Article

Language and Speech

Language *and* Speech
2022, Vol. 65(3) 755–780

DOI: 10.1177/00238309211065789
journals.sagepub.com/home/las
SAGE

Acoustic and Kinematic Correlates of Heterosyllabicity in Different Phonological Contexts

Chiara Celata
Università degli Studi di Urbino 'Carlo Bo', Italy

Chiara Meluzzi
Università degli Studi di Milano, Italy

Chiara Bertini
Scuola Normale Superiore, Pisa, Italy

Abstract
We investigate the temporal and kinematic properties of consonant gemination and heterosyllabic clusters as opposed to singletons and tautosyllabic clusters in Italian. The data show that the singleton versus geminate contrast is conveyed by specific kinematic properties in addition to systematic durational differences in both the consonantal and vocalic intervals; by contrast, tautosyllabic and heterosyllabic clusters differ significantly for the duration of the consonantal interval but do not vary systematically with respect to the vocalic interval and cannot be consistently differentiated at the kinematic level. We conclude that systematic variations in acoustic vowel duration and the kinematics of tongue tip gestures represent the phonetic correlates of the segmental phonological contrast between short and long consonants, rather than of syllable structure. Data are only partly consistent with the predictions of both moraic and gesture-based models of the syllable about the effects of syllable structure on speech production dynamics and call for a more gradient view of syllabification.

Keywords
Gemination, consonant clusters, syllable, gestural coordination, vowel duration, ultrasound tongue imaging (UTI), Italian phonology

Corresponding author:
Chiara Celata, Università degli Studi di Urbino 'Carlo Bo', Via San Girolamo 6/8, 61029 Urbino PU, Italy.
Email: chiara.celata@uniurb.it

1 Introduction

1.1 Heterosyllabicity in Italian

In Italian, paroxytone disyllables containing a geminate consonant show a significantly shorter stressed vowel than disyllables containing a singleton (e.g., D'Imperio & Rosenthall, 1999; Esposito & Di Benedetto, 1999; Fava & Magno Caldognetto, 1976; Mairano & De Iacovo, 2020; McCrary, 2004; Vayra et al., 1999; Vogel, 1982). This timing characteristic has made Italian a good example of a language in which a vowel-to-vowel coordination pattern applies (i.e., where the interval between the targets of two consecutive vocalic gestures does not depend on the size of the intervening consonantal gesture; Fowler, 1983; Smith, 1992). Conversely, in canonical phonological accounts, this coordination is reflected in the generalization that a stressed penultimate syllable in standard Italian is bimoraic (e.g., Gordon, 2016; Krämer, 2009). Besides the singleton versus geminate distinction, the duration of preceding stressed vowels is considered among the factors that are used to parse Italian consonant clusters into tautosyllabic and heterosyllabic categories (e.g., Kenstowicz, 2017; Maddieson, 1984; Vogel, 1982). Based on the view that the syllable with its moraic structure determines the phonetic duration of the segments, it is assumed that stressed vowels before tautosyllabic onset clusters (stop-liquid clusters, also called *muta cum liquida* clusters) are long as they are the nucleus of an open syllable, whereas vowels before heterosyllabic clusters (sonorant-stop and stop-stop clusters) are short as they pertain to closed syllables. However, Italian phonotactics constrains the phonetic shape of clusters in such a way that only a reduced set of consonant combinations produce tautosyllabic clusters, compared with heterosyllabic ones that are more varied. For these reasons, as it will be shown below, directly comparing the singleton/geminate alternation with the tauto-/heterosyllabic cluster alternation is far from being a straightforward operation. Capitalizing on this structural observation, and providing detailed acoustic and articulatory data on the temporal organization of vowel-consonant sequences, this study questions that the same timing patterns are at the origin of the vowel shortenings that can be observed before geminates and before heterosyllabic clusters (compared with, respectively, singletons and tautosyllabic clusters). We explore an alternative view, according to which vowel shortenings, showing different acoustic and articulatory characteristics in the two contexts, should be interpreted as separate phenomena, bearing different functions in the phonology of the language and indexing the phonological length feature in the specific case of gemination.

It is well known from previous studies that besides syllable parsing, other factors also play a role in shaping the timing relationships between consecutive vowels and consonants. For instance, Farnetani and Kori (1986) found that if a cluster contained a voiceless plosive, the preceding vowel was shorter than when the cluster contained a voiced plosive, probably to compensate the increased duration of the consonantal interval. Interestingly, the effect was present irrespective of whether the cluster was phonologically tautosyllabic (stop-liquid, for example, /tr/ vs. /dr/) or heterosyllabic (sonorant-stop, for example, /rt/ vs. /rd/). These findings show that syllables in a word interact with each other, at least at the temporal level: vowel duration may be predicted by the phonological characteristics of segments pertaining to the *following* syllable. Evidence of cross-syllable effects in segments' duration can be found even elsewhere in Italian. For instance, the effect of stop voicing on the duration of vowels followed by nasal-stop clusters spans syllable boundaries (VNCV sequences, with C varying for voicing; Calamai & Celata, 2011). However, in that case, subphonemic vowel variation could also be said to depend on the nasal which also varies, being longer before a /t/ than before a /d/. By contrast, in the /tr/-/dr/ and /rt/-/rd/ examples mentioned above, the temporal effect occurs unquestionably across syllables. Farnetani and Kori (1986) additionally reported that vowels preceding palatals /ʃ ɳ ʎ/ were significantly longer than vowels preceding geminates /sː lː nː/, but shorter than vowels preceding singletons /s l n/. This was explained as a consequence of the fact that palatals, despite behaving phonologically like geminates and thus closing the syllable, are considerably shorter than geminates. Similar patterns of gradual variation were found in laboratory speech (e.g., McCrary, 2004; Vayra et al., 1999) as well as spoken colloquial Italian (Celata & Mairano, 2014).

As geminates have a significantly greater duration compared with singletons, vowel shortening has been viewed as a phenomenon of compensatory variation (or complementarity), although imperfect (e.g., D'Imperio & Rosenthall, 1999; Fowler, 1981; Marotta, 1985). Compensation would imply a cross-syllable account of vowel durational variations: intervocalic geminates, which span a syllable boundary, are longer than intervocalic singletons and *therefore* induce a shortening of the preceding vowel. This view is incompatible with any theory (either moraic or gesture-based) that looks at syllable constituency as the predictor of durational-rhythmical variations. On the one hand, mora theory is not designed to model gradual lengthening or shortening phenomena. On the other hand, gesture-based approaches do account for gradual compression effects, but only within syllables (see below, §1.2) or across syllable in the case of prosodic juncture effects (e.g., Byrd & Saltzman's, 2003, π-gesture hypothesis); cross-syllable morpheme-internal effects would not qualify for gesture-based explanations and should therefore be explained in another way (Browman & Goldstein, 1992). Interpretations of these facts have therefore varied from strong to mild rejection of a strictly moraic account (without, however, bolstering the case of gesture-based explanations). Among the strongest rejections was McCrary's (2004, p. 263) conclusion that "the standard syllable-based analyses of [. . .] segment duration are not supported by experimental evidence. Instead, the conditioning factors [. . .] are segmental, contrast-based conditions. Syllable structure is not implicated in these phenomena." Less extreme positions can also be found. For instance, Farnetani and Kori (1986) assumed that although "the syllable is a base speech unit" in Italian, temporal compensations may also occur cross-syllable inasmuch as "two adjacent syllables are not temporally independent"; the conclusion is that "the unit which tends to be constant in duration is the entire V-to-V temporal interval," that the authors call the "rhythmical syllable" (p. 27) (cf. Fowler, 1983).

All these studies further attest that much of the observed graduality is due to clusters. The validity of this statement is also confirmed on both a diachronic and a typological level. Stop-liquid clusters are tautosyllabic in many languages including Italian, but allow heterosyllabic parsing in some other languages and have allowed an oscillation between the two parsing solutions in diachrony. The most relevant example here is perhaps that of Latin, for which a stage with heterosyllabic *muta cum liquida* clusters has been reconstructed, for at least some words (e.g., Timpanaro, 1965, p. 1090); other examples come from contemporary southern Italian dialects (Loporcaro, 2005). By contrast, intervocalic geminates only allow heterosyllabic parsing.

So the question arises of why different consonantal sequences show different degrees of tolerance with respect to oscillation in syllable parsing. According to Venneman (2012, p. 21), historical descriptions are not themselves explanations and the latter have to be sought "by establishing phonetic correlates of speech sounds in context and generalizing over those correlates." More specifically, in the case of variable vowel shortenings and syllabification in Italian, we hypothesize, based on the literature reviewed below, that articulatory-kinematic correlates, besides acoustic ones, may be relevant to correctly understand variation and should therefore be investigated.

