Journal of the International Phonetic Association (2026), **56**, e7, pp. 1–25
doi:10.1017/S0025100325100881

ARTICLE

The acoustic realization of intervocalic /v/ and /v:/ in Italian: evidence of articulatory variability

Angelo Dian[1,*], John Hajek[2] and Janet Fletcher[3]

[1]University of Oxford, Faculty of Linguistics, Philology and Phonetics, United Kingdom, [2]School of Languages and Linguistics, University of Melbourne, Australia, and [3]The University of Melbourne Faculty of Arts, Australia
*Corresponding author. Email: angelo.dian@ling-phil.ox.ac.uk

(Received 21 May 2025; revised 25 August 2025; accepted 6 November 2025)

Abstract

This study examines the variation in the acoustic properties of the allophones of intervocalic /v/ and /v:/ in Italian, a typologically unusual case where voiced labiodental fricatives contrast phonemically in length. Through an acoustic analysis of read speech from 18 speakers across three regional varieties (Veneto, Roman, Calabrian), we investigate whether realization patterns are shaped by prosodic (phase position and stress placement) and indexical factors (regional variety and speaker sex). Results indicate that /v/ frequently (47%) surfaces as an approximant [ʋ]. In contrast with the principle of geminate 'inalterability', /v:/ also exhibits great articulatory variability, with 36% of tokens showing increased constriction including 17% of all tokens realized as a previously undocumented plosive-like labiodental [b̪:]. Bayesian modelling reveals that approximant variants of /v/ occur more frequently phrase-finally than phrase-medially and that male speakers are more likely to produce more constricted allophones of /v:/. No consistent effects of regional variety were observed, suggesting that the variation may be widespread across Italy. Furthermore, phonemic length alone significantly predicts duration. Therefore, notwithstanding the enhanced phonetic differentiation between /v/ and /v:/, it seems most likely speakers rely primarily on duration to distinguish the pair.

Keywords: voiced labiodental fricative; Italian gemination; consonantal strength; acoustic phonetics; articulatory variability

1. Introduction

The phonemic inventory of Italian consonants includes voiced labiodental fricatives /v, v:/, which are contrastively short (singleton) or long (geminate) in this language, e.g., /ˈbeve/ *beve* ('s/he drinks') vs. /ˈbɛv:e/ *bevve* ('s/he drank'), or /avaˈl:are/ *avallare* ('to endorse') vs. /av:aˈl:are/ *avvallare* ('to sink') (Bertinetto & Loporcaro 2005; Rogers & D'Arcangeli 2004). Notably, /v:/ is the only geminate voiced fricative phoneme in standard Italian. Compared to other obstruents, voiced labiodental fricatives, especially geminate /v:/, have been underresearched experimentally in this language as well as cross-linguistically. Below we provide a phonetic description of these sounds followed by an overview of the few relevant previous studies.

1.1. The phonetic characteristics of /v, v:/ *cross-linguistically and in Italian*

In principle, voiced fricatives require the simultaneous presence of voicing (vocal fold vibration) and frication (turbulent noise) (Catford 1977; Stevens et al. 1992; Bárkányi & Kiss 2010). This articulatory requirement makes them more complex to produce than their voiceless counterparts, as frication is generated most efficiently in the absence of voicing – a point discussed in more detail below. This increased production complexity may contribute to the lower cross-linguistic frequency of voiced fricatives.

Focusing on labiodentals, the most recent version of the UCLA Phonological Segment Inventory Database,[1] first published by Maddieson (1984), reports that /v/ occurs in 95 languages, compared to 180 for its voiceless counterpart /f/, based on a sample of 415 languages. Notably, this survey includes no data on geminate /v:/, suggesting that this sound is typologically rare. According to the PHOIBLE database of phonological inventories (Moran & McCloy 2019), long [v:] is attested – either as a derived or underlying form – in only ten of the 2,186 languages surveyed, including Italian. Closer inspection of the relevant language descriptions indicates that [v:] in these cases typically results from phonological processes that apply at morphosyntactic boundaries – such as assimilation across morpheme junctures or external sandhi – rather than functioning as a contrastive segment. In Italian, however, geminate /v:/ is clearly phonemic, exhibiting lexical contrast with singleton /v/.[2] Beyond PHOIBLE, phonemic /v:/ has, to our knowledge, been documented in Miyako Ryukyuan, spoken in Japan (Fujimoto et al. 2023) and in Persian (Hansen & Myers 2017).

Voicing and frication involve conflicting aerodynamic requirements. Voicing requires appropriate adduction of the vocal folds, which reduces the volume velocity of airflow through the glottis (Stevens 2000). This reduction limits the buildup of oral air pressure, thereby diminishing frication noise, as effective frication depends on a significant pressure drop across the constriction (cf. Shadle 2010). The resulting decrease in intraoral pressure may even cause the articulators to come together, producing a brief occlusion. In contrast, the production of audible frication requires a high-volume velocity of airflow through a narrow articulatory constriction – best achieved with a spread glottis. This glottal configuration, however, inhibits vocal fold vibration and thus precludes voicing (Ohala 1983; Stevens 2000).

To illustrate these aerodynamic differences, Figures 1 and 2 present dynamic airflow (top) and intraoral pressure (bottom) measurements over time for the sequences [afa] and [ava], respectively, produced by the first author, a native speaker of Italian. The average airflow measured between the green and red vertical lines is 117 mL/s for [f] and 36 mL/s for [v]. The corresponding intraoral pressure measured at the midpoint of the pressure pulses is 10.10 cmH_2O for [f] and 5.41 cmH_2O for [v]. In this example, [v] exhibits approximately one-third the airflow and half the intraoral pressure of [f].

Given this complex interplay between voicing and frication, non-canonical realizations of short and long voiced labiodental fricatives are likely to occur alongside canonical [v, v:] forms. Acoustically, canonical realizations should exhibit low-frequency periodicity

[1] Available at http://web.phonetik.uni-frankfurt.de/upsid_info.html. Accessed on 17 April 2025.

[2] The languages in addition to Italian that are reported to exhibit [v:] in PHOIBLE are Hungarian, Sudanese Arabic, Kanga, Legbo, Réunion Creole French, Modern Greek, Pite Saami, North Saami, and Campidanese Sardinian. However, based on the references provided on the PHOIBLE website, it is often unclear what status this sound has in a given language's grammar. While it is reported to be clearly phonemic in, e.g., Pite Saami (Wilbur 2014), it is not phonemic in, e.g., Hungarian (Siptár & Törkenczy 2000) and Modern (Cypriot) Greek (Arvaniti 1999). We are also aware of the presence of [v:] in Russian (Dmitrieva 2017), where it surfaces through morpheme concatenation, and in varieties of Ingrian/Finnish (e.g., Sidorkevich 2014), where its phonemic status is unclear.

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

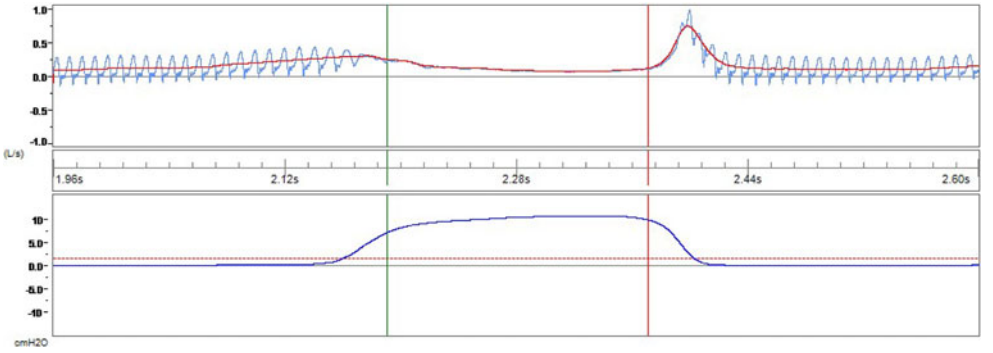

Figure 1. Airflow (top) and intraoral pressure (bottom) of [afa] measured in Aeroview using a Rothenberg mask, a MS-110 transducer electronic unit, and PT-2E and PT-25 pressure transducers, all by Glottal Enterprises, Inc. The displayed time window is 640 ms.

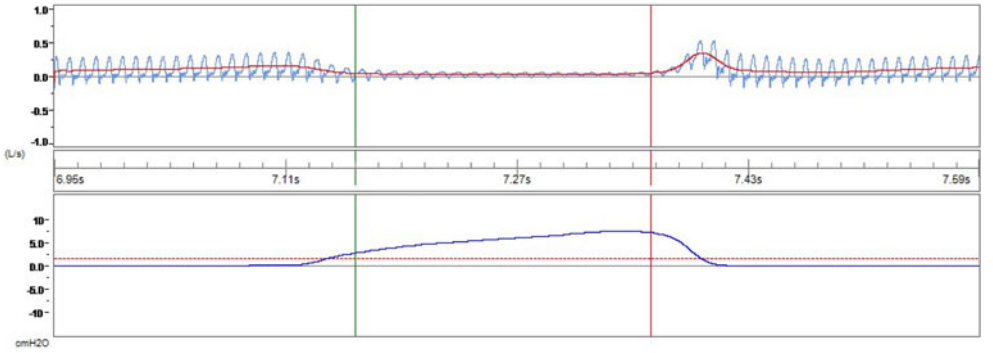

Figure 2. Airflow (top) and intraoral pressure (bottom) of [ava] measured in Aeroview using a Rothenberg mask, a MS-110 transducer electronic unit, and PT-2E and PT-25 pressure transducers, all by Glottal Enterprises, Inc. The displayed time window is 640 ms.

and visible glottal pulses (indicative of voicing) combined with a diffuse, relatively faint high-frequency noise across the spectrum (cf. Ladefoged & Johnson 2014: 212). However, based on the aerodynamic scenarios presented above, non-canonical realizations may include low-frequency periodicity with little or no accompanying mid- or high-frequency energy, suggestive of a more constricted labiodental articulation [ʏ] or a fully occluded realization [b̪] (a voiced labiodental plosive).[3] Alternatively, as also reported by Stevens et al. (1992) for American English and Jesus & Shadle (2002) for European Portuguese, the acoustic signal may show reduced low-frequency periodicity (voicing) alongside strong high-frequency energy (frication), particularly in the middle portion of the consonant, indicating a partially devoiced allophone [ʏ̥]. As explained in more detail below, these non-canonical realizations are more likely to surface for /vː/ due to its longer duration.

In the case of singleton /v/, speakers across languages may resort to a reduction in the degree of labiodental constriction, producing approximant variants [ʋ], particularly in intervocalic positions. This trend aligns with cross-linguistic tendencies of intervocalic consonant lenition, defined as a reduction in constriction degree (cf. Kirchner 2004).

[3] There is almost no cross-linguistic information about this type of stop, but see Ladefoged & Maddieson (1996: 17), and Olson & Hajek (2004: 51) who both make brief mention of it.

