	0.329
Following vowel (mid)	0.007	0.020	0.328	0.743
Speaking rate	0.007	0.002	4.468	**<0.001**
Word status (function)	−0.028	0.021	−1.327	0.185
Preceding vowel (close): Following vowel (close)	0.190	0.076	2.512	**0.012**
Preceding vowel (mid): Following vowel (close)	0.005	0.020	0.257	0.797
Preceding vowel (close): Following vowel (mid)	−0.013	0.047	−0.275	0.784
Preceding vowel (mid): Following vowel (mid)	0.020	0.018	1.087	0.277
Place (bilabial): Preceding vowel (close)	−0.122	0.068	−1.776	0.076
Place (dental): Preceding vowel (close)	0.043	0.046	0.925	0.355
Place (bilabial): Preceding vowel (mid)	0.020	0.018	1.081	0.280
Place (dental): Preceding vowel (mid)	0.035	0.017	2.049	**0.041**
Effects	***SSE***	***MSE***	***F***	***p***
Stress	0.148	0.148	2.8712	0.090
Voicing	112.987	112.987	2193.3072	**<0.001**
Place	3.171	1.586	30.7823	**<0.001**
Preceding vowel	1.036	0.518	10.0561	**<0.001**
Following vowel	0.080	0.040	0.7748	0.461
Speaking rate	1.028	1.028	19.9592	**<0.001**
Word status	0.091	0.091	1.7613	0.186
Preceding vowel: Following vowel	0.460	0.115	2.2338	0.063
Place: Preceding vowel	0.683	0.171	3.3141	**0.010**

Note: significant ps < 0.05 are in bold.

Post hoc, follow-up tests revealed that a significant interaction between place and preceding vowel (see Figure 9a) stemmed from the fact that continuant posterior probabilities for bilabial and dental stops were significantly lower (less lenited) than those of velar stops when preceded by the mid and open vowels [βs = −0.126, 0.180; ts = −7.600, −9.833; ps ≤ 0.0001 for bilabials; βs = −0.143, −0.178; ts = −8.140, −9.322; ps < 0.001, for dentals). In the preceding close vowel context, the difference between the velar stops and the bilabial stops, but not between the velar stops and the dental stops, almost reached a significant level [β = −0.205, t = −3.073, p = 0.055 for bilabials; β = −0.100, 0.047; t = −2.125, p = 0.4563). These results suggested the following ranking from most lenited to least lenited: velar > dental > labial in postvocalic position.

For the significant interaction between the preceding and following vowels (see Figure 9b), post hoc analyses revealed the following. When followed by mid vowels, continuant posterior probabilities were significantly lower (less lenited) when the preceding vowels were close rather than mid [β = −0.109, t = −3.207, p = 0.036] or open [β = −0.135, t = −3.914, p = 0.003]. When followed by open vowels, continuant posterior probabilities are significantly lower only when the preceding vowels are close rather than mid [β = −0.142, t = −4.077, p = 0.002]. On the other hand, when the following vowels are close, continuant posterior probabilities across the three preceding vowel contexts did not reach significance. These results suggest that continuant posterior probabilities are more likely to increase when the preceding and the following vowels are relatively more open (e.g., mid or open, not close). When mid vowels precede the target stops, continuant posterior probabilities increase when the following vowels are of the same or one step higher in the degree of openness. However, the open vowels precede the target stops, and continuant posterior probabilities increase when the following vowels are one step lower in openness.