1.2 Heterosyllabicity in articulation

Vowel shortening in closed syllables has been related to a specific articulatory dynamics, according to which the aperture gesture for the vowel is truncated earlier and the subsequent closing gesture is anticipated, compared with when a vowel is in an open syllable (Munhall et al., 1992). However, most of the studies have dealt with either codas of monosyllables (e.g., Avesani et al., 2009; Marin & Pouplier, 2010; Pastätter & Pouplier, 2014; Shaiman, 2001) or, when heterosyllabicity is considered, syllables closed by geminates (an exception is Smith, 1992, showing data on the /mp/ heterosyllabic cluster as produced by one Italian and three Japanese speakers; see below for details). From the studies on geminates, we know that phonological length in consonants is not only conveyed by increased acoustic duration (associated with an increased duration of the jaw opening gesture), but also by specific articulatory and kinematic characteristics, among which are a more extended linguopalatal contact and a longer, spatially larger and sometimes slower gestures. Moreover, the consonantal gesture for an intervocalic geminate is reported to start earlier than the gesture for the corresponding singleton (e.g., Löfqvist, 2005, 2006, 2007, for Japanese and Swedish; Šimko et al., 2014, for Finnish bilabial stops; Türk et al., 2017, for Estonian bilabial stops; Ridouane & Hallé, 2017, for Tashlhiyt Berber).[1] However, there are also studies reporting little or no kinematic difference for increased duration of consonants and vowels. For instance, Zeroual et al.'s (2008) EMA (electro-magnetic articulography) study of /t tː d dː/ in Moroccan Arabic notably reports that the only characteristic that consistently differentiates geminates from singleton is acoustic closure, which is reflected in a longer tongue tip contact; among the voiceless consonants only, the opening gesture is also faster for the geminate than the singleton. By contrast, the position of the articulatory target, the velocity of the closing gesture, the acoustic duration of the preceding vowel, and the gestural anticipation of the consonant during the preceding vowel do not change across length categories. In a similar way, Löfqvist (2017) reports no statistically significant effect of consonant length in Japanese and Italian as far as the relative timing of the lip gesture for /m/ and the tongue body gesture for the following vowel (/a/ or /i/) is concerned.[2]

The articulatory and kinematic studies on Italian geminates are not many. As for the articulatory correlates of increased duration, they have shown extended linguopalatal contact (Payne, 2006) and higher tongue dorsum and predorsum (Dipino & Celata, 2018). As for gestural timing, Smith (1992, 1995) showed that in non-word production of intervocalic /mː/ and /pː/, the interval between the articulatory target of the preceding vowel and that of the consonant was shorter, and the consonantal gesture started earlier and ended later with respect to the vowel-to-vowel transition gesture, compared with the production of intervocalic /m/ and /p/. According to the authors, these changes in both closing and release gestures would help accommodating longer constriction duration with an ideally equivalent vowel-to-vowel temporal interval (Öhman, 1967). The same effects were found in the production of the intervocalic cluster /mp/, which was not different, in articulatory timing, from geminates /mː/ and /pː/.[3] Zmarich et al. (2007) measured the vertical displacement of the lower lip in the production of /pː bː mː fː vː/ included in pseudo words with preceding and following /a/, and found that for only one of the two subjects the velocity of the closing gesture (i.e., at consonantal onset) was higher than for the corresponding singletons, although no difference was found in the amplitude of the gesture; more differences in both velocity and amplitude were found at the release of the consonant. In a similar study, Gili Fivela et al. (2007) analyzed the movements of lower lip, tongue tip, and tongue dorsum during the production of the disyllables /imːa/, /alːa/, and /adːa/. They found that, in the case of the geminate bilabial, the lip gesture started earlier with respect to the preceding /i/ in comparison with when the consonant was a singleton, but the release of the gesture toward the following /a/ was not different in singletons and geminates, which apparently contradicts the data in Zmarich et al. (2007). In the case of coronal laterals and stops, no differences between singletons and geminates were found in the closing gesture (onset of the geminate), whereas the gestures differed significantly at the consonantal release.

Taken together, these studies suggest that kinematic correlates of phonological length can be measured for geminates in various languages, although the effects are evanescent and potentially variable on a cross-linguistic base. For Italian, the available evidence is quantitatively limited, and sometimes contradictory. Coupled with studies of intra-syllabic gestural coordination (since Munhall et al., 1992), they suggest that the gesture for a coda consonant, be it the beginning of a geminate or the coda in a monosyllabic word, starts earlier with respect to the preceding vowel than the gesture for a non-coda consonant. The constriction gesture for a geminate is sometimes found to be slower, other times faster than for a singleton. Such variability in the literature is probably due to language-specific strategies and/or methodological choices (consonant class, phonetic context, etc.).

2 Aims of the current study

This study aims at investigating the motivations of variable vowel shortening and variable kinematics in Italian heterosyllabic sequences. We ask two major questions. The first is whether there are specific kinematic properties that, in addition to the temporal ones, convey the singleton versus geminate contrast in Italian. Based on the evidence reviewed above, we hypothesize that intervocalic geminates are produced with an anticipation of the beginning of the consonantal gesture, compared with intervocalic singletons. We investigate this issue with respect to four anterior coronal consonants, two stops /t t: d d:/ and two sonorants /n n: l l:/, considering that most of the existing evidence on Italian is rather limited to *(bi)labial* consonants and geminates. The second question is whether the temporal and kinematic properties of the singleton versus geminate contrast can also be found in the contrast between an intervocalic tautosyllabic onset cluster and a heterosyllabic one (both beginning with the same consonant). According to a canonical moraic view of syllabic organization, in both cases the contrast is between an open and a closed syllable. If vowel shortening and kinematic variation are explained by the fact that there is a coda in the closed syllable, then the two types of contrasts should not diverge for their temporal and kinematic properties.

To answer these questions, we compared disyllables containing medial singletons (CVCV) and geminates (CVC:V), and disyllables containing medial stop-liquid onset clusters (CVCRV) and medial heterosyllabic clusters (either liquid-stop, CVRCV, or stop-stop, CVCCV).[4] As already anticipated, the phonotactics of Italian does not allow alternative syllable parsing in clusters: a stop-liquid cluster can only be tautosyllabic, and a sonorant-stop or a stop-stop cluster can only be heterosyllabic. Heterosyllabic clusters are phonologically more varied than tautosyllabic ones, whose C1 cannot be but a stop. By contrast, all stop and sonorant consonants (and fricatives as well) may occur as both singletons and geminates in the language. This aspect is of crucial importance because it shows that there are structural asymmetries in the language between the singleton versus geminate alternation, on the one hand, and tautosyllabic versus heterosyllabic cluster alternation, on the other, the latter being phonotactically more constrained and phonetically less varied; moreover, within clusters, an additional asymmetry holds between heterosyllabic clusters (whose C1 may be a stop or a sonorant) and tautosyllabic clusters (whose C1 can only be a stop and C2 can only be a liquid). Speech timing (or more generally, prosody) and phonotactics strongly interact. The distributional differences between the two types of tauto-/heterosyllabic sequences are part of the Italian phonological grammar and should be taken into account when we investigate about heterosyllabicity and its temporal consequences on vowel-consonant organization.

Based on this observation, and on all previous knowledge of the phonetics and phonology of geminates and clusters reviewed above, this study investigates whether the two types of alternation also have different acoustic and articulatory characteristics. More specifically, the hypothesis is that acoustic vowel shortenings and gestural anticipations of the post-vocalic consonant will be found to be more regular in the case of the singleton versus geminate alternation, where they act as secondary phonetic cues of the phonological feature of consonant length, than in the case of clusters, where heterosyllabicity alone is not a sufficient predictor of systematic phonetic variation.

Before describing the experiment, we conclude this section by shortly reviewing three additional issues that we took into account to determine the number of syllables, the amount of segmental variation, and the lexical frequency of the experimental stimuli.

We know from the literature that stressed antepenult and stressed final syllables behave differently from stressed penults in Italian (D'Imperio & Rosenthall, 1999; Hayes, 1995). For this reason, we limited our study to paroxytone disyllables. The implications that will be derived from the experimental results will consequently be relevantly limited to this kind of phonological structure.

Individual segments and sound sequences are known to be more or less variable in duration on a language-specific basis. For instance, languages without phonemic vowel length, like Italian, have been shown to display greater durational differences between vowels of different qualities, compared with languages with phonemic vowel length (e.g., Gordon, 2006, p. 182). Context-sensitivity is an important issue also in articulatory-kinematic studies. For instance, Šimko et al. (2014) reported that the anticipation of the lip closing gesture for /p:/ in Finnish occurred in the /ap:i/ context more than in the /ip:a/ context and when the preceding consonant was /t/ more than when it was /p/; similar effects were found by Türk et al. (2017) in Estonian. Features of the consonant are relevant also in Zeroual et al.'s (2008) and Ridouane and Hallé's (2017) studies, according to which linguopalatal contact and gestural velocity differences are significant for voiceless plosives but not for their voiced cognates. Based on this literature, in this study we introduced some degree of contextual variation. More specifically, the stimuli included two different preceding vowels, that is, low central /a/ or low-mid back /ɔ/. High or mid-high vowels were not used because their tongue configuration would have made the tracking of the tongue tip movement for the apical consonant too uncertain. Moreover, the stimuli included three different phonetic conditions in the context that followed the target consonants: there could be either a low central /a/ vowel or a sound articulated with a back constriction (a vowel such as /o/ or a consonant such as /k/ in a /dk/ cluster), or even a sound articulated with the anterior part of the tongue (a vowel such as /i/ or /e/ or a consonant such as /r/ in a /tr/ cluster).

We opted for real words, instead of purpose-made zero-frequency stimuli, under the hypothesis that a more ecologically valid elicitation paradigm will prevent the speakers from establishing articulatory routines that might differ from those of real speech production (e.g., McMillan et al., 2009). This, however, had a negative impact on the number of available test items, as our phonetic control parameters had to be very strict and not very many Italian words could finally meet those requirements. We also know that words differing for their frequency of occurrence may differ in how individual articulatory movements are produced; for instance, alveolar gestures may be reduced or even deleted in high-frequency words (e.g., Bybee, 2000; Lin et al., 2014), while studies on vowels have reported more anticipatory coarticulation in high-frequency words and more extreme articulatory trajectories in low-frequency words (e.g., Tomaschek et al., 2018). For this reason, we calculated the frequency of our word stimuli with respect to a reference corpus and entered this factor into the analysis. However, as it will be explained below, no effect of frequency was found on any of the experimental variables.

3 Method

3.1 Speech materials

Twenty-eight paroxytone disyllables with target consonants /t/, /d/, /l/, or /n/ were used for the experiment. These were all real Italian words, including toponyms (e.g., *Adria*), names of renowned

Table 1. Materials of the Experiment: Phonological Transcriptions Followed by Orthographic Transcriptions and English Glosses.