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

Intervocalic lenition is an active phonetic process in different regional varieties of Italian, as in a number of other languages. In Italian obstruents, it typically affects singleton but not geminate consonants (but see Giannelli & Savoia 1979 on geminate lenition in fast speech in Florentine) with patterns of application varying across regions. In northern Italy and Sardinia, lenition predominantly targets voiced obstruents (Canepari 1992), whereas in the centre-south, it may affect all singletons, and is associated with partial or complete voicing of voiceless stops, e.g., Roman /i ˈpini/ > [i ˈp̬i:ni] ('the pine trees') (Dian et al. 2024; Hualde & Nadeu 2012; Nocchi & Schmid 2008). In certain varieties spoken in and around Tuscany, central Italy, most notably Florentine and surrounding lects, lenition of intervocalic voiceless singleton stops through spirantization, typically without voicing especially for /p t/ > [ɸ θ], has even become phonologized – a phenomenon known as *gorgia toscana*, e.g., /i ˈpini/ > [i ˈɸi:ni] (Giannelli & Savoia 1979; Marotta 2008; Russo 2022). As for /v/, for which variants other than fricative [v] have not been reported in the Standard variety of Italian (e.g., Bertinetto & Loporcaro 2005), an approximant allophone, [ʋ], has been reported to occur intervocalically in some regional varieties (see § 1.2 below), as well as in various European languages including, e.g., Serbian (Bjorndahl 2022), Slovak (Hanulíková & Hamann 2010), Danish (Allan 2014), Faroese (Árnason 2011), and Valencian (Saborit Vilar 2009). In the specific case of Italian, this approximant realization also has the advantage of avoiding perceptual confusion with other phonemes, a risk posed by [ɣ̞] and [b̞], which may potentially be perceived as /f/ and /b/, respectively, leading to possible perceptual phonemic mergers.

Regarding geminate /v:/, the aerodynamic challenges already discussed are further intensified by the need to maintain the voicing-frication balance across a longer constriction period. Based on the aerodynamic processes referred to above, possible allophones of /v:/ may exhibit devoicing [ɣ̞̊:] or alternatively decreased or no visible frication noise, indicative of a narrower constriction [ɣ̞:] or complete occlusion [b̞:], which could also be followed by a release burst, characteristic of plosives. In contrast to /v/ lenition, these more constricted /v:/ forms can be considered as instances of articulatory strengthening, assuming that strengthening goes in the opposite direction of weakening, showing an *increase* in constriction degree (cf. the extensive discussion on this point by Bybee & Easterday 2019, who aptly argue that the labial environment is particularly conducive to strengthening in terms of articulatory overshooting). This aspect as well as the issue of devoicing are discussed more in detail in § 3.4. Additionally, approximant realizations of long /v:/ are also possible in principle but disfavoured in practice as evidenced by the cross-linguistic rarity of approximant geminates (Maddieson 2008; but see Celata et al. 2019, on a long approximant variant of /r:/ in Italian).

1.2. Previous experimental phonetic studies on /v, v:/

In an early investigation of Italian, Ferrero et al. (1979) noted a significant reduction of frication noise for singleton /v/, which was often not visible in spectrographic analysis, compared to other voiced fricative phonemes. This is in contrast with the findings for English and European Portuguese discussed in § 1.1 above, whereby frication noise was stronger in non-canonical forms. However, as Ladefoged & Maddieson (1996: 176) explain, frication noise reduction is also to be expected in cases where voicing is maintained during constriction due to the strong low-frequency energy stemming from voicing, which tends to mask the weaker high-frequency frication noise of /v/. This is partly consistent with [ɣ̞] (less constricted) realizations, on a continuum of intervocalic lenition between fricative [v] and approximant [ʋ] (or even deleted [ø]; see below).

Only a few studies have examined the allophonic variation of intervocalic /v/ in specific regional Italian varieties, particularly in the context of intervocalic lenition. For instance,

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

Marotta (2005) focused on Roman Italian, reporting a high incidence (68%) of the approximant allophone [ʋ] compared to canonical [v] (24%), along with some occurrences of segment deletion [ø] (8%). Marotta's corpus, consisting of a spontaneous conversation between two speakers, showed a short average duration of /v/ (55 ms), with [v] and [ʋ] forms pooled. An identical average value is reported in Dian et al. (2024) for /'CVCV/ words in a controlled corpus of target words embedded in carrier phrases for the same Roman variety. For Veneto Italian, spoken in the Veneto, in north-east Italy, Trumper & Maddalon (1982), as cited by Vietti (2019), identify the same [v ʋ ø] allophones, although no distributional data are available for this study. However, durational data (but not allophonic distributions) for Veneto Italian /v/ can be found in Dian et al. (2024), with average values ranging around 65 ms.

Pape & Jesus (2015) compared durational and devoicing patterns in voiced stops and fricatives, including /v/, in nonce /'CVCV/ words across European Portuguese, German, and Italian (again spoken by Veneto speakers) using six speakers per language. Their findings indicate that while intervocalic /v/ does not differ significantly in duration across the three languages, it tends to remain fully voiced only in Veneto Italian. Dian et al. (2024) report similar results using real words, examining Veneto and Roman Italian, where /v/ exhibited comparable durations and a lack of devoicing across both regional varieties. Both studies report average durations that are roughly aligned across languages and varieties within each study, although the values differ between studies, likely due to methodological differences.

Surprisingly, none of the studies we reviewed examine patterns of allophonic variation for geminate /vː/ in Italian or other languages. For instance, while Marotta (2005) provides durational data for both /v/ and /vː/, she reports allophonic distributions for the former (see above) but not the latter. Beyond Italian, we have to date found only one study that has investigated the spectral characteristics of /vː/: Shinohara & Fujimoto (2018: 259) report that for /vː/ in the Ikema dialect of Miyako Ryukyuan, 'pre-voicing without frication noise in the high-frequency area occurred with seven tokens out of eleven among five speakers', thus pointing to frequent occurrences of either [v̬ː] or [b̬ː]. This scarcity of phonetic studies on /vː/ is consistent with its presumed typological rarity across the world's languages, highlighting the need for further experimental investigation.

2. Aims and research questions

This study explores potential allophonic variation in the articulation of the contrastive pair /v/ and /vː/ in word-medial position in Italian, as observable in the acoustic signal. For the first time regarding these segments, a qualitative assessment of phonetic variation is complemented with a quantitative analysis using established acoustic measures of consonantal strength and phonetic voicing.

While previous research has documented considerable variability in the surface realization of /v/ within and across languages, much less is known about its long counterpart /vː/. This study also considers the potential influence of prosodic factors, such as position within the intonational phrase and relative to lexical stress. To assess the generalizability of the findings for Italian, speakers were selected from three geographically and dialectologically distinct areas of Italy (cf. Pellegrini 1977), chosen to reflect regional variation reported in previous studies (e.g., Crocco 2017; Vietti 2019): (i) central Veneto in the northeast, (ii) the city of Rome in the centre-south, and (iii) centro-southern Calabria in the far south.

Our research questions (RQs) are as follows:

1) What are the possible allophonic variants of intervocalic /v/ and /vː/ across regional varieties of Italian? Is there variation in their relative frequence of occurrence

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

triggered by (a) regional variety; (b) speaker sex; or (c) prosodic factors (i.e., post- vs. pre-stress, e.g., /ˈbeve/ vs. /doˈvere/, and phrase-final vs. phrase-medial positions)?

2) Can variants identified qualitatively through visual spectrographic inspection be reliably classified into consonantal strength categories? Specifically, do approximant realizations represent a weaker form than canonical fricatives, and, if present, do seemingly more constricted realizations represent a stronger form than canonical variants?
3) Do non-canonical variants of each length category (/v/ and /vː/) differ durationally from canonical forms?

Based on the reviewed literature, we predict:

a) Intervocalic singleton /v/ will surface as either canonical [v], approximant [ʋ], or deleted [ø] across varieties.
b) Intervocalic geminate /vː/ will surface as either canonical [vː], more constricted [v̞ː] or potentially even completely occluded [b̪ː] across varieties. Non-canonical realizations may also include some partial devoicing independent of the degree of supraglottal constriction (e.g., [v̞ː v̊ː b̪̊ː]), due to an expected tendency for long voiced obstruents to optionally devoice, as posited by the aerodynamic voicing constraint (AVC), e.g., Ohala (1983).
c) Lenited allophones will be shorter than canonical or more constricted forms, in line with the definition of intervocalic lenition as 'a reduction in constriction degree or duration' (Kirchner 2004).

As for potential variation due to speaker sex or prosodic factors we do not have previous studies that allow us to make predictions specific for /v, vː/.

3. Method

3.1. Participants

Eighteen native speakers of Italian (nine female, nine male) took part in the study. Participants were equally distributed across the three target regional varieties: six speakers of Veneto Italian (VI), six of Roman Italian (RI), and six of Calabrian Italian (CI). Speakers ranged in age from 27 to 68 years (mean age: 41.7 years), and none reported any speech, language, or hearing impairments. All participants were born and raised in their respective regions, where they had spent most of their lives. Specifically, all VI speakers and one CI speaker had lived in their home regions their entire lives, whereas all RI participants and five CI participants had spent part of their lives abroad and were either residing in or travelling through Melbourne, Australia, at the time of recording. This distribution of residential circumstances was due to constraints imposed by the COVID-19 pandemic, which limited travel and access to suitable recording environments in Italy during data collection. All participants reported daily use of Italian.

3.2. Materials and procedure

The target items were real disyllabic and trisyllabic Italian words featuring either a /ˈCVC(ː)V/ structure, with post-stress /C(ː)/ (namely *beve* /ˈbeve/ 's/he drinks' and *bevve* /ˈbevːe/ 's/he drank'), or a /CVˈC(ː)VCV/ structure, with pre-stress /C(ː)/ (*dovere* /doˈvere/ 'duty' and *davvero* /daˈvːero/ 'really'). Thus, all items contained the consonants /v/ or /vː/

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

in word-medial, intervocalic position. These words were embedded in carrier sentences designed to manipulate their position within the intonational phrase. Two frame sentences were used: *Ho detto WORD* ('I said WORD') for the phrase-final condition, and *Ho detto WORD prima* ('I said WORD before') for the phrase-medial condition. To minimize a potential confounding effect of nuclear placement, target words were produced with nuclear focus. This was elicited by instructing speakers to respond as if answering the questions *Che hai detto?* ('What did you say?') and *Che hai detto prima?* ('What did you say before?'), respectively. Each speaker produced four repetitions of each target word in both phrase conditions. Items were presented visually on a laptop screen using Microsoft PowerPoint, with presentation order randomized across participants. Participants were instructed to read each sentence aloud at a natural pace and speaking style. Filler words were inserted among target tokens at the start of each recording session to familiarize participants with the task. In total, each of the eighteen participants produced 32 target tokens (two phonemes × four repetitions × two stress conditions × two phrase conditions), yielding 575 tokens overall after excluding one disfluent production.

3.3. Recording protocol

The studio recordings in Melbourne were conducted at the Horwood Recording Studio, part of the University of Melbourne, using a Charter Oak E700 dual-diaphragm solid-state condenser microphone and a Focusrite Scarlett 18i20 (3rd generation) audio interface. Field recordings in Italy were carried out in quiet, softly furnished rooms in the speakers' home environments, ideal to reduce echo, using a ZOOM H1n Handy solid-state recorder. In both recording environments, the microphone was positioned at approximately 20 cm from the speakers' mouth. All speech was digitized at a sampling rate of 44.1 kHz with 16-bit quantization and saved in WAV format. Recordings were continuously monitored to ensure clarity; any trials containing disfluencies, hesitation, or background noise were discarded and re-recorded as needed.