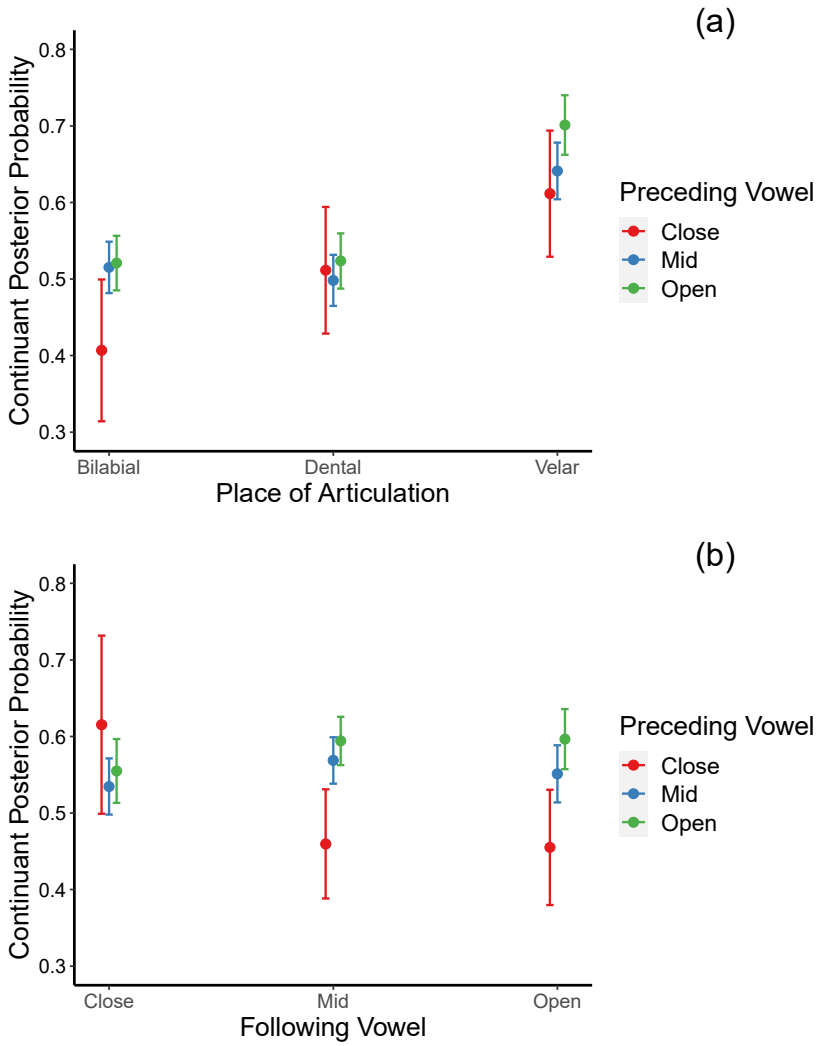

Figure 9. Estimated marginal means of continuant posterior probability by place of articulation and preceding vowel (**a**, upper), and by preceding and following vowel (**b**, lower). The dots represent the estimated marginal means and the interval lines display 95% confidence interval.

4. Results, Summary and Discussion

The degree of lenition of Spanish voiced stop varies as a function of several factors, including stress, place of articulation, quality of surrounding vowels, word status (content or function), and speaking rate. Despite extensive research, no standard method to quantify the degree of lenition has emerged. Under the quantitative acoustic approach, different acoustic dimensions have been employed by different researchers as correlates of lenition. In this study, we compared five acoustic indices of lenition of stops in an Argentinian Spanish corpus (harmonic-to-noise ratio (HNR), duration of the target stops relative to the sum duration of the preceding vowel + target stop + following vowel, intensity difference between the target stops and their preceding and following vowels and mean intensity of the target stops) to the posterior probabilities of the sonorant and continuant phonological features derived from Phonet, a deep recurrent neural network model. The seven lenition metrics are entered as the dependent variables in a series of linear mixed-effect regression models with known factors of lenition, including stress, place of articulation, voicing, preceding and following vowel height/openness, word status (function or content) and speaking rate, as fixed factors. The degree of lenition is predicted to be higher (e.g., higher HNR, lower duration ratio, lower relative intensity differences, higher mean intensity, higher sonorant posterior probability, and higher continuant posterior probability values) in unstressed syllables relative to stressed syllables. Similarly, a higher degree of lenition is

predicted for voiced stops relative to voiceless stops, for function words relative to content words, and for a faster speaking rate relative to a slower speaking rate (Broś et al. 2021). Regarding the place of articulation and flanking vowels, mixed results have been reported. For example, Simonet et al. (2012) reported that /d/ is more lenited after a low vowel than after a higher vowel among (Iberian, Majorcan) Spanish, and (Majorcan) Catalan bilinguals. In contrast, Cole et al. (1999) and Ortega-Llebaria (2004) found Spanish /g/ to be less lenited between low vowels than between high vowels, while no effect of vowel height was found for /b/ (Ortega-Llebaria 2004, 2003).