	Target C	Singleton	Geminate	Tautosyllabic cluster	Heterosyllabic cluster
Stops	/t/	/ˈbata/ *Bata* (1,010; 0.17)	/ˈbatːa/ *batta* "tap (subjunctive)" (3,462; 0.59)	/ˈpatron/ *Pàtron* (442; 0.08)	/ˈbatman/ *Batman* (20,793; 3.55)
		/ˈmɔta/ *mota* "mud" (762; 0.13)	/ˈmɔtːo/ *motto* "motto" (44,669; 7.62)	/ˈbɔtro/ *botro* "pond" (434; 0.07)	/ˈbɔtman/ *Botman* (27; 0.005)
	/d/	/ˈada/ *Ada* (21,006; 3.58)	/ˈadːa/ *Adda* (24,816; 4.23)	/ˈadria/ *Adria* (14,486; 2.47)	/ˈadmin/ *admin* "admin" (25,599; 4.37)
		/ˈbɔdi/ *body* "small suit for babies" (32,294; 5.51)	/ˈbɔdːa/ *bodda* "fatso (feminine)" (98; 0.02)	/ˈbɔdrum/ *Bodrum* (1,337; 0.23)	/ˈpɔdkast/ *podcast* "podcast" (13,141; 2.24)
sonorants	/l/	/ˈpala/ *pala* "shovel" (36,314; 6.19)	/ˈpalːa/ *palla* "ball" (159,925; 27.27)	impossible	/ˈbaltsa/ *balza* "flounce" (9,222; 1.57)
		/ˈkalo/ *calo* "drop" (146,356; 24.96)	/ˈkalːo/ *callo* "foot corn" (2,590; 0.44)	impossible	/ˈkaltsa/ *calza* "sock" (14,712; 2.51)
	/n/	/ˈvano/ *vano* "room" (65,047; 11.09)	/ˈvanːo/ *vanno* "they go" (635,328; 88.01)	impossible	/ˈvanto/ *vanto* "pride" (18,197; 3.1)
		/ˈgana/ *Ghana* (565; 0.1)	/ˈkanːe/ *canne* "reeds" (37,180; 6.34)	impossible	/ˈgandze/ *ganze* "clever (fem. pl.)" (202; 0.03)

When gloss is absent, the word is a proper noun. Numbers in brackets: lexical frequencies, expressed as number of occurrences in the corpus and occurrences per million of words (see text for details).

commercial companies (e.g., *Bata*), proper names (e.g., *Batman*), or common abbreviations used in everyday language (e.g., *admin*). These items were organized in series of four (for /t/, /d/) or three items (for /l/, /n/) as shown in Table 1. In each series, the target obstruent occurred intervocalically as a singleton (e.g., /ˈbata/), geminate (e.g., /ˈbatːa/), first consonant of an onset cluster (tautosyllabic cluster, e.g., /ˈpatron/), or coda (heterosyllabic cluster, for example, /ˈbatman/). Stop-initial heterosyllabic clusters are not frequent in the Italian lexicon, particularly if compared with sonorant-initial ones; most often, they are included in words of foreign origin. Nevertheless, they do represent a possible option for Italian speakers, and for this reason they were included in the corpus. As sonorants are not allowed as initial segments of a tautosyllabic cluster in Italian, target /l/ and /n/ only occurred intervocalically as singletons and geminates (e.g., /ˈpala/, /ˈpalːa/) or as initial consonants of a heterosyllabic cluster (e.g., /ˈbaltsa/).[5] The unbalance of the speech corpus therefore reflects directly the asymmetry of the Italian phonotactics: both stops and liquids may occur as first member of heterosyllabic clusters (V-stop-stop-V, V-liquid-stop-V), but only a stop can occur as first member of a tautosyllabic cluster (stop-liquid-V). Gaps in the corpus correspond to gaps in the phonological competence of the speakers.

Word-initial consonants did not change within each series, except for voicing distinctions in some of the series, which was judged irrelevant to the purposes of the current investigation. The vowel of the first syllable was always lexically stressed and could be /a/ (in six series) or /ɔ/ (in 2 series); the vowel did not change within each series. As anticipated in §2, some variation was also introduced in the articulatory characteristics of the segment that followed the medial consonant. We controlled statistically for the effects of the segments that followed the target consonant, as well as of those that preceded, by including two independent factors in the analysis, respectively, for the preceding and the following phonetic context (see below, §3.4). Our kinematic analysis was limited to the timing and velocity of the cluster-initial apical gesture; we leave the question of cluster-internal gestural organization (which could also be interesting for the purposes of the current investigation) to future and more focused analyses.

The frequency of each word form (number of tokens and number of occurrences for millions of words) was calculated from the itTenTen16 web corpus (5.9 billion tokens; Jakubíček et al., 2013) and included in the random part of the two models that were run for the articulatory variable Ant (§3.4). As however, the proportion of the random variance did not change when this factor was included with respect to when it was not, we concluded that the item frequency was not a relevant factor in our dataset and we therefore excluded it from further analyses.

3.2 Speakers, equipment, recording procedure

Ten native Italian speakers, five males and five females, aged 24–35 and reporting no speech or hearing deficit were recorded according to the following procedure. They were speakers of a Tuscan variety of Italian, in which gemination and vowel shortening before geminates are said to show up maximally as far as consonant duration ratios are concerned (Mairano & De Iacovo, 2020). The task was a self-paced word reading of isolated target words and was run in an anechoic chamber. Synchronous audio and ultrasound tongue imaging (UTI) data were collected at 60 Hz via a Mindray ultrasound machine equipped with a microconvex probe (Mindray 65EC10EA 6.5 MHz) and a stabilization headset (Articulate Instruments Ltd [AIL], 2008). The position of the probe was supposed to be orthogonal to the tongue surface so that the ultrasonic wave emitted from the probe was able to catch the entire lingual configuration during the speech movements. The acquisition platform was based on the AAA software, version 2.14 (AIL, 2012). During the recording session, the speaker read the prompt items one by one, as they appeared in the AAA window on the computer screen in front of her or him. Three repetitions of the prompt list were recorded, except for one participant who repeated the prompt list four times. The order of the prompts was randomized across participants. The total duration of each recording session varied between 25 and 30 minutes.

3.3 Phonetic analysis and annotation

Once completed, the recordings were exported from AAA into *.wav format and imported into Praat, version 6.0.37 (Boersma & Weenink, 2018) for manual segmentation and annotation. For each stimulus, the target consonant and the preceding vowel were segmented. The annotated acoustic signal was then reimported into AAA for the semi-automatic tracing of midsagittal tongue profiles (see below).

For each item, four phonetic measures were taken; two were acoustic measures and two were articulatory-kinematic.

In the acoustic domain, the duration of the stressed vowel and that of the following consonantal interval were measured. The onset and offset of the stressed vowel were established as the onset and offset, respectively, of higher frequency components (typically, F2). In particular, vowel onset was identified as the zero crossing of the first positive peak in the digitized waveform, alongside clearly visible formant patterns in the spectrogram. Vowel offset, in turn, was defined as the last well-formed period with a visible F2. Where the flanking consonants were nasals or laterals, intensity curves were examined alongside changes in F2 trajectories. The following consonantal interval was the target consonant (singleton or geminate) in some stimuli and a cluster formed by the target consonant and a following consonant in other stimuli. The onset of the consonantal interval coincided with the offset of the stressed vowel. When the consonantal interval corresponded to or ended in a plosive, its offset was established as the zero crossing of the first positive peak in the digitized waveform of the following unstressed vowel, alongside clearly visible formant patterns in the spectrogram. When the consonantal interval corresponded to or ended in a nasal, lateral, or rhotic consonant, intensity curves were examined alongside changes in F2 trajectories to determine the boundary between the consonantal interval and the following unstressed vowel. Affricates were treated as contour segments and therefore included as such in the consonantal interval.

In the articulatory-kinematic domain, peak velocity of tongue tip gestures (henceforth: MaxVel) and the degree of anticipation of the apical gesture during the production of the preceding vowel (henceforth: Ant) were measured, based on UTI data. The procedure was inspired by Zeroual et al. (2008), who measured the gestural anticipation of alveolar plosives as the distance from the acoustic onset of the preceding /a/ to the onset of the tongue tip closing gesture. We replicated the procedure described in Strycharczuk and Scobbie (2015) for extracting dynamic information from midsagittal ultrasound images of tongue shape and location.

More specifically, UTI data were processed as follows. The fan set-up (i.e., the search area within which the software operates the first gross tongue profile tracking) was customized for each speaker. Tongue splines (i.e., tongue surface contours) in each UTI frame included in the acoustically annotated intervals were traced according to a semi-automatic procedure: first, a gross recognition of the brightest points in each ultrasound frame, potentially corresponding to the 42 points composing the tongue profile, was automatically run by the AAA software; then, careful manual correction was carried out, again frame by frame. The hard palate of each speaker was also tracked and then superimposed to the lingual profiles for reference; the palate images were obtained from the ultrasound frames relative to the moment of swallowing some water. Once tongue splines were traced in the acoustically annotated intervals, we first identified the relevant fan radials for tongue vertical displacement measurement (Figure 1, left). The fan is the coordinate frame, centered at the origin of the ultrasound signal (the probe) and consisting of 42 equidistant radials. Depending on the different articulatory properties of consonants and phonetic contexts, one of the fan lines between the 7th and the 12th from the right was selected to maximally capture the articulatory differences between the vocalic and consonantal gesture in terms of tongue vertical displacement in the specific region of interest (tongue tip / blade). The procedure was entirely manual and qualitative, that is, based on observation of concrete articulatory gestures, on a speaker-specific basis. As a second step, we extracted tongue displacement and absolute velocity values along the selected fan (Figure 1, center). Peak velocity was the maximum velocity positive value of the tongue tip raising toward the palate. In the third step, we determined the instant at which the alveolar gesture for the post-vocalic consonant began (Figure 1, right). The articulatory maximum for the alveolar gesture occurs where the tongue tip velocity reaches its minimum (velocity = 0 cm/sec; the tongue is steady against the palate); this is preceded by a velocity peak that identifies the movement toward the target. We defined the articulatory beginning of the alveolar gesture as the point before the peak where the velocity reached at least 20% of the peak velocity. We then measured the duration of the interval from the onset of the vowel and the beginning of the alveolar gesture ("gestural delay" in Figure 1, right). As a fourth and final step, we divided this interval by the duration of the vowel, thus resulting in the proportional measure of how much of the vowel is produced before the gesture for the post-vocalic consonant begins (Ant). Smaller Ant values thus indicated stronger anticipation of the consonant with respect to the beginning of the vowel.