3.4. Acoustic analysis

Segmentation and labelling of the target consonants were carried out manually using the EMU Speech Database Management System (EMU-SDMS; Winkelmann et al. 2017). Segment boundaries were defined based on visual inspection of both the waveform and spectrogram, with procedures varying according to the phonemes' acoustic realization. For fricative variants, the onset and offset of the target segment were placed at the visible boundaries of frication noise, identified by the presence of mid-to-high frequency energy. For approximant realizations, boundaries were placed at points on either side of the segment where a dip in the amplitude envelope was observed in the waveform, accompanied by a slight change in formant structure in the spectrogram. For plosive allophones, the onset was marked at the beginning of the closure – typically indicated by a drop in waveform amplitude and a loss of formant structure – while the offset was placed at the onset of modal voicing in the subsequent vowel, following any post-release interval, as seen in both waveform and spectrogram. Examples of these realizations are shown in § 4.1 below.

For each token, the duration of the target consonant was measured in milliseconds. All tokens were classified visually into nine allophonic categories, four for /v/ ([v], [v̞], [ʋ], and [b̪]) and five for /vː/ ([vː], [v̞ː], [vːˑ], [b̪ː], and [b̪ːˑ]). As these allophones were unevenly distributed, they were subsequently grouped into three broader 'strength' classes – *canonical*, *weakened*, and *strengthened* based on their surface manner of articulation and voicing status – to ensure analytical robustness. For manner, this classification follows the assumption

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

of a consonantal strength hierarchy, in which stops are considered 'stronger' than fricatives, and fricatives 'stronger' than approximants (Honeybone 2008; Ewen & van der Hulst 2001; Lass 1984; Escure 1977; Foley 1977). As for voicing, categorizing devoiced realizations as either weakened or strengthened is less straightforward, as this remains a matter of debate. From a phonological perspective, turning a voiced obstruent into a voiceless one in intervocalic position – a context typically associated with lenition – has traditionally been regarded as a case of strengthening (e.g., Lass 1984). Phonetically, this has been attributed to more favourable aerodynamic conditions for voicing in this environment (Westbury & Keating 1986). However, as noted in § 1.1, the aerodynamic conditions for sustaining voicing in geminates are unfavourable. On this basis, authors such as Kirchner (1998) and Bauer (2008) have argued that maintaining voicing in geminates entails strengthening and therefore classify devoiced variants as weakened realizations. We remain agnostic as to whether devoiced forms are cases of weakening or strengthening from a theoretical standpoint. For the purposes of analysis here, however, we treat these variants as strengthened, based on the results of the preliminary acoustic analysis in § 4.3. In order to distinguish devoiced from voiced allophones objectively, we used REAPER (Talkin 2015) via MacReaper (Dallaston & West 2018) to calculate the proportion of glottal closure instants (GCIs) – indicative of voicing – within the constriction duration. Tokens with 50% voicing or less following a visual check were classified as devoiced, and those with more than 50% as voiced, following Abramson & Whalen (2017). As a result, the *canonical* group consisted of [v(ː)] realizations; the *weakened* group included [ʋ]; the *strengthened* group encompassed [vː], [b̪(ː)], and devoiced [v̥(ː)] and [b̪ː] forms.

As an additional preliminary step, we sought to verify quantitatively whether the allophones we identified fitted within each aforementioned strength category. To this end, we used well-established acoustic metrics of lenition, namely *IntDiff* and *MaxVel* (cf. Hualde & Nadeu 2012, among others, originally adapted from Kingston 2008). Following the methodology in Ennever et al. (2017), after a first check using raw files, the sound files were band-pass filtered between 400 and 1200 Hz for this analysis, as this approach yielded more meaningful results than using unfiltered files. *IntDiff* refers to the raw difference between the minimum amplitude (or intensity) in the consonant and the maximum amplitude in the following vowel, measured in dB. The lower *IntDiff*, the less constricted (and therefore the more weakened) the obstruent. *MaxVel* relates to the maximum velocity of the amplitude contour between consonant and vowel. It is calculated as the highest first derivative – that is, the steepest rate of change – in the amplitude contour (in dB/ms) within the analysis window. Again, the lower *MaxVel*, the more weakened the obstruent. We used the *rmsana* function, part of the *R* package *wrassp* (Winkelmann et al. 2024), to obtain RMS (root mean square) amplitude contours. Consistent with Ennever et al. (2017), we set the window size to 10 ms and the shift to 2.5 ms. To compute *MaxVel*, we divided the amplitude difference between two consecutive measurement steps by the step duration (2.5 ms) and identified the maximum positive value within points of consonant amplitude minima and vowel maxima.

3.5. Statistical analysis

We conducted three statistical analyses in R (R Core Team 2025). The first aimed to verify the grouping of allophones into three strength categories (cf. RQ2); the second modelled the likelihood of occurrence of each of these strength categories (RQ1); the third looked at the effect of length and strength on consonant duration (RQ3). For methodological consistency, all analyses used Bayesian regression via the *brms* package (Bürkner 2017). This choice was motivated by the suitability of this method for the second analysis, as explained more in detail below.

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

For the first analysis, we set up two models, one with *IntDiff* ('M1a') and one with *MaxVel* ('M1b') as the dependent variable. The only fixed factor was *allophone*, with nine levels corresponding to the distinct segmental realizations observed in the corpus: [ʊ], [v], [vː], [b̪], [vˑ], [v̥], [b̪ː], [v̥ː], and [b̥ː]. As a very small number of tokens were available for devoiced singleton [v̥] (only one token) and singleton plosive [b̪] (two tokens), these categories were collapsed with [v] for the purposes of statistical modelling. For each dependent variable, we fitted a Bayesian linear mixed-effects model with *allophone* as a fixed effect and *speaker* as a random intercept to account for repeated measures, yielding the formulae *IntDiff* ~ *allophone* + *(1 | speaker)* and *MaxVel* ~ *allophone* + *(1 | speaker)*. Models assumed Gaussian likelihoods with weakly informative default priors on the intercept and slope coefficients; priors on standard deviations were half-*Student-t* distributed. Four Markov chains were run for 4,000 iterations each, with the first 2,000 iterations used as warm-up, yielding 8,000 post-warmup samples. Convergence was assessed via the potential scale reduction factor ($\hat{R} < 1.01$) and visual inspection of trace plots. Posterior estimates for each *allophone* level were obtained using the *emmeans* package (Lenth et al. 2025), yielding estimated marginal means (EMMs) with 95% credible intervals (CrIs) based on posterior draws. Pairwise comparisons between all levels were computed on the posterior samples, and differences were deemed *credible*[4] if their 95% CrI did not cross or only marginally crossed zero. Where relevant, we also report Pr(Δ>0), or the posterior probability that the difference is greater than zero.

For the second analysis, to model the likelihood of occurrence of each 'strength' category, two categorical logistic regression models – one for /v/ and one for /vː/ – were fitted using Bayesian mixed-effects methods. These are referred to here as models 'M2v' and 'M2vv', respectively. Bayesian methods have been increasingly adopted in phonetic studies for modelling categorical outcomes with more than two levels (e.g., Dilley et al. 2019; Kim 2022), as traditional frequentist approaches often face limitations in estimating multinomial logistic regression models, particularly with complex random effects structures. Following recommendations by Vasishth et al. (2018) and Nalborczyk et al. (2019), as well as the methodology in Flego & Forrest (2021), the *brms* package was chosen over alternatives such as *MCMCglmm* (Hadfield 2010) due to its greater flexibility in model specification and its intuitive formula syntax, which mirrors that of the more widely used *lmerTest* and *lme4* packages (Kuznetsova et al. 2017). Separate models were fitted for /v/ and /vː/ to account for potential structural and distributional differences in the realization of these phonemes, including the fact that they exhibit distinct sets of allophones and may be differentially affected by the tested factors. The /v/ model was specified as a multinomial logistic regression with three outcome categories (canonical, strengthened, weakened), with 'canonical' set as the reference category. The /vː/ model was specified as a Bernoulli logistic regression with two outcome categories (canonical, strengthened), with 'strengthened' coded as the success level. Each model included fixed effects of *regional variety* (CI, RI, VI), *stress condition* (post-stress, pre-stress), *phrase position* (phrase-medial, phrase-final), and *speaker sex* (female, male), with random intercepts for speaker and lexical item. The formula was *strength* ~ *variety* + *stress condition* + *position* + *sex* + *(1 | speaker)* + *(1 | word)* for each model. Four Markov chains were run for 4000 iterations each, with *adapt_delta* set to 0.995 and max_treedepth to 15 to ensure model convergence. Posterior estimates were visualized using conditional effects plots, and fitted probabilities were extracted to compute the average likelihood of each allophonic category across conditions. For both M2v and M2vv, we specified weakly informative priors to ensure convergence: intercepts were given

[4] In the Bayesian framework, results are described as either 'credible' or 'unclear', which is broadly analogous to 'significant' and 'non-significant' in the frequentist framework, albeit based on different underlying principles (see, e.g., Vasishth et al. 2018). We adopt this terminology throughout the paper.

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

Student-t(3, 0, 2.5) distributions, population-level effects *Normal(0, 1)*, and group-level standard deviations *Exponential(1)*. For the multinomial model, priors were applied separately to the strengthened vs. canonical and weakened vs. canonical contrasts. Four Markov chains were run for 4,000 iterations each, with the first 2,000 iterations used as warm-up, yielding 8,000 post-warmup samples. Convergence was assessed via the potential scale reduction factor ($\hat{R} < 1.01$) and visual inspection of trace plots.

For the third analysis, consonant duration (*Cdur*, in milliseconds) was analyzed using a Bayesian linear mixed-effects model ('M3'). The fixed effects were *length* (singleton, geminate), *strength* (canonical, weakened, strengthened), *phrase position* (phrase-medial, phrase-final), *stress condition* (post-stress, pre-stress), *sex* (female, male), and *regional variety* (CI, RI, VI). Two- and three-way interactions between *length*, *strength*, and *phrase position* were also included to account for the combined effects of these factors on allophonic distribution found in the second analysis above (see § 4.4). Random intercepts were specified for *speaker* and *word*, with by-speaker random slopes for *length* to account for inter-speaker variability in the length contrast (see formula in Table 4). The model was fitted with a Gaussian likelihood. To set a weakly informative prior on the intercept centred on the observed scale of the data, the empirical mean (μ) and standard deviation (σ) of *Cdur* were computed from the dataset and inlined directly into the prior specification for reproducibility *(Normal(μ, σ))*. Priors on fixed effects were *Normal(0, 50)*, while priors on random-effect standard deviations and the residual standard deviation were *Student-t(3, 0, 40)*. Correlation parameters were given an *LKJ(2)* prior. These priors regularize estimates on the original millisecond scale without imposing strong constraints, ensuring that posterior distributions were primarily data-driven. Four Markov chains were run for 4,000 iterations each (1,000 warm-up, 3,000 post-warmup), yielding 12,000 posterior samples. Sampling control parameters were set to *adapt_delta* = 0.98 and *max_treedepth* = 12 to avoid divergent transitions and ensure efficient exploration of the posterior. Convergence was assessed via the potential scale reduction factor ($\hat{R} < 1.01$), effective sample size, and visual inspection of trace plots. Posterior predictive checks (*pp_check*) confirmed that the fitted model adequately captured the observed distribution of *Cdur*. To evaluate group-level effects of categorical predictors, we obtained posterior predictions for representative conditions directly from the fitted model (using *posterior_epred*). From these, we computed posterior means, 95% credible intervals, and pairwise contrasts (expressed as differences of posterior predictions).