Table 9 summarizes the significant main effects of the seven regression models, one for each dependent variable. As shown in this table, all seven metrics predicted that intervocalic voiced stops in Argentinian Spanish are lenited to a significantly higher degree than voiceless stops and that the lenition degree increases when the speaking rate becomes faster. Interestingly, unlike Broś et al. (2021), who found that lenition is more likely in function words relative to content words in Spanish stops spoken in the Canary Islands when relative duration is a dependent variable, word status is not predictive of lenition advancement along any of the seven metrics we examined, including relative duration. Besides dialectal differences, the right context of the stops examined in Broś et al. (2021) is not limited to vowels but also includes consonants. These or other currently unknown factors may account for this discrepant finding. Crucially, these results suggest that sonorant and continuant posterior probabilities share a predictive pattern of lenition with the five quantitative acoustic metrics.

Table 9. Summary of the linear mixed-effects regression models. ✓denotes significant main effects ($p < 0.05$).

	Mean HNR	Relative Duration	Relative Intensity (re: Preceding Vowel)	Relative Intensity (re: Following Vowel)	Mean Intensity	Sonorant Posterior Probability	Continuant Posterior Probability
Stress			✓	✓	✓	✓	
Voicing	✓	✓	✓	✓	✓	✓	✓
Place	✓	✓	✓	✓		✓	✓
Preceding vowel	✓	✓	✓			✓	✓
Following vowel	✓	✓	✓	✓	✓		
Speaking rate	✓	✓	✓	✓	✓	✓	✓
Word status							

A similar predictive pattern of lenition between sonorant posterior probability and all three intensity measurements is also observed. Specifically, all four metrics predict the degree of lenition in the expected direction: more advanced in unstressed than in stressed syllables. Interestingly, values along all five acoustic metrics, but not the sonorant and continuant posteriors probabilities, significantly vary according to the following vowels. Crucially, the behavior of the sonorant and continuant posteriors probabilities is consistent with the finding that the height of the preceding vowels rather than the height of the following vowels introduced a varying degree of constriction, at least for /d/, in intervocalic context (Simonet et al. 2012). Moreover, inconsistent effects of the following vowels are observed across the five acoustic measurements. Specifically, HNR, relative duration, and relative intensity (to the following vowel) values predicted a lower degree of stop constriction (higher degree of lenition) when the following vowels are relatively close than when they are relatively more open. However, the opposite pattern was suggested by the relative intensity (to the preceding vowel) and mean intensity values: a more advanced degree of lenition when the following vowels are relatively more open.

However, like three of the acoustic metrics, namely HNR, relative duration, and relative intensity (to the preceding vowel), sonorant and continuant posterior probabilities vary significantly but differently as a function of preceding vowel height. That is, while

values along the three acoustic metrics suggest a higher degree of weakening when the preceding vowels are relatively more close, the opposite is indicated by the sonorant and the continuant posterior probabilities: the more open the preceding vowels are, the greater the degree of lenition. Note, however, that the predictive pattern of sonorant and continuant posterior probabilities, but not of the three acoustic dimensions, is consistent with the articulatory effort-based view of lenition (Kirchner 1998, 2013). More specifically, since the distance that the articulators will travel from (and possibly also to) a lower vowel is likely reduced relative to a higher vowel, lenition should be more prevalent after a more open than a more close vowel.

Place of articulation of the stops exerts influence on four acoustic metrics: HNR, relative duration, relative intensity (to the preceding vowel), relative intensity (to the following vowel), as well as on the sonorant and the continuant posterior probabilities. While the pattern of the influence is similar for the sonorant and the continuant posterior probabilities, it differs across the four acoustic metrics. Like sonorant and continuant posterior probabilities, the two acoustic intensity measures indicate that velar stops are more lenited than dental and bilabial stops. This lenition pattern is consistent with that of Kingston's (2008) finding on Spanish stops produced by two female speakers from Ecuador and Peru. He reasoned that velar stops are more lenited perhaps "because velar closures are more often incomplete" (footnote 20, p. 21).