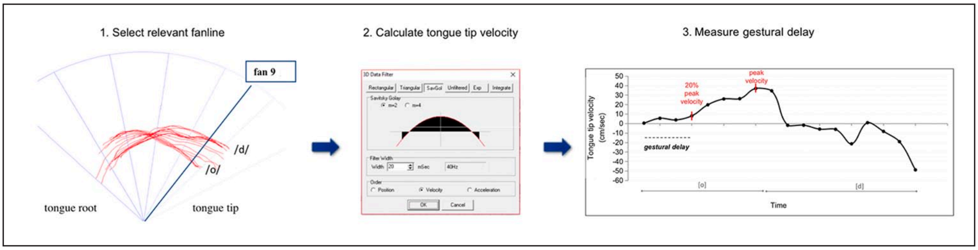

Figure 1. Schematization of the procedure for the extraction of the kinematic-articulatory measures MaxVel and Ant from the stimulus *body*. See text for details.

Table 2. Number (and Percentage) of Outliers in the Datasets.

	C	*V*	*MaxVel*	Ant
Singletons–geminates (*N* = 490)	10 (2%)	9 (1.8%)	15 (3%)	11 (2.2%)
Tautosyllabic–heterosyllabic clusters (*N* = 370)	6 (1.6%)	8 (2.2%)	10 (2.7%)	8 (2.2%)

3.4 Statistics

Linear mixed models were calculated using R (version 3.5.0; R Core Team, 2018) with the lme4 package (Bates et al., 2015) fit by restricted maximum likelihood (REML). We estimated the best fitting model performing likelihood ratio tests between nested models: starting from the most general model and removing the redundant factors via a back-forward procedure. The analysis was split into two parts: in the first part, we analyzed singletons and geminates, and in the second, tautosyllabic and heterosyllabic clusters. In both parts, the four dependent variables mentioned in §3.3 were analyzed, namely, the duration of the stressed vowel (henceforth: V), the duration of the consonantal interval (henceforth: C), MaxVel, and Ant.

Each variable was independently modeled. The factors which were supposed to influence the results were the following: Phonology (singleton vs. geminate, in the first analysis; tautosyllabic vs. heterosyllabic cluster, in the second); Constriction (plosives vs. sonorants); PrecedingV (/a/ vs. /ɔ/); and FollowingGesture (three levels: open, when the phoneme which followed the target consonant was /a/; tongue body constriction, when the following phoneme was /o/ or /k/; and tongue front constriction, when the following phoneme was /i/ or /e/). Speaker, Item, and Repetition were included as random factors. As mentioned above (§3.1), the Frequency of the items was initially included among the random factors in the models with MaxVel and Ant as the dependent variables, but in no case it proved to account for any part of the random variation component; therefore, it was subsequently excluded from the statistics. Observations over and above the standard deviation multiplied by 2.5 were considered outliers and therefore excluded from the statistics: this led to the exclusion of a few items, as summarized in Table 2.

The most general model included all the independent factors as predictors as well as the interactions between Phonology and Constriction, Phonology and PrecedingV, and Phonology and FollowingGesture; the model also included by-Speaker and by-Item random intercepts in its random part as well as by-Speaker random slopes for Phonology and Constriction and by-Item random slopes for Phonology and Repetition, to account for the effects of random variance in these predictors (Baayen et al., 2008). In the case of the cluster dataset, it proved impossible to start from the most general model including all independent factors and interactions as this raised problems of non-convergence, partly related to the unbalance of the experimental conditions across the dataset.

Table 3. Optimized Linear Mixed Model for Variable C.

	Estimate	*SE*	*t* value	Pr(>|*t*|)
(Intercept)	0.1325	0.0059	22.336	<.001***
Phonology	−0.1034	0.0090	−11.485	.0005***
Constriction	0.0204	0.0086	−2.379	.0307*

Data subset: singletons and geminates. Number of observations: 480; speakers: 10; items: 16. Correlation between Phonology and Constriction: .035.
****p* ≤ 0.001; ***p* ≤ 0.01; **p* ≤ 0.05.

Table 4. Optimized Linear Mixed Model for the Variable V.

	Estimate	*SE*	*t* value	Pr(>|*t*|)
(Intercept)	0.1701	0.0074	22.857	<.001***
Phonology	0.0373	0.0066	5.617	.0009***
Constriction	0.0194	0.0065	2.959	.0076**

Data subset: singletons and geminates. Number of observations: 480; speakers: 10; items: 16. Correlation between Phonology and Constriction: .090.
****p* ≤ 0.001; ***p* ≤ 0.01; **p* ≤ 0.05.

Therefore, we ran a general model that was equivalent to that of the singleton/geminate dataset with the exception of the interactions, which were not included. Then, to verify whether the difference between tautosyllabicity and heterosyllabicity was shaped by the other predictors, we ran post hoc comparisons (independent-samples *t*-tests or Wilcoxon paired signed-rank test when the normality assumption was not met) crossing Phonology with Constriction, PrecedingV, or FollowingGesture.

The degree of correlation between C and V was calculated using the Kendall rank correlation test.

4 Results

4.1 Singletons and geminates

4.1.1 Acoustics. The difference in duration between singletons and geminates was extremely robust, as expected, both over the entire dataset (singletons: N=243, μ=0.081 s, σ=23; geminates: N=247, μ=0.185 s, σ=29) and in the speech of each participant. The linear model that best accounted for the data, calculated according to the procedure in §2.4, included only two predictors, that is, Phonology and Constriction. The output of the model is shown in Table 3 (deviation coding). The factor Phonology accounted for a difference of 0.103 seconds between singletons and geminates; in addition, plosives were overall 0.020 seconds longer than sonorants. No significant interactions were found, which meant that the difference between singletons and geminates was equally represented among plosives and sonorants; at the same time, plosives were longer than sonorants both when singletons and when geminates.

The difference between singletons and geminates was also evident in the acoustic duration of the preceding vowel, with geminates showing shorter vowels (N=247, μ=0.151 s, σ=30) than singletons (N=243, μ=0.188 s, σ=29). The model that best accounted for the data included the two predictors Phonology and Constriction (Table 4). The factor Phonology accounted for a difference of 0.037 seconds (recall that the difference was over 0.100 seconds for Variable C); by contrast, the effect of Constriction on vowel duration was approximately of the same magnitude as for

Table 5. Optimized Linear Mixed Model for the Variable MaxVel.

	Estimate	*SE*	*t* value	Pr(>|*t*|)
(Intercept)	0.958	0.0544	22.112	<.001***
Phonology	−0.6218	0.1033	3.617	.0085**
PrecedingV	−0.0985	0.0376	−1.034	.0581 ns

Data subset: singletons and geminates. Number of observations: 475; speakers: 10; items: 16. Correlation between Phonology and PrecedingV: .127.
****p* ≤ 0.001; ***p* ≤ 0.01; **p* ≤ 0.05.

Table 6. Optimized Linear Mixed Model for the Variable Ant.

	Estimate	Std. Error	*t* value	Pr(>|*t*|)
(Intercept)	0.5614	0.0244	22.973	<.001***
Phonology	0.1207	0.0274	4.394	.0004***
Constriction	0.0917	0.0380	−2.414	.0280*
PrecedingV	−0.1344	0.0269	4.992	.0003***

Data subset: singletons and geminates. Number of observations: 479; speakers: 10; items: 16. Maximum correlation between variables: .258.
****p* ≤ 0.001; ***p* ≤ 0.01; **p* ≤ 0.05.

consonant duration, vowels before plosives being on average 0.019 seconds shorter than before sonorant (recall that plosives were estimated 0.020 seconds longer than sonorants in the C model). No significant interaction was found.

Kendall's correlation test showed that V was inversely correlated with C (τ=–0.338, $p<.001$). This indicated that, in the acoustic dimension, the longer the consonant, the shorter the preceding vowel.

4.1.2 Tongue tip kinematics. MaxVel was higher in geminates (N=243, μ=38.8 cm/s, σ=10.5) than in singletons (N=247, μ=32.1 cm/s, σ=8.6), which suggested that the apical gesture was faster in the realization of geminate constriction, compared with singletons. The statistics confirmed that the only significant predictor was Phonology (Table 5), which accounted for a difference of 6.2 cm/s. The velocity of the apical gesture appeared not to be influenced by phonetic factors such as the following articulatory gesture; there was a statistically non-significant tendency for consonants preceded by /ɔ/ to be initiated with a faster movement compared with consonants preceded by /a/.

Average Ant values were found to be higher in singletons (N=247, μ=0.656, σ=0.145) than in geminates (N=243, μ=0.530, σ=0.178), suggesting that the anticipation of the apical gesture occurred proportionally earlier in geminates than in singletons. The model that best accounted for the data included three predictors, that is, Phonology, Constriction, and PrecedingV. Thus, compared with the acoustic variables of C and V and the kinematic variable of MaxVel, Ant appeared to be influenced by the shape assumed by the oral cavity during the production of the preceding vowel. The output of the model is shown in Table 6. In addition to the singleton versus geminate distinction, Ant was longer for sonorants than for plosives, thus suggesting that the apical gesture began comparatively earlier when the consonant was a plosive with respect to a sonorant. Recall that Ant is a proportional estimation uninfluenced by the absolute duration of the vowel. Moreover, the apical gesture began comparatively earlier when the vowel was /ɔ/ than when it was /a/, indicating that, all other things begin equal, the tongue tip started earlier its raising gesture when the

Table 7. Optimized Linear Mixed Model for Variable C.