4. Results

This section begins with illustrative examples of the allophonic variants of /v/ and /v:/ (§ 4.1), followed by a descriptive overview of their distribution across the three strength categories (§ 4.2). The results are then presented in three parts: (i) a preliminary statistical analysis validating the categorization of variants into strength classes (§ 4.3), (ii) a statistical analysis of categorical strength distributions (§ 4.4), and (iii) a separate analysis of the duration patterns associated with each strength class (§ 4.5).

4.1. Illustrations of variants

RQ1 asked what allophones of /v/ and /v:/ occurred in the data and how they were distributed across target conditions. Before turning to a quantitative analysis, Figures 3 and 4 illustrate representative examples of allophonic variants observed across the corpus. As described above, canonical [v(:)] realizations exhibit low-frequency periodicity in the waveform and a visible voice bar across the constriction, but no formant structure, with

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

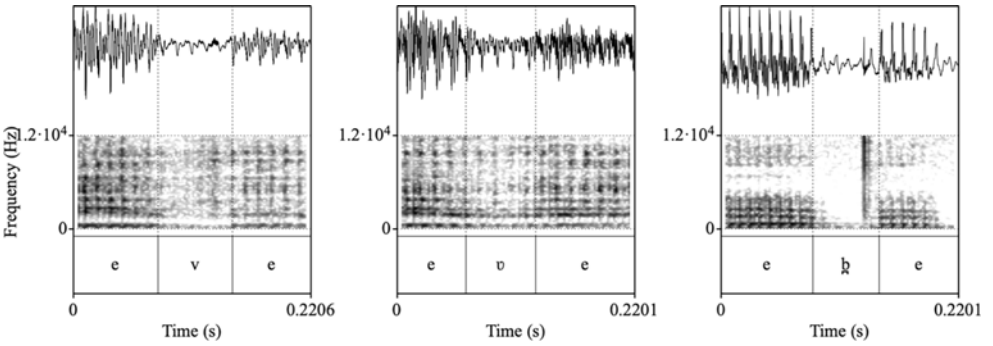

Figure 3. Annotated examples of allophonic variants of /v/ embedded in /ˈbeve/ ('s/he drinks') from a VI (left and middle panel) and a CI (right panel) speaker.

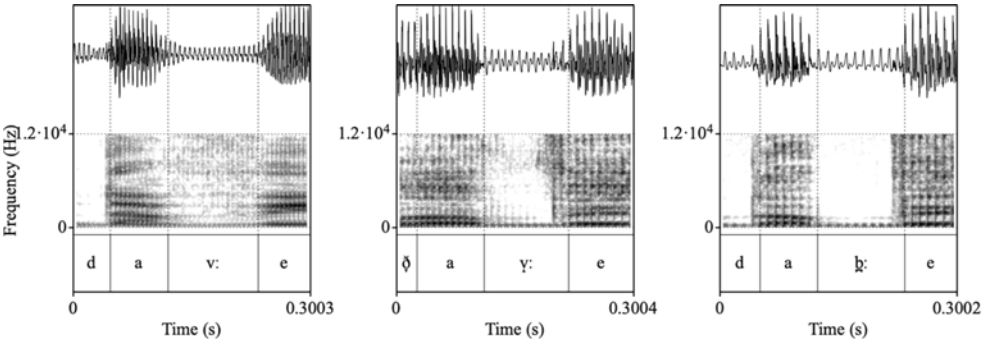

Figure 4. Annotated examples of allophonic variants of /v:/ embedded in /daˈvːero/ ('really'), each from a different RI speaker.

frication noise present across the frequency spectrum. Approximant [ʋ] variants show a lower-amplitude, vowel-like waveform, with visible lower formants (F1 and F2) during the constriction phase. Plosive [b̪(ː)] realizations are characterized by low-frequency periodicity, a clear voice bar, absence of frication noise during constriction, a burst at release, and a short post-release phase. Finally, more constricted [ʏ̞(ː)] variants typically display a voice bar and weaker frication noise than canonical forms, with noise energy concentrated near the offset of constriction, reflecting a release of narrowing rather than a complete articulatory closure. In addition, a few devoiced fricative [ʏ̥(ː)] and plosive [b̪̊ː] realizations were also noted.

4.2. Descriptive summary of allophonic distributions

Table 1 details the allophonic distributions of /v/ and /v:/ across pre- and post-stress tokens, with regional varieties pooled. For singleton /v/, canonical short voiced fricative [v] represents only 52% (149/288) of the tokens, distributed similarly across stress conditions, while a notable 47% (136/288) of cases were realized as approximant [ʋ], a very small fraction (2/288, or 1%) as plosive [b̪], and in only one case as a devoiced [ʏ̥]. We found no instances of full deletion [ø]. For geminate /v:/, 59% (170/287) of total tokens surfaced as canonical [v:], while 36% (103/287) of occurrences showed increased constriction, with 19% (54/287)

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

Table 1. Overall counts and rates of occurrence (%) of allophonic variants for /v/ and /v:/ in the corpus.

	Stress	[ø]	[ʋ]	[v(:)]	[ʏ̞(:)]	[ʏ̟(:)]	[b̥̞(:)]	[b̥̟(:)]	tot
/v/	Post-	–	67 (47%)	74 (51%)	1 (1%)	–	–	2 (1%)	144
	Pre-	–	69 (48%)	75 (52%)	–	–	–	–	144
	All	–	136 (47%)	149 (52%)	1 (~0%)	–	–	2 (1%)	288
/v:/	Post-	–	–	85 (59 %)	9 (6%)	28 (20%)	6 (4%)	15 (11%)	143
	Pre-	–	–	85 (59%)	5 (4%)	26 (18%)	–	28 (19%)	144
	All	–	–	170 (59%)	14 (5%)	54 (19%)	6 (2%)	43 (15%)	287

Variants are indicated with decreasing degree of articulatory lenition from left to right.

Table 2. Distribution of allophones by strength category for /v/ and /v:/.

	/v/		/v:/	
Canonical	[v]	149/288 (51.7%)	[v:]	170/287 (59.2%)
Weakened	[ʋ]	136/288 (47.2%)	–	
Strengthened	[ʏ̞], [b̥̟]	3/288 (1.1%)	[ʏ̟:], [ʏ̞:], [b̥̟:], [b̥̞:]	117/287 (40.8%)

realized as more constricted fricative [ʏ̟:] and 17% (49/287) as plosive [b̥̟:]. 20 out of 287 (7%) /v:/ tokens showed substantial devoicing: 14 surfaced as long devoiced fricatives [ʏ̞:] and 6 as plosive-like [b̥̞:].

As outlined in Section 3.4, the allophones listed in Table 1 were tentatively grouped into three 'strength' categories for analysis. Table 2 presents the distribution of individual allophones within each category for /v/ and /v:/.

4.3. Statistical results 1 – Diagnostic analysis of variant categorization

Moving on to RQ2, after an initial qualitative grouping of variants, we now assess whether these differ in their *IntDiff* and *MaxVel* values (cf. § 3.4) according to their qualitatively assessed strength category. We provide descriptive plots showing raw data distributions overlaid with estimated marginal means (EMMs) with 95% credible intervals (CrIs), ordered by increasing EMM.

For *IntDiff*, model M1a revealed systematic differences between allophones on this parameter, as evidenced by posterior estimates of marginal means reported in the Appendix (Table A1). Figure 5 illustrates a visual summary of the results. The approximant singleton [ʋ] has the lowest mean (16.57 dB), followed by the canonical singleton [v] (25.63 dB), with a statistically credible difference of 9.03 dB (cf. Table A1). The canonical geminate [v:] ranks next (36.74 dB), differing credibly from [v] by +11.11 dB. The devoiced geminate [ʏ̞:] shows a marginally higher EMM than [v:] (37.78 dB), with a non-credible difference of +1.04 dB, suggesting these two variants may be grouped together in terms of *IntDiff* patterns. The devoiced geminate plosive [b̥̟:] follows (42.30 dB), exceeding [ʏ̞:] by +4.53 dB, although not credibly, likely due to its wider CrI range resulting from lower token frequency. Its more frequent voiced counterpart [b̥̟:] (42.50 dB) is, however, credibly higher than [ʏ̞:] (+4.83 dB). The highest EMM is observed for [ʏ̟:] (43.95 dB), which does not differ credibly from either [b̥̟:] or [b̥̞:].

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

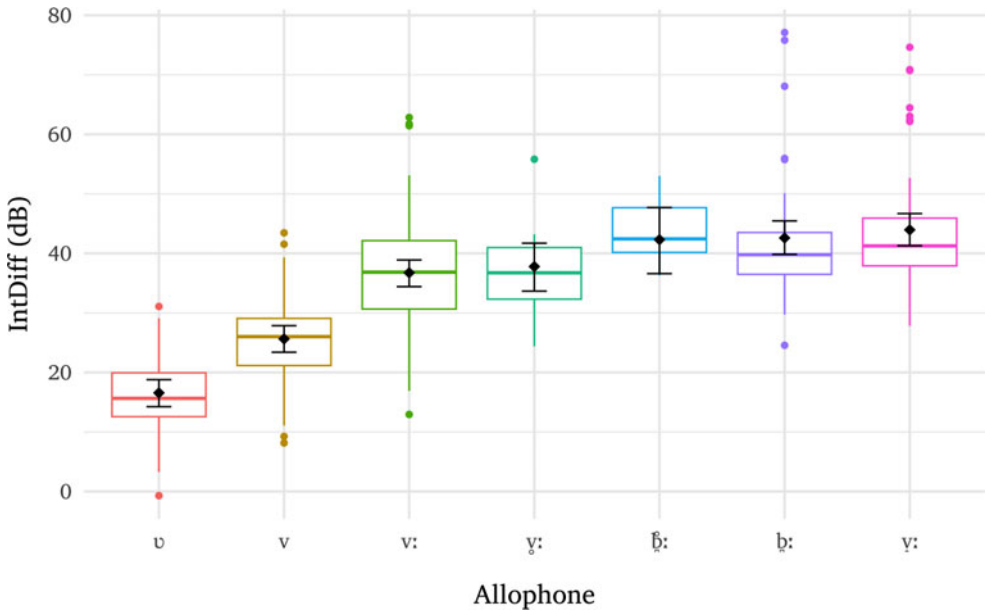

Figure 5. Raw *IntDiff* distributions for each *allophone*, overlaid with model-predicted means (black diamonds) and corresponding 95% CrIs (error bars).

For *MaxVel*, model M1b also indicated systematic differences between allophones (cf. Table A2 in the Appendix), which are summarized in Figure 6. The EMMs increase from the approximant singleton [ʋ] (1.42 dB/ms; lowest) to the canonical singleton [v] (1.87 dB/ms; +0.45, a credible difference), then to the canonical geminate [vː] (2.17 dB/ms; +0.30 compared to [v], again a credible difference). A mid-range cluster follows – [b̪ː] and [yː̥] – whose EMMs are equivalent (both at 2.26 dB/ms) and both credibly different (+0.39) from [v] but not from [vː]. Above these is the more constricted geminate [yː], showing a higher EMM (2.56 dB/ms; +0.30), which is, however, not statistically different from [yː̥] but is different from [b̪ː]. Finally, the devoiced geminate plosive [b̪̊ː] is highest overall (2.74 dB/ms). Albeit not credibly different from [yː] or [yː̥], it is statistically different from [b̪ː] (+0.48) and preceding variants.