However, the two sets of measures (two acoustic intensity measures vs. sonorant and continuant posterior probabilities) differ in their ranking of the bilabial and the dental stops. Consistent with Kingston (2008)'s prediction, dentals are more lenited than bilabials according to the two acoustic intensity measures, while the opposite is predicted by the sonorant and the continuant posterior probabilities. Alternatively, contrary to Kingston's prediction, it is possible that bilabial closures are less complete than dental closures and are thus more lenited. The fact that the two intensity measures are computed relative to their immediately preceding and following vowels only while sonorant and continuant posterior probabilities are estimated globally based on all sonorant, and continuant segments in the corpus may also explain their differing predictive pattern. On the other hand, HNR and relative duration measures suggest that dental stops are (non-significantly) more lenited than velar stops and significantly more lenited than bilabial stops, while relative duration indicates that dental stops are significantly more lenited than both bilabial velar stops. These results are inconsistent with the effort-based account of lenition (Kirchner 1998, 2013) and Kingston's (2008) findings.

In addition to significant main effects, our regression models also yielded significant interactions mainly between the place of articulation and the preceding and following vowels as well as between preceding and following vowels. The significant place × preceding vowel interaction was found for HNR, relative duration, relative intensity (to the preceding vowel), as well as for sonorant and continuant posterior probabilities. Post hoc pair-wise comparisons indicated different interaction patterns for different metrics, with the patterns for sonorant and continuant posteriors probabilities being more uniform and consistent with previous findings (e.g., Kingston 2008) than the three acoustic metrics. More specifically, the more open the preceding vowels are, the more lenited the velar stops are predicted by the sonorant posterior probabilities. Further, velar stops are predicted to be more lenited than bilabial and dental stops when preceded by open vowels. Similarly, continuant posterior probabilities predicted that velar stops are more lenited than bilabial and dental stops when preceded by mid and open vowels but not close vowels. On the other hand, HNRs predicted a higher degree of lenition for dental and bilabial stops after close vowels, while preceding low vowels trigger a higher degree of lenition for velar stops. Relative duration predicts a higher degree of lenition for dental stops compared to bilabial and velar stops after open vowels, while relative intensity (to the preceding vowels) predicts a more advanced degree of lenition of velar stops relative to bilabial and dental stops when the preceding vowels are relatively more close than open.

Significant place × following vowel interactions are found for every metric, except for mean intensity and continuant posterior probabilities. Overall, post hoc analyses suggested that following close vowels introduced a higher degree of variation in the acoustic metric values and predicted a higher degree of lenition of stops in this context. On the other hand, sonorant posterior probabilities are more affected by following open vowels and predicted higher degrees of lenition of velar stops in this environment.

Significant preceding × following vowel interactions are found for HNR, relative duration, relative intensity (to the preceding vowel), relative intensity (to the following vowels), and sonorant and continuant posterior probabilities. Post hoc analyses indicated that a higher degree of lenition is predicted by all acoustic metrics when flanking vowels are relatively close in height and when they are relatively more close than relatively more open. On the other hand, relatively more open flanking vowels are predicted to trigger a higher degree of lenition, particularly for velar stops, by the sonorant posterior probabilities. On the contrary, continuant posterior probabilities predicted a higher degree of lenition when either the preceding or the following vowel height is relatively high. This finding may be explained by the fact that smaller oral opening is more conducive to friction generation, characteristics of fricatives, and members of the [+continuant] phonological class.

In conclusion, the degree of lenition predicted by different lenition metrics vary. However, lenition patterns predicted by the sonorant and continuant posterior probabilities are more consistent and in the direction expected by previous findings and the effort-based view of lenition. As far as the main effects are concerned, lenition patterns predicted by sonorant and continuant posterior probabilities are largely consistent with the relative acoustic intensity measures. This is not surprising given that inputs to the Phonet model that generate the sonorant and continuant posterior probabilities are feature sequences based on log energy distributed across 33 triangular Mel filters of each 0.5 s chunk of the input signals. Some differences between these two sets of metrics may lie in the fact that acoustic intensity measures are relative to the target stops immediate left and right contexts, while sonorant and posterior probability estimates are relative to the whole class of sonorant and continuant segments in the corpus. Sonorant and continuant posterior probabilities relative to the preceding and following segment only could be used in future research to see if minor discrepancies in lenition predictive patterns found between these two sets of metrics could be eliminated. In addition, the approach could be further improved by replacing forced alignment with the automated segmentation method proposed by (Ennever et al. 2017).