	Estimate	*SE*	*t* value	Pr(>|*t*|)
(Intercept)	0.2170	0.0059	36.341	<.001***
Phonology	−0.0603	0.0041	−14.597	<.001***
FollowingGesture (tfront)	−0.0183	0.0038	−4.721	.050**
FollowingGesture (tbody)	−0.0078	0.0038	−2.032	>.050 ns

Data subset: clusters. Number of observations: 364; speakers: 10; items: 12. Correlation between Phonology and FollowingGesture: –0.26.
****p* ≤ 0.001; ***p* ≤ 0.01; **p* ≤ 0.05.

Table 8. Optimized Linear Mixed Model for Variable V.

	Estimate	*SE*	*t* value	Pr(>|*t*|)
(Intercept)	0.1651	0.0069	23.694	<.001***
Phonology	0.0333	0.0048	6.918	<.001***
Constriction	−0.0334	0.0048	−6.933	<.001***
FollowingGesture (tfront)	0.0147	0.0034	4.250	<.050**
FollowingGesture (tbody)	0.0065	0.0034	1.903	>.050

Data subset: clusters. Number of observations: 362; speakers: 10; items: 12. Maximum correlation between factors: –.566.
****p* ≤ 0.001; ***p* ≤ 0.01; **p* ≤ 0.05.

preceding articulation involved a raised posterior tongue than when the preceding vowel was articulated with a flattened tongue dorsum and a lowered jaw. No significant interactions among any of these factors were found.

4.2. Tautosyllabic and heterosyllabic clusters

4.2.1 Acoustics. The difference in duration between tautosyllabic and heterosyllabic clusters was robust (tautosyllabic: $N=125$, $\mu=0.148$ s, $\sigma=24$; heterosyllabic: $N=245$, $\mu=0.213$ s, $\sigma=28$). The linear model that best accounted for the data included only two predictors, that is, Phonology and FollowingGesture (Table 7). The factor Phonology accounted for a difference of 0.060 seconds; in addition, clusters followed by high vowel (i.e., those in /admin/, /adria/, and /gandze/) were 0.010 seconds shorter than clusters followed by /a/ (i.e., those in /baltsa/, /kaltsa/, /batman/, and /bɔtman/).

Comparing tautosyllabic and heterosyllabic clusters within subsets of data defined by the other predictors, all comparisons turned out to be significant. We found a significant difference between tautosyllabic and stop-initial heterosyllabic clusters ($t=-13.208$, $df=78$, $p<.001$) as well as between tautosyllabic and sonorant-initial heterosyllabic clusters ($W=14$, $p<.001$); the difference was equally significant within the group of items with a preceding /a/ ($W=7$, $p<.001$) as well as for those with a preceding /o/ ($t=9.6646$, $df=38$, $p<.001$); finally, the difference was significant for items defined by both a tongue body ($t=-10.606$, $df=48$, $p<.001$) and a tongue front ($W=2$, $p<.001$) following gesture. These data confirmed that the C durational difference between tautosyllabic and heterosyllabic clusters was strong and consistently represented across the dataset.

The difference between tautosyllabic and heterosyllabic clusters was also evident in the acoustic duration of the preceding vowel, with heterosyllabic clusters showing shorter vowels ($N=245$, $\mu=0.151$ s, $\sigma=23$) than tautosyllabic ones ($N=125$, $\mu=0.172$ s, $\sigma=25$). The model that best accounted for the data included three predictors, that is, Phonology, Constriction, and FollowingGesture (Table 8). The factor Phonology accounted for a difference of 0.033 seconds

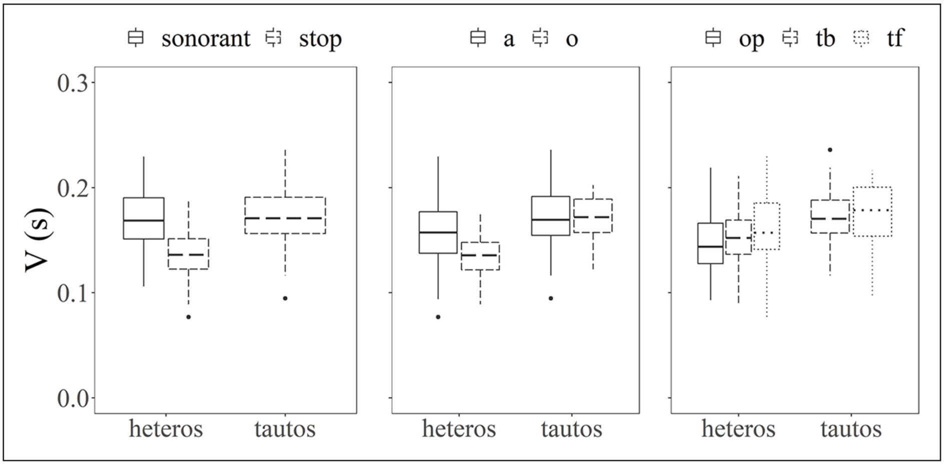

Figure 2. V duration in tautosyllabic and heterosyllabic clusters, as a function of Constriction (sonorant vs. stop), PrecedingV (/a/ vs. /o/), and FollowingGesture (opening vs. tongue body constriction vs. tongue fronting). Data subset: clusters.

(recall that the difference was of 0.060 seconds for variable C); there was an additional effect of Constriction, showing that vowels before clusters beginning with plosives (/dm/, /tm/, /dr/, /tr/, /dk/) were estimated 0.033 seconds shorter than before clusters beginning with a sonorant (/nt/, /nts/, /lts/); recall that the two groups of clusters were not found to differ significantly for C. Finally, as a mirror image of what was found in C, vowels before clusters followed by high vowels (i.e., those in /admin/, /adria/, and /gandze/) were estimated 0.015 seconds longer than vowels before clusters followed by /a/ (i.e., those in /baltsa/, /kaltsa/, /batman/, and /bɔtman/).

Independent-samples *t*-tests showed that, when the dataset was split by Constriction (stop vs. sonorant), vowels were significantly shorter before stop-initial heterosyllabic clusters /dm/, /tm/, and /dk/ than before stop-initial tautosyllabic clusters /dr/ and /tr/ ($t=8.5695$, $df=78$, $p<.001$), but, interestingly, vowels before sonorant-initial heterosyllabic clusters /nt/, /nts/, and /lts/ were not different from vowels before stop-initial tautosyllabic clusters ($t=0.41876$, $df=78$, $p=.6765$). This is illustrated in Figure 2. The heterosyllabic versus tautosyllabic difference in V duration was significant within the subset of items with a preceding /a/ ($t=2.1276$, $df=78$, $p<.05$) as well as within the subset of items with a preceding /o/ ($t=6.699$, $df=38$, $p<.001$). Finally, concerning the following articulatory gesture, the difference was significant when the following gesture involved the tongue body (/patron/, /bɔtro/, /bɔdrum/ as opposed to /pɔdkast/, /vanto/; $t=2.8202$, $df=48$, $p<.05$) but not when the following gesture involved the tongue front (/adria/ as opposed to /gandze/, /admin/).

Altogether, these data suggested that, as far as V was concerned, the difference between the two types of clusters disappeared in selected data subsets. In particular, vowels before heterosyllabic sonorant clusters showed the same duration than before tautosyllabic clusters.

Kendall's correlation test showed that V was significantly but mildly correlated inversely with C ($\tau=-0.11$, $p<.050$).

4.2.2 Tongue tip kinematics. MaxVel was on average higher in tautosyllabic clusters ($N=125$, $\mu=34.3$ cm/s, $\sigma=14.3$) than in heterosyllabic ones ($N=245$, $\mu=32.8$ cm/s, $\sigma=9.9$). However, the difference was not statistically significant. As illustrated in Table 9, there was no effect of Phonology on MaxVel variation in clusters. The model that best accounted for the data included

Table 9. Optimized Linear Mixed Model for Variable MaxVel.

	Estimate	*SE*	*t* value	Pr(>|*t*|)
(Intercept)	0.888	0.0544	20.012	<.001***
PrecedingV	−0.0854	0.0116	−4.034	.0431*

Data subset: clusters. Number of observations: 360; speakers: 10; items: 12.
***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$.

Table 10. Optimized Linear Mixed Model for the Variable Ant.

	Estimate	*SE*	*t* value	Pr(>|*t*|)
(Intercept)	0.4702	0.0191	24.68	<.001***

Data subset: clusters. Number of observations: 363; speakers: 10; items: 12.
***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$.

PrecedingV as the only significant predictor, which accounted for the fact that consonants preceded by /ɔ/ were initiated with a faster apical movement than consonants preceded by /a/; the estimated effect amounted to 8.5 cm/s.

Comparing tautosyllabic and heterosyllabic clusters within subsets of data defined by the other predictors, no comparison turned out to be significant. This confirmed that the MaxVel difference between tautosyllabic and heterosyllabic clusters was absent from the whole dataset, as well as from its subparts defined by Constriction, PrecedingV, and FollowingGesture.

Coming now to the second kinematic measure, Average Ant values were found to be higher in tautosyllabic (N=125, μ=0.496, σ=0.118) than in heterosyllabic clusters (N=245, μ=0.463, σ=0.131). However, the model that best accounted for the data was one in which none of the predictors were found to be significant (Table 10). Thus, compared with the two acoustic variables and MaxVel, Ant appeared not to be influenced by any of the factors that were taken into consideration. The difference between tautosyllabic and heterosyllabic clusters depicted above for the average Ant values was therefore irrelevant at the statistical level when the whole group of data was considered.

When we looked at subsets of data defined by the factors Constriction, Preceding V, and FollowingGesture, we found that Ant was significantly different in tautosyllabic versus heterosyllabic clusters when stop-initial clusters were considered (t=2.1299, df=71.036, p<.05), but not when sonorant-initial heterosyllabic clusters were compared with stop-initial tautosyllabic clusters (t=0.45135, df=70.785, p=.6531), a pattern which closely resembled the one found for V (§4.2.3) and which is represented in Figure 3. Also similar to V was the fact that the difference between clusters was significant when the following gesture involved the tongue body (patron/, /bɔtro/, /bɔdrum/ as opposed to /pɔdkast/, /vanto/; t=2.1391, df=48, p<.05) but not when the gesture involved the tongue front (/adria/ as opposed to /gandze/, /admin/). Finally, Ant differences were not significant in either subset of items with a preceding /a/ or /ɔ/.