Overall, the results for both *IntDiff* and *MaxVel* support the view that approximants in this study are weaker than canonical fricatives in singleton form, and that canonical geminates tend to be weaker than more constricted variants. For singletons, [ʋ] consistently shows lower values than [v], with credible differences for both measures. Among geminates, [vː] yields lower values than more constricted categories, with [yː̥] and the plosive variants ([b̪ː], [b̪̊ː]) generally occupying the upper end of the scale. These results corroborate the categorization of [ʋ] as weakened, [v, vː] as canonical, and [yː, b̪ː, b̪̊ː] as strengthened forms, at least from an acoustic viewpoint. The classification of devoiced geminate [yː̥] is less straightforward, as it patterns with the canonical variant in terms of *IntDiff* and also with more constricted ones in terms of *MaxVel*. Since there is some indication of increased strength for this allophone, and for consistency with [b̪̊ː], we include it as part of the strengthened category. Finally, we could not test whether the very few singleton tokens showing increased constriction can be classified as strengthened, but we will assume this to be the case as raw means for [b̪] (32.16 dB for *IntDiff* and 3.07 dB/ms for *MaxVel*) are

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

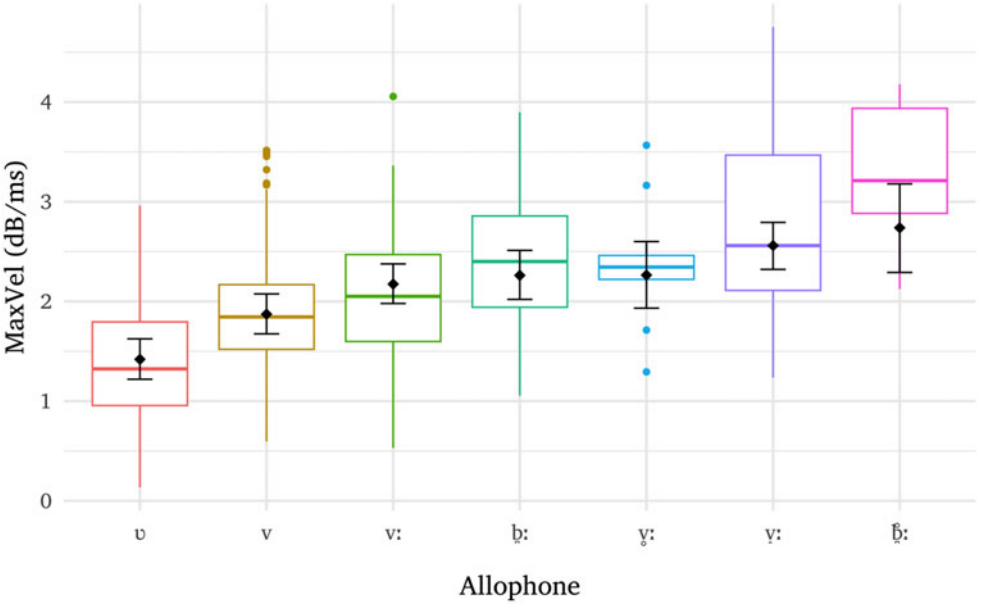

Figure 6. Raw *MaxVel* distributions for each *allophone*, overlaid with model-predicted means (black diamonds) and corresponding 95% CrIs (error bars).

considerably higher than those reported above for [v] (25.73 dB and 1.87 dB/ms). As for devoiced [ɣ̥], we believe that a single token does not allow any type of categorization.

4.4. Statistical results 2 – Allophonic strength distributions

We now return to RQ1, having verified the classification of allophonic variants into their respective strength categories. Table 3 shows the results of model M2v for /v/ and model M2vv for /vː/ combined (cf. § 3.5). Since the distributions presented in Table 2 show a strong tendency for /v/ to exhibit either canonical or weakened forms, and no weakened forms for /vː/, we focus our interpretation on the canonical vs. weakened contrast for /v/, and on the canonical vs. strengthened contrast for /vː/. Estimates for the less frequent categories, which yielded unstable or non-credible effects, are not discussed here.

As can be seen, there is no statistically credible effect of *regional variety* or *stress condition* on either model, suggesting that these factors do not influence the allophonic distributions in the data. Conversely, *position in the phrase* has an effect on /v/ and *speaker sex* exerts a strong influence on /vː/. Specifically, phrase-medial tokens of /v/ are less likely to be weakened, while male speakers are considerably more likely than females to produce strengthened variants of /vː/. The posterior probability that the effect of sex on /vː/ is positive is 0.97, corresponding to a ∼40 percentage-point increase in the probability of strengthened realizations for males relative to females. Each of these phonemes is addressed below.

For /v/, Table 3 shows that *position in the phrase* is the only significant predictor. Specifically, /v/ in phrase-final words tended to be realized as a weaker [ʊ] more often than in phrase-medial positions, with model-estimated probabilities of weaker realizations equalling 56.2% for the former and 38.4% for the latter. This effect of *position in the phrase* on

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

Table 3. Bayesian multinomial and Bernoulli logistic regression (models M2v and M2vv) results. Effects whose 95% CrI exclude zero are shown in bold.

Predictor	/v/ Estimate (log-odds)	/v/ 95% CrI	/vː/ Estimate (log-odds)	/vː/ 95% CrI	Interpretation
Intercept	0.36	[−1.50, 2.20]	−1.64	[−3.54, 0.43]	–
Variety (CI – RI)	0.13	[−1.32, 1.55]	0.43	[−1.06, 1.88]	–
Variety (CI – VI)	−0.54	[−1.98, 0.91]	−0.08	[−1.55, 1.38]	–
Stress (post – pre)	0.04	[−1.23, 1.32]	0.03	[−1.33, 1.33]	–
Position(PF – PM)	**−1.12**	[**−1.71, −0.54**]	0.24	[−0.38, 0.85]	↓ /v/ weakening in phrase-medial position
Sex (F – M)	0.19	[−1.14, 1.55]	**1.49**	[**−0.01, 2.87**]	↑ /vː/ strengthening for male speakers (Pr > 0 = 0.97)

CrI = credible interval. Fixed effects estimates from Bayesian multinomial logistic regression models for /v/ (weakened vs. canonical) and /vː/ (strengthened vs. canonical). The reference level for both models is 'canonical'. Credible effects (95% CrI excludes or only marginally touches 0) are shown in bold. Only the contrasts relevant to each phoneme are reported.

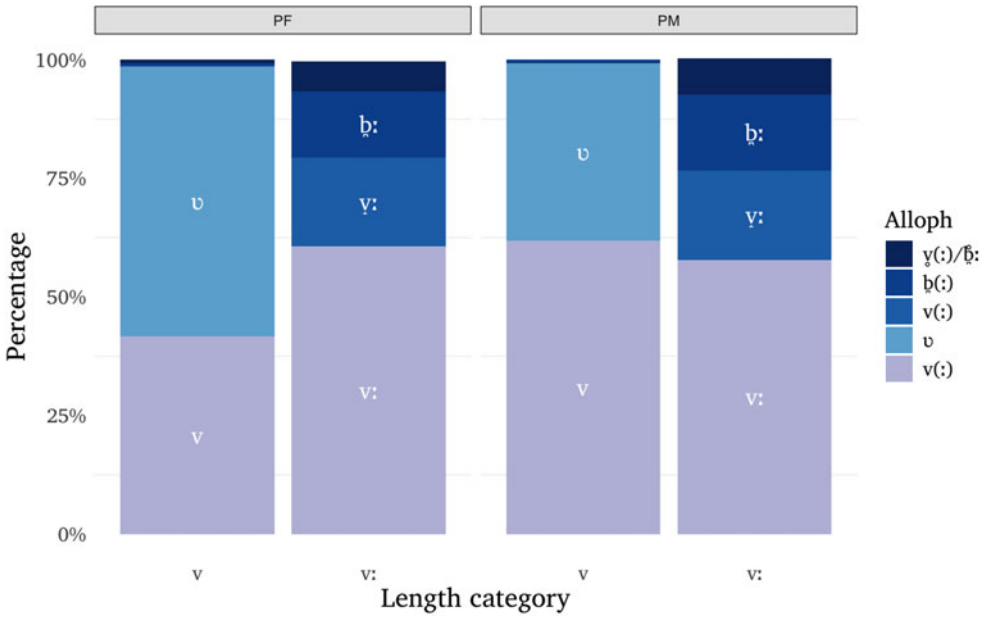

Figure 7. Distribution of allophonic variants of intervocalic /v/ and /vː/ across phrase-final (PF) and phrase-medial (PM) positions, expressed as percentages. All other conditions are pooled.

the allophonic variation of /v/ can be observed in Figure 7 showing raw data, while the same figure shows that distributions are similar for /vː/ across phrase-final and phrase-medial categories.

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

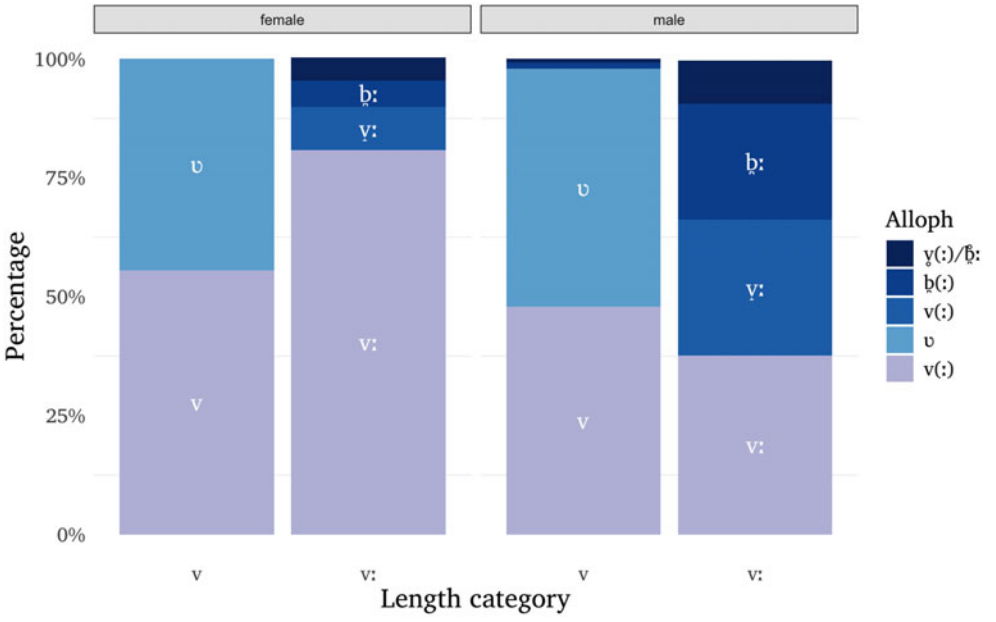

Figure 8. Distribution of allophonic variants of intervocalic /v/ and /v:/ across female and male speakers, expressed as percentages. All other conditions are pooled.

For /v:/, Table 3 demonstrates that the only significant effect among the tested factors is that of *speaker sex*, with a much higher predicted probability for male speakers to produce strengthened forms of /v:/ (61.3%) as compared to female speakers (20.5%). This pattern for /v:/ is clearly shown in Figure 8, while for /v/ the distributions do not differ statistically between male and female speakers.