5. Conclusions

The degree of intervocalic Argentinian Spanish stop weakening across known lenition factors predicted by five quantitative acoustic metrics and two metrics, posteriors probabilities of the sonorant and the continuant phonological features, derived from a deep neural network Phonet model, were compared. As expected, all seven metrics predicted a higher degree of lenition in stressed syllables relative to unstressed syllables and in a faster speaking rate compared to a slower speaking rate. On the contrary, the effects of flanking vowel height and stop place of articulation on the lenition patterns were differentially predicted by the five acoustic metrics. However, lenition patterns predicted by the sonorant and the continuant posterior probabilities are largely consistent with those of the relative acoustic intensity measures confirming, on the one hand, the superiority of the intensity measures and, on the other hand, the reliability of Phonet as an alternative or additional approach to investigate the degree of lenition.

Author Contributions: 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, R.W.; writing—review and editing, K.T. and F.W.; visualization, F.W.; supervision, R.W. and K.T.; project administration, R.W. and K.T.; funding acquisition, R.W. and K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NSF-National Science Foundation (SenSE). Award No. 8522 037266—SenSE.

Data Availability Statement: The data presented in this study are available upon request.

Acknowledgments: We thank Lara Rüter for her assistance in typesetting the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HNR	Harmonic-to-noise ratio
dB	Decibel

References

Backley, Phillip. 2011. *Introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Bailey, George. 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 (NWAV 44)*. Philadelphia: University of Pennsylvania Working Papers in Linguistics (PWPL), vol. 22, pp. 10–20.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models Using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Broś, Karolina, Marzena Żygis, Adam Sikorski, and Jan Wołłejko. 2021. Phonological contrasts and gradient effects in ongoing lenition in the Spanish of Gran Canaria. *Phonology* 38: 1–40. [CrossRef]

Carrasco, Patricio, José I. Hualde, and Miquel Simonet. 2012. Dialectal differences in Spanish voiced obstruent allophony: Costa Rican versus Iberian Spanish. *Phonetica* 69: 149–79. [CrossRef]

Celdrán, Eugenio Martínez. 1984. Cantidad e intensidad en los sonidos obstruyentes del castellano: Hacia una caracterización acústica de los sonidos aproximantes. *Estudios de Fonética Experimental* 71–129.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Cohen Priva, Uriel, and Emily Gleason. 2020. The causal structure of lenition: A case for the causal precedence of durational shortening. *Language* 96: 413–48. [CrossRef]

Colantoni, Laura, and Irina Marinescu. 2010. The scope of stop weakening in Argentine Spanish. In *Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*. Austin: Cascadilla Press, pp. 100–14.

Cole, Jennifer, José Ignacio Hualde, and Khalil Iskarous. 1999. Effects of prosodic and segmental context on /g/-lenition in Spanish. In *Proceedings of the Fourth International Linguistics and Phonetics Conference*. Prague: The Karolinum Press, vol. 2, pp. 575–89.

Dalcher, Christina Villafaña. 2008. Consonant weakening in Florentine Italian: A cross-disciplinary approach to gradient and variable sound change. *Language Variation and Change* 20: 275–316. [CrossRef]

Davis, Steven, and Paul Mermelstein. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28: 357–66. [CrossRef]

Eddington, David. 2011. What are the contextual phonetic variants of in colloquial Spanish? *Probus* 23: 1–19. [CrossRef]

Ennever, Thomas, Felicity Meakins, and Erich R. Round. 2017. A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology* 8: 20. [CrossRef]

Figueroa, Mauricio, and Bronwen G. Evans. 2015. Evaluation of segmentation approaches and constriction degree correlates for spirant approximant consonants. Paper presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14.