5 Discussion

This study had two principal purposes: one was that of determining whether the singleton versus geminate contrast in the case of selected Italian apical consonants showed specific articulatory-kinematic correlates, and the other was that of verifying whether syllables closed by heterosyllabic clusters showed similar or different temporal and kinematic characteristics compared with

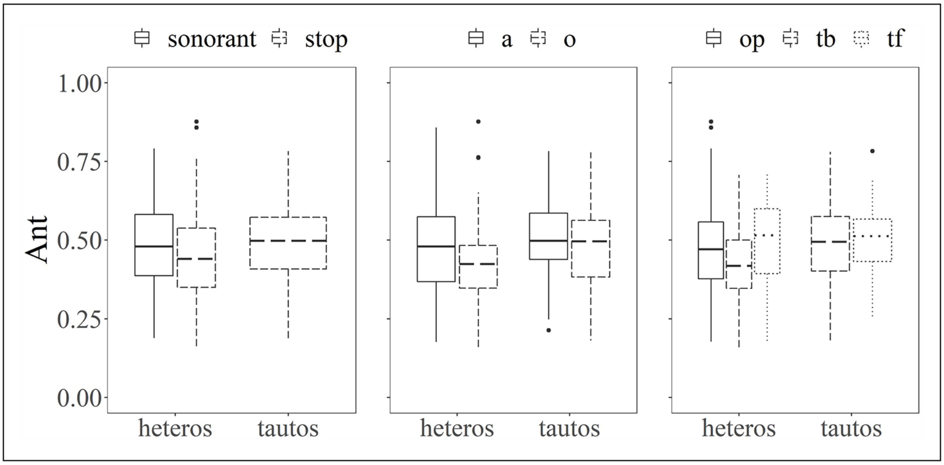

Figure 3. Ant values in tautosyllabic and heterosyllabic clusters, as a function of Constriction (sonorant vs. stop), PrecedingV (/a/ vs. /o/), and FollowingGesture (opening vs. tongue body constriction vs. tongue fronting). Data subset: clusters.

syllables closed by geminates. Concerning the first point, we could document specific kinematic correlates that suggest a different timing of the tongue tip raising gesture in intervocalic geminates as opposed to singletons; by contrast, the difference between tautosyllabic and heterosyllabic clusters was not clearly indexed by different kinematic correlates. We will now itemize and discuss the specific findings starting from the singleton versus geminate contrast and then coming to clusters.

5.1 Singletons and geminates

Our experiment has confirmed that, when gemination is investigated in isolated target disyllables, vowels before geminates are significantly shorter, compared with when they are before singletons. Unsurprisingly, the durational difference between singleton and geminate consonants was much bigger (estimated around 0.100 seconds) than the difference between the preceding vowels (0.037 seconds). The significant negative correlation between C and V additionally suggested that there is complementarity between the two, that is, the duration of the vowel is partly predicted by the duration of the entire consonantal interval, including when, as in the case of geminates, it spans a syllable boundary (cross-syllable compensation as a result of increased duration). The correlation coefficient was equal to −0.34, thus similar to what was found in bigger corpora of spoken Italian including clusters (e.g., Celata & Mairano, 2014, reported −0.36 as the maximal by-speaker value, –0.39 as the maximal by-context value, and −0.42 for the singleton/geminate subset). The absence of strict isochrony in languages is generally attributed to phonetic arguments such as the fact that vowels are incompressible beyond an absolute minimum duration to maintain the word's perceptual identity (since Klatt, 1976). However, there may be additional aspects to be considered, such as data distribution.

In this respect, Figure 4 shows that C is bimodally distributed, with one frequency peak at 0.0715 seconds and a second peak at 0.1765 seconds, whereas V is normally distributed (W = 0.99788, p = .8052), with values centered around one peak at 0.1618 seconds. Gemination is a categorical phenomenon: this is evidenced by the bimodal distribution of singletons and geminates found in the data. By contrast, vowel shortening is not categorical; rather, it is gradient, as

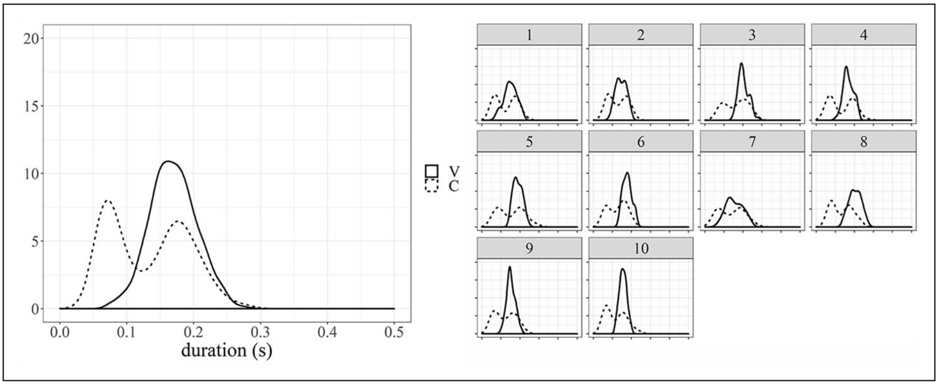

Figure 4. Density curves for variables V and C, overall (left) and split by subjects (right). Data subset: singletons and geminates.

evidenced by the non-bimodal distribution of values. This difference can already be seen as an indirect piece of evidence against a strictly moraic account of vowel duration in Italian, because moraicity predicts categoricity (vowels are either "short" or "long" if they appear in closed or open syllables, respectively), whereas vowel differences are actually gradual in the language. Moreover, the distributional characteristics of vowels and consonants directly reflect their functional difference: phonological distinction in consonants, non-phonological variation in vowels. Therefore, the data suggest that the production of the consonantal (i.e., segmental) phonological contrast is anchored to specific production dynamics that include a principle of cross-syllable compensation of the durations of vocalic and consonantal intervals.

An additional finding also pointed toward the existence of durational compensations in the data. Sonorants (both singletons and geminates) were shorter than stops (Table 3); symmetrically, vowels before sonorants were longer than before stops (Table 4). In both cases, the difference was about 0.020 seconds, thus confirming that temporal compensation was pervasive in this dataset.

In the kinematic domain, the apical gesture for a geminate was faster and began comparatively earlier during the acoustic vowel than the gesture for a singleton, indicating an increased anticipation between the nucleus and the following consonantal gesture in the case of geminates. (Recall that the Ant variable was normalized over vowel duration, and therefore this effect cannot be seen as an artifact of having acoustically shorter vowels before geminates.)

This finding was consistent with data from various languages that suggested that the closing gesture for a geminate consonant starts earlier than for a singleton (e.g., Šimko et al., 2014, for Finnish, Türk et al., 2017, for Estonian, Ridouane & Hallé, 2017, for Tashlhiyt Berber; see review in §1). However, in other languages no effect of gemination on articulatory timing was found (e.g., Moroccan Arabic, according to Zeroual et al., 2008). This suggests that the length contrasts may be differently implemented across languages not only at the level of segments' duration but also at the level of gestural organization, which has an impact on the degree of cohesion of adjacent segments (and, therefore, of consecutive syllables). Moreover, our kinematic data are only partly consistent with previous findings on Italian (§1.2). In particular, the current results are consistent with the data for the bilabial nasals and stops reported in Gili Fivela et al. (2007), Smith (1995), and Zmarich et al. (2007, one subject), but they are not consistent with the data for apical /l:/ and /n:/ in Gili Fivela et al. (2007), for various bilabial and labiodental consonants in Zmarich et al. (2007), and for /m:/ in Löfqvist (2017). As anticipated in §1, inconsistencies in the results may be due to methodological differences across studies; see, for instance, the arguments by Löfqvist (2017) about the way of quantifying the synchronization of lower lip and tongue movements in the production of /m:/. Another complication in the case of (bilabial) nasals might be related to the presence of a velum gesture, which is not necessarily synchronous with the oral constriction gesture but nevertheless ensures the realization of acoustically prominent resonances. In this respect, our UTI-based analysis being focused on apical stops and sonorants (preceded by low or mid-low vowels) could rather easily track tongue tip movements, as opposed to dorsal movements required for the preceding vowels.

Both kinematic effects of gemination (increased closure velocity and increased temporal anticipation) were extremely robust in our data, as they appeared to be consistently present across consonants (sonorants, stops) and phonetic contexts.

Apart from correlating to the singleton versus geminate distinction, Ant was also found to vary significantly as a function of vowel quality and consonantal constriction. Stops, which were overall acoustically longer than sonorants, were preceded by acoustically shorter vowels and articulated with a proportionally greater anticipation of the constriction gesture, compared with sonorants. Although both stops and sonorants in our dataset were anterior coronals, their manner of distinction turned out to be consistently manifested by acoustic as well as kinematic differences. Concerning vowel variations, when the vowel involved a posterior tongue movement, the tongue tip for the consonant started raising earlier (and tended to move faster, Table 5) than when the vowel was produced with a low, flat tongue dorsum. Such contextual variation suggests that apical consonants are differently timed depending on the articulatory nature of the adjacent segments, even if—as in the case of vowels—the tongue tip is not actively involved. Our findings are therefore consistent with previous knowledge of apicals' sensitivity to contextual variations documented in the acoustic and articulatory domains (since, for example, Keating, 1991; Recasens et al., 1997), furthermore providing new evidence in the kinematic domain.