4.5. Statistical results 3 – durational patterns

RQ3 addressed whether non-canonical 'strength' variants differ durationally from canonical forms. As shown in Table 4, Model M3 (cf. § 3.5) found *length* to be the only credible predictor: posterior predictions indicated that /v:/ was longer than /v/, with an estimated average difference of 75 ms across speakers and conditions (/v:/: 138.5 ms, 95% CrI [105.8, 166.9]; /v/: 63.9 ms, 95% CrI [36.1, 95.9]; $\Delta = 74.6$ ms, 95% CrI [38.3, 102.5], $Pr(\Delta > 0) = 0.995$). None of the other predictors yielded credible effects on duration, either as main factors or in interactions.

This robust effect of *length* is shown in Figure 9, plotting raw data: /v:/ is consistently longer than /v/ on average across all strength categories. By contrast, no clear effect of consonantal strength emerges. Even the comparison between canonical [v] and weakened [ʋ] reveals only a marginal difference, with average durations of 64 ms (standard deviation = 14 ms) and 53 ms (standard deviation = 12 ms), respectively. In other words, whether /v/ or /v:/ undergo strengthening or lenition does not appear to influence their duration. This pattern holds both within and across gemination categories.

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

Table 4. Bayesian linear mixed-effects regression (model M3) results. Effects whose 95% credible intervals exclude zero are shown in bold.

Predictor	Estimate	Est. Error	95% CrI
Intercept	96.04	17.49	[61.54, 130.11]
Length (singleton – geminate)	**−32.66**	**16.63**	**[–64.93, –0.62]**
Strength1 (canonical – weakened)	0.40	15.17	[–29.35, 29.92]
Strength2 (canonical – strengthened)	−9.75	30.20	[–69.08, 49.47]
Position (PF – PM)	2.81	14.50	[–26.01, 30.93]
Stress (post – pre)	1.32	7.69	[–15.38, 18.03]
Sex (F – M)	4.26	3.28	[–2.11, 10.78]
Variety1 (CI – RI)	−0.58	4.05	[–8.72, 7.24]
Variety2 (CI – VI)	−4.77	4.07	[–12.83, 3.37]
Length × Strength1	−1.98	15.17	[–31.67, 27.72]
Length × Strength2	3.37	30.18	[–55.75, 62.98]
Length × Position	−2.18	14.51	[–30.56, 26.60]
Strength1 × Position	1.14	14.51	[–27.03, 29.99]
Strength2 × Position	0.01	28.88	[–57.53, 56.35]
Length × Strength1 × Position	−0.70	14.50	[–29.38, 27.64]
Length × Strength2 × Position	0.23	28.88	[–56.00, 57.56]

Formula: C duration ~ length * strength * position + stress condition + speaker sex + regional variety + (1 + length | speaker) + (1 | word). CrI = credible interval; PM = phrase medial; PF = phrase final. Predictors are coded with treatment contrasts. Reference levels are: singleton (gemination), canonical (strength), phrase-final (position), post-stress (stress condition), female (sex), and CI (regional variety).

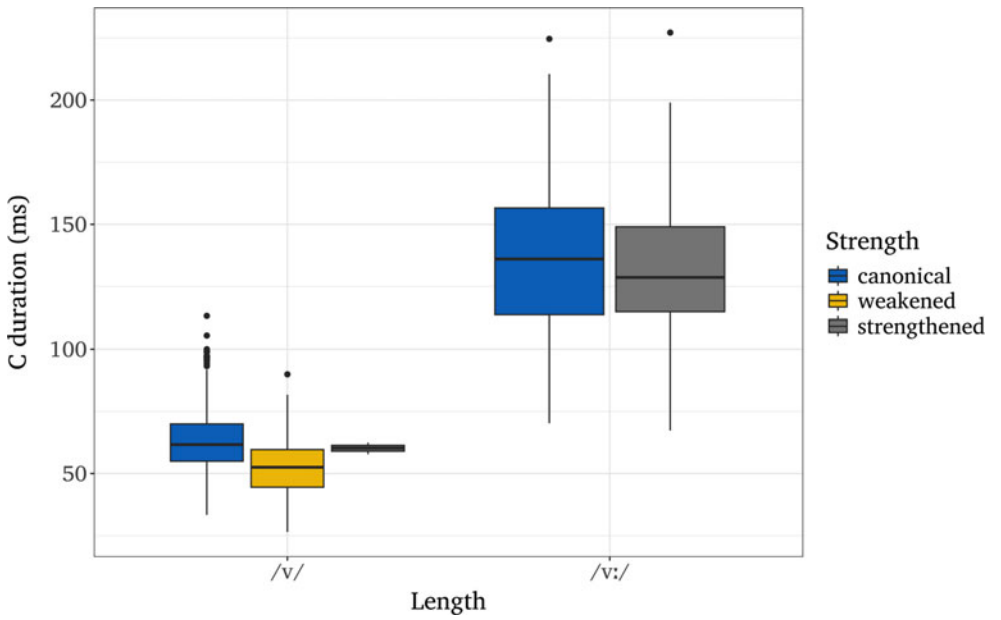

Figure 9. C duration distributions by consonantal *length* and *strength*. All other conditions are pooled.

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

5. Discussion

This study investigated the allophonic distribution and durational patterns of intervocalic /v/ and /v:/ – two segments that are known to pose articulatory challenges – in three regional varieties of Italian. While considerable allophonic variation linked to articulatory weakening has previously been reported for singleton /v/ in intervocalic position, little is known about its geminate counterpart /v:/, which remains phonetically underexplored cross-linguistically, likely due to its typological rarity.

In relation to Research Question 1 (RQ1), concerning the allophonic variation of these phonemes, the results show that intervocalic /v/ is frequently realized as an approximant [ʋ], accounting for nearly half of all tokens across the speaker sample with no significant cross-regional differences in this proportion, while it surfaced as a plosive in only two cases and as a devoiced fricative in only one instance. These findings are largely in line with our predictions in § 2, although no deletions occurred. Moreover, the plosive and devoiced variants, despite being infrequent, were unexpected for short /v/. We interpret these few occurrences are possible cases of overshooting (cf. § 1.1) due to hyperarticulation given the controlled nature of the experimental task.

The frequent occurrence of [ʋ] is consistent with earlier studies of specific regional varieties of Italian (e.g., Marotta 2005; Trumper & Maddalon 1982) but the present study expands on these by showing that frequent approximant realizations are not confined to a particular dialectal area. Rather, they appear to be a more general feature of Italian speech, potentially reflecting a widespread lenition process affecting most voiced singleton obstruent phonemes in intervocalic contexts (cf. for instance Hualde & Nadeu 2012 on Roman Italian). This finding is also consistent with the observation by Ferrero et al. (1979) of reduced frication noise in the production of /v/ in Italian.

Among the tested predictors, only position within the intonational phrase had a significant effect on /v/: approximant forms were more frequent when target words were uttered in phrase-final than in phrase-medial position. Although this finding might seem surprising under the assumption of phrase-final lengthening (cf. Cho 2016; Fletcher 2010), which in turn should lead to less frequent segment lenition (e.g., Katz & Pitzanti 2019), our data show that phrase-final tokens are not significantly longer than phrase-medial ones. This can be accounted for by the observation that phrase-final lengthening predominantly affects the rhyme portion (i.e., the nucleus and any coda) of syllables directly adjacent to the phrase boundary and of stressed syllables (Turk & Shattuck-Hufnagel 2007; also cf. Petrone et al. 2014 for Italian specifically), whereas the target /v/ segments in the present study always occurred in onset position (i.e., /ˈbe.ve/, /doˈve.re/).

The behaviour of geminate /v:/ showed a degree of articulatory variability that, while not entirely unexpected (cf. § 2), was nonetheless substantial. Although the majority of tokens were produced as canonical [v:], over one-third displayed increased constriction, with a notable 17% across all tokens realized as labiodental plosives, with visible occlusion, release, and post-release phases. Furthermore, twenty tokens, accounting for around 7% of total realizations, exhibited substantial devoicing. This is in contrast with singleton /v/, for which this devoicing was almost never observed, as also previously noted by Pape & Jesus (2015) for Italian. Again, variation in /v:/ did not pattern with region, suggesting that the observed differences are not variety-specific. Lexical factors, such as word frequency or phonological structure, are also unlikely to account for the variation, as the two target words differ on both dimensions. *Davvero* is substantially more frequent than *bevve*, whose frequency is also regionally conditioned;[5] the words also differ in onset composition, with /d/ and /b/ respectively, ruling out any potential segmental harmony effects. Instead, the

[5] Juilland & Traversa (2019) report *davvero* to occur 58 times out of 5,014 entries in their corpus, making it 14.5 times more frequent than *bevve*, which occurs 4 times. *Bevve* is a form of the *passato remoto* in Italian, a tense that is not typically used in everyday speech in northern varieties.

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

observed variation likely reflects aerodynamic and articulatory challenges associated with sustaining voicing and frication over a longer segmental duration (Ohala 1983; Shadle 2010). Speaker sex was the only tested factor to influence /vː/ realization, with male speakers more likely to produce constricted or plosive-like variants than female speakers, possibly pointing to sociophonetic differences which need however to be tested further.

Importantly, these results call into question the principle of geminate inalterability (e.g., Hayes 1986; Kirchner 2000). In its theoretical sense, the term refers to the resistance of geminate consonants to lenition processes – a pattern that is indeed borne out in the present data. At the same time, however, *inalterability* also carries the implication that geminates remain invariant in their realization, which is not supported here. While /vː/ does not lenite, its realizations are far from uniform: non-canonical tokens frequently display acoustic evidence of increased constriction, pointing to strengthening rather than weakening. All in all, the variability observed for /vː/ may be interpreted as a by-product of maintaining contrastively long durations, as also discussed for Research Question 3 (RQ3) below, especially under articulatory conditions that make long voiced fricatives difficult to produce. Crucially, this variability cannot be observed when focusing on voiceless obstruents, for which sustaining longer durations presents fewer articulatory challenges, and which have formed the primary basis for theoretical accounts of geminate inalterability such as Kirchner (2000).[6]

With respect to strengthening and weakening, Research Question 2 (RQ2) asked whether the allophones observed in this study could be grouped into strength categories, with approximant variants being classified as weakened and more constricted forms – including plosive realizations – as stronger compared to canonical forms. The acoustic analysis of strength measures (*IntDiff* and *MaxVel*) broadly confirmed this classification, while also revealing patterns specific to individual allophones. Whereas approximant realizations, only observed for singleton /v/, were clearly separate from canonical ones, more constricted variants often patterned together, suggesting that they may be different manifestations of the same phenomenon. In particular, [ʋː], [b̪ː], and [b̥̪ː] displayed similar mean *IntDiff* values, while for *MaxVel*, [ʋː] was slightly higher than [b̪ː], yet statistically equivalent to devoiced [b̥̪ː]. Acoustically, this similarity is unsurprising, as the analysis targeted a frequency band (400–1200 Hz) that excluded frication noise but also most of voicing-related energy (cf. Ennever et al. 2017). Qualitatively, however, the presence of high-frequency noise clearly distinguishes [ʋː] from [b̪ː], which we believe are best understood as part of a strength-conditioned continuum spanning canonical fricative to fully plosive realizations. The articulatory-based reason we give for this is that we expect /vː/ to be inherently characterized by a high degree of constriction, as also shown through real-time magnetic resonance imaging (rt-MRI) data by Fujimoto et al. (2023) for the Ikema dialect of Miyako Ryukyuan (see also Payne 2006 and Burroni et al. 2024 for articulatory evidence of greater constriction in Italian coronal and bilabial geminates). This increased constriction, paired with a reduced intraoral pressure buildup due to voicing (cf. § 1.1), forms a challenging barrier for airflow to pass through, which may result in the observed acoustic realizations.