González, Carolina. 2002. Phonetic variation in voiced obstruents in North-Central Peninsular Spanish. *Journal of the International Phonetic Association* 32: 17–31. [CrossRef]

Guevara-Rukoz, Adriana, Isin Demirsahin, Fei He, Shan Hui Cathy Chu, Supheakmungkol Sarin, Knot Pipatsrisawat, Alexander Gutkin, Alena Butryna, and Oddur Kjartansson. 2020. Crowdsourcing Latin American Spanish for Low-Resource Text-to-Speech. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC)*. Marseille: European Language Resources Association (ELRA), pp. 6504–13.

Gurevich, Naomi. 2011. *The Blackwell Companion to Phonology*. Chester: John Wiley & Sons, Ltd., vol. 3, Chp. 66.

Hammond, Robert M. 2001. *The Sounds of Spanish: Analysis and Application (with Special Reference to American English)*. Somerville: Cascadilla Press.

Harris, James Wesley. 1969. *Spanish Phonology*. Cambridge: Press Research Monographs, MIT Press.

Harris, John, and Eno-Abasi Urua. 2001. Lenition degrades information: Consonant allophony in Ibibio. *Speech, Hearing and Language: Work in Progress* 13: 72–105.

Harris, John, Eno-Abasi Urua, and Kevin Tang. Forthcoming. A unified model of lenition as modulation reduction: Gauging consonant strength in Ibibio. *Phonology*. (Preprint on PsyArXiv). Available online: https://psyarxiv.com/a25yw/ (accessed on 1 March 2023).

Hayes, Bruce. 2008. *Introductory Phonology*. Hoboken: John Wiley & Sons, vol. 7.

Honeybone, Patrick. 2012. Lenition in English. In *The Oxford Handbook of the History of English*. Oxford: Oxford University Press. [CrossRef]
Hualde, José Ignacio. 2005. *The Sounds of Spanish with Audio CD*. Cambridge: Cambridge University Press.
Hualde, José Ignacio. 2013. *Los Sonidos del español: Spanish Language Edition*. Cambridge: Cambridge University Press.
Hualde, José Ignacio, Miquel Simonet, and Marianna Nadeu. 2012. Consonant lenition and phonological recategorization. *Laboratory Phonology* 2: 301–29. [CrossRef]
Hualde, José Ignacio, Ryan Shosted, and Daniel Scarpace. 2011. Acoustics and Articulation of Spanish /d/ Spirantization. Paper presented at the 17th International Congress of Phonetic Scsiences, Hong Kong, China, August 17–21; vol. 27, pp. 906–9.
Huang, Xuedong, Alex Acero, Hsiao-Wuen Hon, and Raj Reddy. 2001. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Upper Saddle River: Prentice Hall PTR.
Jakobson, Roman, C. Gunnar Fant, and Morris Halle. 1951. *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. Cambridge: MIT Press.
Kendall, Tyler, Charlotte Vaughn, Charlie Farrington, Kaylynn Gunter, Jaidan McLean, Chloe Tacata, and Shelby Arnson. 2021. Considering performance in the automated and manual coding of sociolinguistic variables: Lessons from variable (ing). *Frontiers in Artificial Intelligence* 4: 648543. [CrossRef]
Kingma, Diederik P., and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv* arXiv:1412.6980.
Kingston, J. Lenition. 2008. *Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Somerville: Cascadilla Press, pp. 1–31.
Kirchner, Robert Martin. 1998. An Effort Based Approach to Consonant Lenition. Ph.D. thesis, University of California, Los Angeles, CA, USA.
Kirchner, Robert Martin. 2013. *An Effort Based Approach to Consonant Lenition*. New York: Routledge.
Lahiri, Aditi, and Henning Reetz. 2002. Underspecified recognition. In *Laboratory Phonology 7*. Edited by Carlos Gussenhoven and Natasha Warner. Berlin: Mouton de Gruyter, pp. 637–76.
Lahiri, Aditi, and Henning Reetz. 2010. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38: 44–59. [CrossRef]
Lavoie, Lisa. M. 2001. *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. New York: Garland.
Lenth, Russell V., Paul Buerkner, Maxime Herve, Jonathon Love, Hannes Riebl, and Henrik Singmann. 2021. Emmeans: Estimated Marginal Means, aka Least-Squares Means [R Package]. Available online: https://www.scinapse.io/papers/3089169625 (accessed on 1 March 2023).
Lewis, Anthony Murray. 2001. *Weakening of Intervocalic /p, t, k/ in two Spanish Dialects: Toward the Quantification of Lenition Processes*. Champaign: University of Illinois at Urbana-Champaign.
Lozano, Maria Del Carmen. 1978. Stop and Spirant Alternations: Fortition and Spirantization Processes in Phonology. Ph.D. thesis, Indiana University, Bloomington, IN, USA.
Magloughlin, Lyra. 2018. /tɹ/ and /dɹ/ in North American English: Phonologization of a Coarticulatory Effect. Ph.D. thesis, University of Ottawa, Ottawa, ON, Canada.
Martínez-Celdrán, Eugenio, and Xosé Luís Regueira. 2008. Spirant approximants in Galician. *Journal of the International Phonetic Association* 38: 51–68. [CrossRef]
Mascaró, Joan, and Mark Aronoff. 1984. Continuant spreading in Basque, Catalan, and Spanish. *Language Sound Structure. Studies in Phonology Presented to Morris Halle by His Teacher and His Students* 8: 287–98.
McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In *Proceedings of the Eighteenth Annual Conference of the International Speech Communication Association*. Stockholm: International Speech Community Association (ISCA), pp. 498–502. [CrossRef]
McLarty, Jason, Taylor Jones, and Christopher Hall. 2019. Corpus-Based Sociophonetic Approaches to Postvocalic R-Lessness in African American Language. *American Speech* 94: 91–109. [CrossRef]
Navarro Tomás, Tomás. 1977. *Manual de Pronunciación Española*, 1st ed. Madrid: Centro de Estudios Históricos.
Ortega-Llebaria, Marta. 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 (CPhS-15)*. Barcelona: Causal Productions, vol. 7, pp. 2817–20.
Ortega-Llebaria, Marta. 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 Timothy L. Face. Berlin: De Gruyter Mouton, pp. 237–53.
Pandey, Ayushi, Pamir Gogoi, and Kevin Tang. 2020. Understanding forced alignment errors in Hindi-English code-mixed speech—A feature analysis. In *Proceedings of the First Workshop on Speech Technologies for Code-Switching in Multilingual Communities 2020*. Online: INCOMA Ltd., pp. 13–17.
Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and et al. 2011. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* 12: 2825–30. [CrossRef]
R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Romero, Gallego J. 1995. Gestural Organization in Spanish: An Experimental Study of Spirantization and Aspiration. Ph.D. thesis, University of Connecticut, Storrs, CT, USA.