5.2 Clusters

We compared intervocalic tautosyllabic and heterosyllabic clusters to see whether the temporal and kinematic characteristics defining the singleton versus geminate contrast also held for the cluster contrast, under the assumption that in both cases there was an opposition between an open and a closed syllable. There is a substantial asymmetry in Italian as far as heterosyllabic clusters are phonetically more varied (both stop-stop and sonorant-stop clusters) than tautosyllabic clusters (only stop-liquid clusters), and this asymmetry was directly reflected in the choice of the materials for the current experiment. These distributional differences between different types of clusters make the cluster alternation phonetically and phonotactically less regular than the singleton versus geminate alternation. Assuming that they are part of the phonological grammar of the language, we hypothesized that acoustic vowel shortenings and gestural anticipations of the post-vocalic consonant were less regular in the case of the cluster alternation, compared with the singleton versus geminate alternation, despite the fact that traditional moraic accounts treat both alternations as a contrast between an open and a closed syllable.

A first result in the domain of acoustic measurements concerned the fact that like geminates were significantly longer than singletons, so heterosyllabic clusters were significantly longer than tautosyllabic ones; moreover, vowels before heterosyllabic clusters were significantly shorter. The difference between clusters was estimated around 0.060 seconds, thus lower than between singletons and geminates, but still relevant[6] and statistically significant. By contrast, the difference between the preceding vowels amounted to 0.033 seconds, that is, somehow smaller than in the singleton versus geminate contrast (where it amounted to 0.037 seconds). Then at the group level,

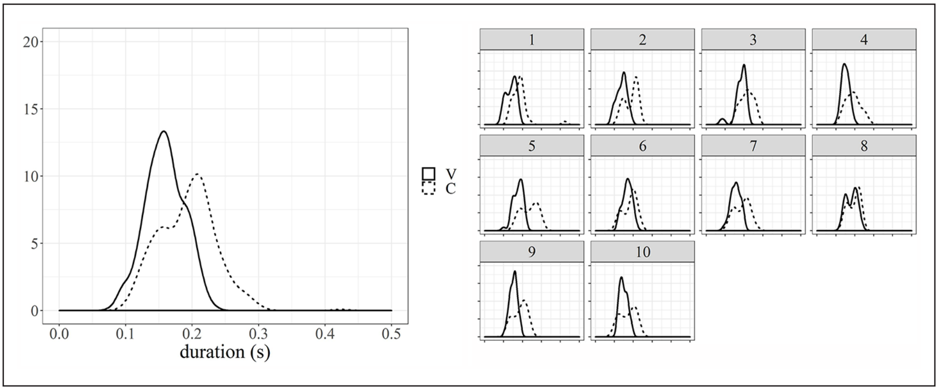

Figure 5. Density curves for variables V and C, overall (left) and split by subjects (right). Data subset: clusters.

we can conclude that the temporal properties of the singleton versus geminate contrast were also present in the tautosyllabic versus heterosyllabic cluster comparison. Figure 5 additionally shows that data distribution is also similar at the group level: C shows a non-normal distribution (W=0.97527, p < .001, with a main peak at 0.2081 seconds and a secondary lower peak at 0.1523 seconds), and V is normally distributed (W=0.99492, p=.2649). However, there was much more individual variation, with plots on the right of Figure 5 showing inconsistent evidence of a bimodal distribution in the C variable and inconsistent evidence of a normal distribution in the V variable.

A second relevant aspect concerns the difference between stop-initial and sonorant-initial clusters. While the two groups did not differ as far as the duration of the consonantal interval was concerned (Table 7), V was significantly different, with vowels before stop-initial clusters showing a shorter duration than vowels before sonorant-initial clusters (Table 8). This result suggests that vowel variations before clusters may be determined by factors that are unrelated to the cluster's tautosyllabicity versus heterosyllabicity. Vowel duration varied according to the sonorant versus stop distinction also in the singleton/geminate dataset (see above, §5.1). However, in that case the consonants also varied, consistently with the principle of durational compensation that we found in that dataset. By contrast, in the cluster dataset which is at issue here, only vowels vary, whereas stop-initial and sonorant-initial clusters show the same average duration. In this respect, then, the cluster dataset shows evidence of vowel variation which is less predictable and apparently irreducible to heterosyllabicity.

A third source of evidence for the rhythmical difference between pre-geminates and pre-cluster vowels is that the difference between tautosyllabicity and heterosyllabicity had an impact on the duration of the preceding vowel only when stop-initial clusters were taken into account (Figure 2). Finally, the fact that the inverse correlation between cluster and vowel duration was very low (τ=–0.11) further suggests that another characteristics of the singleton versus geminate contrast, that is, the tendency to distributing vocalic and consonantal intervals into slots of complementary duration, is not implemented in the sequences with clusters to the same extent. Recall that compensation is inherently cross-syllabic, as it presupposes that the duration of the entire consonantal interval, spanning from the coda of a syllable to the onset of the subsequent syllable, influences the duration of the nucleus of the first syllable (Farnetani & Kori, 1986; Fowler, 1983; Smith, 1992). In our data, we can thus conclude that complementarity characterizes the segmental phonological contrast more clearly than the contrast between tautosyllabic and heterosyllabic clusters.

The following phonetic context was found not to affect the duration of acoustic vowels and consonants in the singleton/geminate dataset, yet it did affect the duration of both consonants and vowels in the cluster dataset (Tables 7 and 8): there was a symmetric effect of vowel shortening and cluster lengthening before /i/, whose motivation will not be investigated given that the clusters were composed of different segments. More importantly, there was no interaction between the effects of the phonetic context and the effects of the phonological contrast between tautosyllabicity and heterosyllabicity. It should therefore be concluded that in the cluster dataset, the effects of contextual variability were overall stronger than in the singleton/geminate dataset.

Concerning the kinematic findings, the models did not provide any statistical evidence of an effect of either MaxVel or Ant. Therefore, differently from singletons and geminates, tautosyllabic and heterosyllabic clusters did not differ systematically for the kinematic properties investigated here.

There was, however, an important effect of the factor Constriction. Within stop-initial clusters, tautosyllabicity was manifested by less gestural anticipation (higher Ant values) and heterosyllabicity by increased gestural anticipation (lower Ant values) (Figure 3). Given that vowel duration also changed in the same way (Figure 2), we may conclude that, among stop-initial clusters, heterosyllabic clusters were preceded by shorter vowels and were produced with a proportionally greater anticipation of the consonantal gesture, compared with tautosyllabic clusters. However, when sonorant-initial heterosyllabic clusters were considered, vowel duration and gestural anticipation did not differ from tautosyllabic stop initial clusters. Moreover, Figure 3 shows that Ant distinguished between clusters only in the context of low vowels. Therefore, the contrast between tautosyllabicity and heterosyllabicity in clusters is manifested at the levels of vowel shortening and gestural anticipation in only a small subset of the data, which correspond to a small subset of the lexicon of the language. At the level of peak velocity, there is no significant difference between clusters.

6 Conclusion

This study has shown that heterosyllabicity may be related to different properties in the speech signal and in speech articulation. This means that, in a phonetically grounded view of phonological relations, the notion of heterosyllabicity has to be specified in terms of the rhythmical and gestural consequences that it bears in the language.

Our results suggest that the contrast between tautosyllabic and heterosyllabic clusters in Italian disyllables is much more weakly cued at the level of acoustic and articulatory correlates, especially if compared with the contrast between singletons and geminates. Gemination has robust correlates not only in the durational variations of the consonant and of the preceding vowel, but also in the velocity and the relative timing of the tongue tip constriction gesture, which occurs comparatively earlier than in intervocalic singletons. Robustness derives from systematicity across subjects and phonetic conditions. Peak velocity and gestural anticipation, besides vowel shortening, should all be considered as phonetic ways to enhance the contrast between phonologically long and short consonants. Contrast enhancing through systematic variation of production dynamics occurs frequently in languages, and the speakers have an auditory advantage from that (e.g., Dmitrieva, 2019; Storme, 2019). All stops and sonorants (plus fricatives) can be geminated in Italian; such widespread distribution of the phonological length feature corresponds to systematic acoustic and kinematic variations between singletons and geminates.

By contrast, heterosyllabic clusters show variable vowel shortening, with optionality at the level of individual subjects, and absent or weak kinematic correlates, with limitations related to the specific type of cluster and the quality of the surrounding phonetic context. In clusters, the phonetic context has a much stronger effect in modulating both acoustic durations and gesture timing relations. Clusters themselves are complex phonetic objects, compared with geminates whose production differences with respect to the corresponding singletons are primarily related to the temporal extension of the constriction. Moreover, our dataset was asymmetric to the extent that it directly incorporated the asymmetry of the Italian phonotactic system which allows two types of heterosyllabic clusters (i.e., stop-initial and sonorant-initial clusters) and only one type of tautosyllabic clusters (stop-initial clusters), thus including a source of inescapable variability in the two groups of items. The phonetic complexity of clusters, alongside the asymmetric distribution of clusters in the dataset, can thus explain their variable behavior and the lack of articulatory routines related to cluster syllabification, as well as the weakness of the vowel duration effects. There is no kinematic and very little acoustic evidence of heterosyllabicity as such in clusters.

The compensatory mechanisms which make vowels to shorten proportionally to the amount of lengthening in consonants (although strict isochrony is far to be reached) is also more effective in the case of intervocalic singletons and geminates than in the case of clusters, consistently with the primary role that the duration of the constriction gesture plays in the singleton versus geminate distinction.

Taken together, these data suggest that, in the production of disyllables by Italian speakers, systematic variations in acoustic vowel duration and in the kinematics of tongue tip gestures represent the phonetic correlates of the segmental phonological contrast between short and long consonants, rather than of syllable structure. To be considered phonetic correlates of heterosyllabicity, the same systematic variations should have been found in vowel-cluster sequences, which was not the case in this study. Stressed vowel shortening, showing different acoustic and articulatory characteristics before geminates and before heterosyllabic clusters, should be considered as bearing different functions in the phonology of the language and indexing the length feature in the case of gemination. This view is consistent with typological and historical data that establish a clear link between vowel shortening and gemination. For instance, in various languages contrastive vowel quantity arose as a reaction to the process of consonant degemination (e.g. in Breton, Bothorel, 1982; in Northern Italian dialects, Loporcaro, 2007). Similarly, typological surveys show that 26 languages have vowel shortening in the context of gemination but only 9 have vowel shortening in the context of a singleton syllable coda (Maddieson, 1984).