As for RQ3, focusing on consonant duration, the analysis showed that consonantal strength had no significant effect on this parameter. Despite clear differences in manner of articulation – for example, between fricative and approximant /v/ variants – segmental duration tended to remain stable within each length category. Although this finding runs contrary to our expectation of shorter lenited realizations (cf. § 2), it provides evidence that Italian speakers maintain duration as the primary cue to the singleton-geminate contrast, even when the degree of supraglottal constriction or voicing properties fluctuate. The lack

[6] Notably, Hayes (1986) presents the inalterability of geminate /vː/ in Persian in support of his argument; however, the analysis remains purely phonological, without accompanying experimental data.

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

of durational impact from allophonic variation supports the view that segment length in Italian is phonologically robust cross-regionally and actively maintained across a range of phonetic conditions, which nevertheless can help enhance the length contrast.

6. Conclusion

Taken together, the above findings suggest that intervocalic lenition in Italian operates independently of regional boundaries in the case of singleton /v/, although a larger speaker pool for each regional variety is necessary to substantiate this conclusion. At the same time, the data reveal that geminate /v:/ is not uniformly stable in surface form, and that more constricted forms, e.g., [b̪:], rather than lenited ones, may emerge as a result of the inherent strength, coupled with the aerodynamic demands, of long voiced fricatives. Nevertheless, articulatory investigations of the type conducted by Fujimoto et al. (2023) for Miyako Ryukyuan are required to confirm these hypotheses for Italian.

These findings suggest a need to refine our understanding of geminate inalterability, particularly in the context of marked segments such as /v:/, and call for further research on the interplay between segmental duration, constriction degree, and voicing in the phonetic realization of length contrasts. At the same time, our results have also uncovered a new allophone, i.e. [b̪:], not previously reported for any variety of Italian (nor indeed for any other language we are currently aware of). We think this omission is most likely due to the typological rarity of /v:/, and a lack of acoustic investigation more generally of voiced labiodental fricatives.

Regarding possible limitations, the use of read speech may have reduced the degree of naturalistic variation observable in spontaneous discourse. Furthermore, although the sample size (18 speakers in total, with six per regional variety) is consistent with experimental standards, a larger and more demographically diverse pool would allow for more robust generalizations, particularly in relation to sociophonetic factors such as regional variety, speaker sex or age.

Future research should address these limitations by investigating the same phenomena in spontaneous or semi-spontaneous speech, testing a broader range of regional varieties, and exploring the contribution of additional factors such as speech rate and style, or vowel context. Moreover, the observed sex differences in /v:/ production merit further sociophonetic inquiry to determine whether they can be generalized to a broader population.

Acknowledgements We are grateful to the participants in this study and to the two anonymous reviewers for their insightful feedback.

Appendix A

Table A1. Pairwise comparisons of mean *IntDiff* values between levels for Model M1a. Statistically credible differences are indicated by 95% credible intervals that do not include zero.

Contrast	Mean (left)	Mean (right)	Diff.	95% CrI Lower	95% CrI Upper	Evidence
ʋ – v	16.57	25.63	−9.03	−10.66	−7.48	Credible
ʋ – v:	16.57	36.74	−20.15	−21.69	−18.57	Credible
ʋ – b̪:	16.57	42.60	−26.03	−28.38	−23.87	Credible
ʋ – v̥:	16.57	37.78	−21.21	−24.93	−17.54	Credible
ʋ – b̥̪:	16.57	42.30	−25.71	−31.33	−20.64	Credible
ʋ – v̞:	16.57	43.95	−27.37	−29.50	−25.21	Credible
v – v:	25.63	36.74	−11.11	−12.56	−9.68	Credible

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

Table A1. Continued.

Contrast	Mean (left)	Mean (right)	Diff.	95% CrI Lower	95% CrI Upper	Evidence
v – b̪:	25.63	42.60	−16.99	−19.32	−14.67	Credible
v – y̥:	25.63	37.78	−12.17	−15.85	−8.55	Credible
v – b̥:	25.63	42.30	−16.69	−22.18	−11.57	Credible
v – y̬:	25.63	43.95	−18.33	−20.48	−16.23	Credible
vː – b̪:	36.74	42.60	−5.88	−8.28	−3.59	Credible
vː – y̥:	36.74	37.78	−1.04	−4.85	2.46	Unclear
vː – b̥:	36.74	42.30	−5.56	−11.02	−0.33	Credible
vː – y̬:	36.74	43.95	−7.21	−9.39	−5.02	Credible
b̪: – y̥:	42.60	37.78	4.83	0.91	8.92	Credible
b̪: – b̥:	42.60	42.30	0.32	−5.24	5.93	Unclear
b̪: – y̬:	42.60	43.95	−1.35	−3.94	1.44	Unclear
y̥: – b̥:	37.78	42.30	−4.53	−10.85	1.59	Unclear
y̥: – y̬:	37.78	43.95	−6.16	−10.07	−2.04	Credible
b̥: – y̬:	42.30	43.95	−1.66	−7.33	3.69	Unclear

Table A2. Pairwise comparisons of mean *MaxVel* values between levels for Model M1b. Statistically credible differences are indicated by 95% credible intervals that do not include zero.

Contrast	Mean (left)	Mean (right)	Diff.	95% CrI Lower	95% CrI Upper	Evidence
ʋ – v	1.42	1.87	−0.45	−0.59	−0.32	Credible
ʋ – vː	1.42	2.17	−0.75	−0.88	−0.63	Credible
ʋ – b̪:	1.42	2.26	−0.84	−1.03	−0.66	Credible
ʋ – y̥:	1.42	2.26	−0.85	−1.14	−0.54	Credible
ʋ – b̥:	1.42	2.74	−1.32	−1.72	−0.90	Credible
ʋ – y̬:	1.42	2.56	−1.14	−1.31	−0.96	Credible
v – vː	1.87	2.17	−0.30	−0.42	−0.18	Credible
v – b̪:	1.87	2.26	−0.39	−0.59	−0.20	Credible
v – y̥:	1.87	2.26	−0.39	−0.69	−0.09	Credible
v – b̥:	1.87	2.74	−0.87	−1.29	−0.46	Credible
v – y̬:	1.87	2.56	−0.69	−0.85	−0.50	Credible
vː – b̪:	2.17	2.26	−0.09	−0.29	0.10	Unclear
vː – y̥:	2.17	2.26	−0.09	−0.38	0.22	Unclear
vː – b̥:	2.17	2.74	−0.56	−0.97	−0.12	Credible
vː – y̬:	2.17	2.56	−0.39	−0.57	−0.21	Credible
b̪: – y̥:	2.26	2.26	0.00	−0.32	0.33	Unclear
b̪: – b̥:	2.26	2.74	−0.48	−0.92	−0.05	Credible
b̪: – y̬:	2.26	2.56	−0.30	−0.52	−0.08	Credible
y̥: – b̥:	2.26	2.74	−0.48	−0.96	0.03	Unclear
y̥: – y̬:	2.26	2.56	−0.30	−0.63	0.02	Unclear
b̥: – y̬:	2.74	2.56	0.18	−0.26	0.60	Unclear

https://doi.org/10.1017/S0025100325100881 Published online by Cambridge University Press

References

Abramson, Arthur S. & D.H. Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002.

Allan, Robin. 2014. *Danish: An essential grammar*. London: Routledge.

Árnason, Kristján. 2011. *The phonology of Icelandic and Faroese*. Oxford & New York: Oxford University Press.

Arvaniti, Amalia. 1999. Illustrations of the IPA: Cypriot Greek. *Journal of the International Phonetic Association* 29, 173–178.

Bárkányi, Zsuzsanna & Zoltán Kiss. 2010. A phonetic approach to the phonology of v: A case study from Hungarian and Slovak. In Susanne Fuchs, Martine Toda & Marzena Zygis (eds.), *Turbulent Sounds*, 103–142. Berlin & New York: De Gruyter Mouton. https://doi.org/10.1515/9783110226584.103.

Bauer, Laurie. 2008. Lenition revisited. *Journal of Linguistics* 44(3), 605–624. https://doi.org/10.1017/S0022226708005331.

Bertinetto, Pier Marco & Michele Loporcaro. 2005. The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association* 35(2), 131–151. https://doi.org/10.1017/S0025100305002148.

Bjorndahl, Christina. 2022. Voicing and frication at the phonetics-phonology interface: An acoustic study of Greek, Serbian, Russian, and English. *Journal of Phonetics* 92, 101–136. https://doi.org/10.1016/j.wocn.2022.101136.

Burroni, Francesco, Sireemas Maspong, Nicole Benker, Philip A. Hoole & James Kirby. 2024. Spatiotemporal features of bilabial geminate and singleton consonants in Italian. In Cécile Fougeron & Pascal Perrier (eds.), *Proceedings of the 13th International Seminar on Speech Production*, Autrans, France, vol. 13, 181–184.

Bürkner, Paul-Christian. 2017. *brms*: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). https://doi.org/10.18637/jss.v080.i01. Published online by the Foundation for Open Access Statistics, 29 August 2017.

Bybee, Joan & Shelece Easterday. 2019. Consonant strengthening: A crosslinguistic survey and articulatory proposal. *Linguistic Typology* 23(2), 263–302. https://doi.org/10.1515/lingty-2019-0015

Canepari, Luciano. 1992. *Manuale di Pronuncia Italiana*. Bologna: Zanichelli.

Catford, John Cunnison. 1977. *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Celata, Chiara, Alessandro Vietti & Lorenzo Spreafico. 2019. An articulatory account of rhotic variation in Tuscan Italian: Synchronized UTI and EPG data. In Mark Gibson & Juana Gil (eds.), *Romance Phonetics and Phonology*, 91–117. Oxford: Oxford University Press.

Cho, Taehong. 2016. Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass* 10(3), 120–141. https://doi.org/10.1111/lnc3.12178.

Crocco, Claudia. 2017. Everyone has an accent. Standard Italian and regional pronunciation. In Massimo Cerruti, Claudia Crocco & Stefania Marzo (eds.), *Towards a new standard: Theoretical and empirical studies on the restandardization of Italian*, 89–117. Berlin & Boston: De Gruyter Mouton.

Dallaston, Katherine & J West. 2018. MacReaper (version 2). https://kjdallaston.com/projects/ (accessed 20 May 2025).

Dian, Angelo, John Hajek & Janet Fletcher. 2024. Cross-regional patterns of obstruent voicing and gemination: The case of Roman and Veneto Italian. *Languages* 9(12), 383. https://doi.org/10.3390/languages9120383. Published online by MDPI, 20 December 2024.

Dilley, Laura, Jessica Gamache, Yuanyuan Wang, Derek M. Houston & Tonya R. Bergeson. 2019. Statistical distributions of consonant variants in infant-directed speech: Evidence that /t/ may be exceptional. *Journal of Phonetics* 75, 73–87. https://doi.org/10.1016/j.wocn.2019.05.004.

Dmitrieva, Olga. 2017. Production of geminate consonants in Russian: Implications for typology. In Haruo Kubozono (ed.), *The phonetics and phonology of geminate consonants*, 1st edn, 34–65. Oxford: Oxford University Press.

Ennever, Thomas, Felicity Meakins & Erich Round. 2017. A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology* 8(1), 20. https://doi.org/10.5334/labphon.18

Escure, Geneviève. 1977. Hierarchies and phonological weakening. *Lingua* 43, 55–64.