Simonet, Miquel, José I. Hualde, and Marianna Nadeu. 2012. Lenition of /d/ in spontaneous Spanish and Catalan. In *Proceedings of the Thirteenth Annual Conference of the International Speech Communication Association*. Portland: International Speech Communication Association (ISCA), vol. 2, pp. 1414–17.

Soler, Antonia, and Joaquín Romero. 1999. The role of duration in stop lenition in Spanish. In *Proceedings of the 14th International Congress of Phonetic Sciences*. Oakland: The Regents of the University of California, vol. 1, pp. 483–86.

Vásquez-Correa, Juan Camilo, Philipp Klumpp, Juan Rafael Orozco-Arroyave, and Elmar Nöth. 2019. Phonet: A Tool Based on Gated Recurrent Neural Networks to Extract Phonological Posteriors from Speech. In *Proceedings of the Twentieth Annual Conference of the International Speech Communication Association*. Graz: International Speech Community Association (ISCA), pp. 549–53. [CrossRef]

Villarreal, Dan, Lynn Clark, Jennifer Hay, and Kevin Watson. 2020. From categories to gradience: Auto-coding sociophonetic variation with random forests. *Laboratory Phonology* 11: 6. [CrossRef]

Yuan, Jiahong, and Mark Liberman. 2009. Investigating /l/ variation in English through forced alignment. In *Proceedings of the Tenth Annual Conference of the International Speech Communication Association*. Brighton: International Speech Community Association (ISCA), pp. 2215–18. [CrossRef]

Yuan, Jiahong, and Mark Liberman. 2011. /l/ variation in American English: A corpus approach. *Journal of Speech Sciences* 1: 35–46. [CrossRef]