Ewen, Colin & Harry van der Hulst. 2001. *The phonological structure of words: An introduction*. Cambridge: Cambridge University Press.

Ferrero, Franco, Arturo Genre, Louis-Jean Boë & Michel Contini. 1979. *Nozioni di fonetica acustica*. Turin: Omega Edizioni.

Flego, Stefon & Jon Forrest. 2021. Leveraging the temporal dynamics of anticipatory vowel-to-vowel coarticulation in linguistic prediction: A statistical modeling approach. *Journal of Phonetics* 88. 101093. https://doi.org/10.1016/j.wocn.2021.101093.

Fletcher, Janet. 2010. The prosody of speech: Timing and rhythm. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The handbook of phonetic sciences*, 1st edn, 521–602. Hoboken, NJ: Wiley. https://doi.org/10.1002/9781444317251.

Foley, James. 1977. *Foundations of theoretical phonology*. Cambridge & New York: Cambridge University Press.

Fujimoto, Masako, Shigeko Shinohara, and Daichi Mochihashi. 2023. Articulation of geminate obstruents in the Ikema Dialect of Miyako Ryukyuan: A real-time MRI analysis. *Journal of the International Phonetic Association* 53(1), 69–93. doi: 10.1017/S0025100321000013.

Giannelli, Luciano & Leonardo Maria Savoia. 1979. L'indebolimento consonantico in Toscana. *Rivista italiana di dialettologia* 2, 25–58.

Hadfield, Jarrod D. 2010. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software* 33(2). https://doi.org/10.18637/jss.v033.i02. Published online by the Foundation for Open Access Statistics, 2 February 2010.

Hansen, Benjamin B. & Scott Myers. 2017. The consonant length contrast in Persian: Production and perception. *Journal of the International Phonetic Association* 47(2), 183–205. https://doi.org/10.1017/S0025100316000244.

Hanulíková, Adriana & Silke Hamann. 2010. Slovak. *Journal of the International Phonetic Association* 40(3), 373–378. https://doi.org/10.1017/S0025100310000162.

Hayes, Bruce. 1986. Inalterability in CV Phonology. *Language* 62(2), 321. https://doi.org/10.2307/414676.

Honeybone, Patrick. 2008. Lenition, weaking and consonantal strength: tracing concepts through the history of phonology. In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and Fortition*, vol. 99, 9–92. Berlin & New York: Mouton de Gruyter. https://doi.org/10.1515/9783110211443.1.9.

Hualde, José Ignacio & Marianna Nadeu. 2012. Lenition and phonemic overlap in Rome Italian. *Phonetica* 68(4), 215–242. https://doi.org/10.1159/000334303.

Jesus, Luis M.T. & Christine H. Shadle. 2002. A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics* 30(3), 437–464. https://doi.org/10.1006/jpho.2002.0169.

Juilland, Alphonse G. & Vicenzo Traversa. 2019. *Frequency dictionary of Italian words (The Romance languages and their structures*. First series). Reprint 2018. Berlin & Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110868937.

Katz, Jonah & Gianmarco Pitzanti. 2019. The phonetics and phonology of lenition: A Campidanese Sardinian case study. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1), 16. https://doi.org/10.5334/labphon.184. Published online by Open Library of Humanities, 26 September 2019.

Kim, Jungyeon. 2022. Multinomial regression modeling of vowel insertion patterns: adaptation of coda stops from English to Korean. *Linguistics* 60(6), 1989–2017. https://doi.org/10.1515/ling-2021-0031.

Kingston, John. 2008. Lenition. In Laura Colantoni & Jeffrey Steele, (eds.), *Proceedings of the Third Conference on Laboratory Approaches to Spanish Phonology*, 1–31. Somerville, MA: Cascadilla Press.

Kirchner, Robert. 1998. *An effort-based approach to consonant lenition*. Ph.D. dissertation, University of California, Los Angeles.

Kirchner, Robert. 2000. Geminate inalterability and lenition. *Language* 76(3), 509–545. https://doi.org/10.2307/417134.

Kirchner, Robert. 2004. Consonant lenition. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 313–345. Cambridge: Cambridge University Press.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). https://doi.org/10.18637/jss.v082.i13. Published online by the Foundation for Open Access Statistics, 6 December 2017.

Ladefoged, Peter & Keith Johnson. 2014. *A course in phonetics*. 7th edn. Stamford, CT: Cengage Learning.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford, UK & Cambridge, Mass.: Blackwell Publishers.

Lass, Roger. 1984. *Phonology: an introduction to basic concepts*. Cambridge: Cambridge University Press.

Lenth, Russel V., Balazs Banfai, Ben Bolker, Paul Buerkner, Iago Giné-Vázquez, Maxime Herve, Maarten Jung, Jonathon Love, Fernando Miguez, Julia Piaskowski, Hannes Riebl & Henrik Singmann. 2025. Package 'emmeans' (version 1.11.1). https://cran.r-project.org/web/packages/emmeans/emmeans.pdf (accessed 20 May 2025).

Maddieson, Ian. 1984. *Patterns of sounds*, doi: 10.1017/CBO9780511753459. Published online by Cambridge University Press, August 2010.

Maddieson, Ian. 2008. Glides and gemination. *Lingua* 118(12), 1926–1936. https://doi.org/10.1016/j.lingua.2007.10.005.

Marotta, Giovanna. 2005. Il consonantismo romano. Processi fonologici e aspetti acustici. In F. Albano Leoni & R. Giordano (eds.), *Italiano parlato. Analisi di un dialogo*, 1–24. Naples: Liguori.

Marotta, Giovanna. 2008. Lenition in Tuscan Italian (Gorgia Toscana). In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition*, vol. 99, 235–272. Berlin & New York: Mouton de Gruyter.

Moran, Steven & Daniel McCloy. 2019. PHOIBLE (version 2.0). https://doi.org/10.5281/ZENODO.2593234. Published online by Zenodo, 14 March 2019. (Available online at http://phoible.org, accessed 20 May 2025).
Nalborczyk, Ladislas, Cédric Batailler, Hélène Loevenbruck, Anne Vilain & Paul-Christian Bürkner. 2019. An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research* 62(5), 1225–1242. https://doi.org/10.1044/2018_JSLHR-S-18-0006.
Nocchi, Nadia & Stephan Schmid. 2008. Aspetti della lenizione in alcune varietà dell'italiano meridionale. In Massimo Pettorino, Antonella Giannini, Marianna Vallone, Renata Savy (eds.), *La comunicazione parlata. Atti del congresso internazionale*. Naples: Liguori, 109–36. https://doi.org/10.5167/UZH-16407.
Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The production of speech*, 189–216. New York: Springer New York. https://doi.org/10.1007/978-1-4613-8202-7_9.
Olson, Kenneth S. & John Hajek. 2004. A crosslinguistic lexicon of the labial flap. *Linguistic Discovery* 2(2), 21–57. http://linguistic-discovery.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/2/xmlpage/1/article/262. Published online by the Dartmouth College Library, 2004.
Pape, Daniel & Luis M.T. Jesus. 2015. Stop and fricative devoicing in European Portuguese, Italian and German. *Language and Speech* 58(2), 224–246. https://doi.org/10.1177/0023830914530604.
Payne, Elinor. 2006. Non-durational indices in Italian geminate consonants. *Journal of the International Phonetic Association* 36(1), 83–95.
Pellegrini, Giovan Battista. 1977. *Carta dei dialetti d'Italia*. Pisa: Pacini.
Petrone, Caterina, Mariapaola D'Imperio, Leonardo Lancia & Susanne Fuchs. 2014. The interplay between prosodic phrasing and accentual prominence on articulatory lengthening in Italian. In Nick Campbell, Dafydd Gibbon & Daniel Hirst (eds.), *Proceedings of the 7th International Conference on Speech Prosody*, Trinity College Dublin, 192–196.
R Core Team. 2025. *R: A language and environment for statistical computing* (version 4.5.0). Vienna: R Foundation for Statistical Computing.
Rogers, Derek & Luciana D'Arcangeli. 2004. Italian. *Journal of the International Phonetic Association* 34(1), 117–121. https://doi.org/10.1017/S0025100304001628.
Russo, Michela. 2022. Locality domains on lenition. Spirantization (gorgia) and voicing in Tuscan dialects. *Linx* (84). https://doi.org/10.4000/linx.9107. Published online by OpenEdition, 30 August 2022.
Saborit Vilar, Josep. 2009. *Millorem la pronúncia*. Valencia: Acadèmia Valenciana de la Llengua.
Shadle, Christine H. 2010. The aerodynamics of speech. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The handbook of phonetic sciences*, 1st edn, 39–80. Hoboken, NJ: Wiley. https://doi.org/10.1002/9781444317251.ch2.
Shinohara, Shigeko & Masako Fujimoto. 2018. Acoustic characteristics of the obstruent and nasal geminates in the Ikema dialect of Miyako Ryukyuan. In Elena Babatsouli (ed.), *Crosslinguistic research in monolingual and bilingual speech*, 253–270. Chania, Greece: Institute of Monolingual and Bilingual Speech.
Sidorkevich, Daria. 2014. *Язык ингерманландских переселенцев в Сибири: структура, диалектные особенности, контактные явления* [The language of the Ingrian settlers in Siberia: structure, dialectal features, contact phenomena]. Ph.D. dissertation, the Russian Academy of Sciences, Saint Petersburg.
Siptár, Peter & Miklos Törkenczy. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.
Stevens, Kenneth N. 2000. *Acoustic phonetics*. 1st paperback edn. Cambridge, Mass.: MIT Press.
Stevens, Kenneth N., Sheila E. Blumstein, Laura Glicksman, Martha Burton & Kathleen Kurowski. 1992. Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *The Journal of the Acoustical Society of America* 91(5), 2979–3000. https://doi.org/10.1121/1.402933.
Talkin, David. 2015. REAPER: Robust Epoch And Pitch EstimatoR. https://github.com/google/REAPER (accessed 20 May 2025).
Trumper, John B. & M. Maddalon. 1982. *L'italiano regionale tra lingua e dialetto. Presupposti ed analisi*. Cosenza: Brenner.
Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35(4), 445–472. https://doi.org/10.1016/j.wocn.2006.12.001.
Vasishth, Shravan, Bruno Nicenboim, Mary E. Beckman, Fangfang Li & Eun Jong Kong. 2018. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics* 71, 147–161. https://doi.org/10.1016/j.wocn.2018.07.008.
Vietti, Alessandro. 2019. Phonological variation and change in Italian. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.494. Published online by Oxford University Press, 25 February 2019.
Westbury, John R. & Patricia A. Keating. 1986. On the naturalness of stop consonant voicing. *Journal of linguistics* 22(1), 145–166.
Wilbur, Joshua. 2014. *A grammar of Pite Saami*. Berlin: Language Science Press.

Winkelmann, Raphael, Jonathan Harrington & Klaus Jänsch. 2017. EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language* 45, 392–410.
Winkelmann, Raphael, Lasse Bombien, Michel Scheffers & Marcus Jochim. 2024. wrassp: Interface to the 'ASSP' Library. doi:10.32614/CRAN.package.wrassp, R package version 1.0.5, https://CRAN.R-project.org/package=wrassp.

Cite this article: Dian Angelo, Hajek John, and Fletcher Janet (2026). The acoustic realization of intervocalic /v/ and /v:/ in Italian: evidence of articulatory variability. *Journal of the International Phonetic Association* **56**(e7), 1–25. https://doi.org/10.1017/S0025100325100881