Our data confirm Smith's (1992) and Farnetani and Kori's (1986) findings about the tendency for Italian to show isochrony within the vowel-to-vowel interval, especially in disyllables with singletons and geminates. However, our data are only partly consistent with the predictions of both moraic and gesture-based models of the syllable. With respect to the moraic model, it may be problematic that not all types of closed syllables show the same rhythmic and kinematic correlates. In particular, the traditional assumption that the duration of the preceding stressed vowel can be directly predicted by syllabic constituency appears, in the light of the current data, inadequate to account for the entire picture of heterosyllabic consonantal sequences in Italian. Our kinematic data, in particular, show that no gestural timing routines can be established to account for the tautosyllabicity versus heterosyllabicity contrast in clusters; in parallel, variation at the acoustic output is linked to various phonetic and contextual factors that are unrelated to syllabification. Moreover, the fact that heterosyllabic clusters do not necessarily imply vowel shortening might suggest that some of them give rise to extra-heavy syllables, something that is generally considered dispreferred in languages, although not impossible (Gordon, 2006). For instance, Moosmüller and Brandstätter (2014) show that Middle Bavarian dialects, generally considered to exhibit a rigid pattern of durational compensation (short vowels before *fortis* stops, long vowels before *lenis* stops), do indeed include extra-heavy syllables with *fortis* stops preceded by long vowels; the authors derive from these data that a strictly moraic account is insufficient to describe the temporal organization of these dialects and that although "the quantity interaction [. . .] is [. . .] reflected in a strong correlation between vowel and consonants, [. . .] the type V:F is not integrated in this pattern, but takes an intermediate position" (Moosmüller and Brandstätter, 2014, p. 93).

By contrast, in the case of gestural models of syllable structure, the fact that the increased gestural anticipation of codas is variable in the phonological contexts (geminates vs. clusters) can only be accounted for if the timing relationships between gestures within the syllable are conceived of as very general templates over which language-, context-, and possibly also speaker-specific constraints apply. As a matter of fact, this view is receiving increasing support from cross-linguistic investigation of a variety of phenomena related to the internal organization of syllables (e.g., Chen et al., 2017; Pouplier, 2015; Pouplier et al., 2020), showing the role of phonetic segmental factors in shaping the articulatory cohesion within the syllable. Gestural cohesion effects spanning syllable boundaries also call for a more gradient view of articulatory syllabification.

Acknowledgements

This contribution is in loving memory of Sylvia Moosmüller, of her warm friendship, and of the vivacity of our discussions which greatly contributed to the implementation of this study. The authors are especially grateful to the Editors and two anonymous Reviewers for their helpful comments on a previous version of this contribution.

Author contributions

Conceptualization: C. Celata; design of the experiment: C. Celata and C. Meluzzi; recordings, data pre-processing, and annotation: C. Meluzzi; acoustic and articulatory phonetic analysis: C. Celata and C. Meluzzi; statistical analysis: C. Bertini; visualization: C. Bertini; writing—original draft preparation: C. Celata (§1, 2, 4, 5, 6), C. Meluzzi (§3.1, 3.2, 3.3) and C. Bertini (§3.4); writing—review and editing: all authors; supervision: C. Celata. All authors have read and agreed to the published version of the manuscript.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Scuola Normale Superiore grant SNS15C_A to C. Celata for the project "Gestural coordination and speech rhythm: Implications for pathological speech" and by the Italian Ministry of Education, University and Research (MIUR)—FFABR 2018 grant to C. Celata.

ORCID iD

Chiara Celata https://orcid.org/0000-0001-6748-7675

Notes

1. For the sake of completeness, it should be recalled that phonological length has articulatory correlates also in the case of vowels: gestures have been shown to be longer, faster, and spatially more extended in long versus short vowels (e.g., Beňuš, 2010, for Slovak; Türk et al., 2017, for Estonian).
2. Again to be complete, there are mixed results also for vowels: for instance, German show little or no effect of quantity on the gesture velocity to the following consonantal constriction (although gestures are overall longer for tense/long vowels than for lax/short ones; Hertrich & Ackermann, 1997).
3. Löfqvist (2017) failed at replicating the findings in Smith (1995) for Italian intervocalic /m/ and /m:/, showing that the interval between the lip release for the consonant and the tongue movement for the vowel did not change as a function of phonological length.
4. C stands for plosive or nasal consonant, R for liquid, and V for vowel. Note that some of our stimuli were vowel-initial disyllables, so the structures in those cases would be VCV versus VC:V and VCRV versus VRCV/VCCV.

5. /ts/ and /dz/ represent the alveodental affricates, respectively, voiceless and voiced. The distribution of voicing varies across regional varieties of Italian (Meluzzi, 2016). The form in which words with affricates appear in Table 1 corresponds to the Tuscan Italian pronunciation norm, followed by the participants in the experiment.
6. The just noticeable difference (JND) for consonants is considered by Lehiste (1970, p. 13) to be about 10–40 ms.

References

Articulate Instruments Ltd. (2008). *Ultrasound stabilisation headset users manual: Revision 1.4*.

Articulate Instruments Ltd. (2012). *Articulate assistant advanced user guide: Version 2.14*.

Avesani, C., Vayra, M., & Zmarich, C. (2009). Coordinazione vocale-consonante e prominenza accentuale in italiano. La sfida della Articulatory Phonology [Vowel-consonant coordination and accentual prominence in Italian. The challenge of Articultory Phonology]. In G. Ferrari, R. Benatti, & M. Mosca (Eds.), *Linguistica e modelli tecnologici di ricerca, Atti del XL Congresso Internazionale della Società di Linguistica Italiana (SLI)* (pp. 353–386). Bulzoni.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beňuš, S. (2011). Control of phonemic length contrast and speech rate in vocalic and consonantal syllable nuclei. *The Journal of the Acoustical Society of America*, *130*(4), 2116–2127.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (Version 6.0.37.) [Computer program]. http://www.praat.org/

Bothorel, A. (1982). *Étude phonétique et phonologique du breton parlé à Argol (Finistère-sud). [Phonetic and phonological study of Argol Breton]*. Lille: Atelier reproduction des thèses, Lille III.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). CSLI.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*, 149–180.

Calamai, S., & Celata, C. (2011, August 27–31). *Timing in Italian VNC sequences at different speech rates [Conference]. Interspeech2011 – 12th Annual Conference of the International Speech Communication Association*, *Florence*.

Celata, C., & Mairano, P. (2014). On the timing of V-to-V intervals in Italian: A review, and some new hypotheses. *Revista de Filología Románica*, *31*, 37–53.

Chen, C., Celata, C., & Ricci, I. (2017). An EPG + UTI study of syllable onset and coda coordination and coarticulation in Italian. In C. Bertini, C. Celata, C. Meluzzi, G. Lenoci, & I. Ricci (Eds.), *Fattori sociali e biologici nella variazione fonetica—Social and biological factors in speech variation* (Vol. 21, pp. 151–172). Officina.

D'Imperio, M. P., & Rosenthall, S. (1999). Phonetics and phonology of main stress in Italian. *Phonology*, *16*(1), 1–28.

Dipino, D., & Celata, C. (2018). An UTI study of alveolar stops in Italian. In A. Vietti, L. Spreafico, D. Mereu, & V. Galatà (Eds.), *Il parlato nel contesto naturale—Speech in the natural context* (pp. 41–53). Officinaventuno.

Dmitrieva, O. (2019). Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics*, *73*, 128–143.

Esposito, A., & Di Benedetto, M. G. (1999). Acoustical and perceptual study of gemination in Italian stops. *The Journal of the Acoustical Society of America*, *106*(4), 2051–2062.

Farnetani, E., & Kori, S. (1986). Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication*, *5*(1), 17–34.

Fava, E., & Magno Caldognetto, E. (1976). Studio sperimentale delle caratteristiche elettroacustiche delle vocali toniche e atone in bisillabi italiani [An experimental study of the electroacoustic features of stressed and unstressed vowels in Italian disyllables]. In R. Simone, et al. (Eds.), *Atti del Convegno Internazionale di "Studi di Fonetica e Fonologia"* (Padova, 1–2 ottobre 1973) (pp. 35–79). Bulzoni.

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica*, *38*, 35–50.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*(3), 386–412.

Gili Fivela, B., Zmarich, C., Perrier, P., Savariaux, C., & Tisato, G. (2007). *Acoustic and kinematic correlates of phonological length contrast in Italian consonants* [Proceedings]. *XVIth International Congress of Phonetic Sciences, Saarbrücken, Germany*, 6–10 August 2007.

Gordon, M. (2006). *Syllable weight: Phonetics, phonology, typology*. Routledge.

Gordon, M. (2016). *Phonological typology*. Oxford University Press.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Hertrich, I., & Ackermann, H. (1997). Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *Journal of the Acoustical Society of America*, *102*, 523–536.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). *The TenTen corpus family* [*Conference*]. *7th International Corpus Linguistics Conference (CL2013)*, Lancaster, UK, 22–26 July 2013.

Keating, P. (1991). Coronal places of articulation. In C. Paradis & J.-F. Prunet (Eds.), *Phonetics & Phonology: The special status of coronals* (pp. 29–48). Academic Press.

Kenstowicz, M. (2017). A note on the phonology and phonetics of CR, RC, and SC consonant clusters in Italian. In H. Quinn, D. Massam, & L. Matthewson (Eds.), *Linguistic travels in time and space: Festschrift for Liz Pearce Wellington*. Working Papers in Linguistics (Vol. 23, pp. 87–100). Victoria University of Wellington.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221.

Krämer, M. (2009). *The phonology of Italian*. Oxford University Press.

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

Lin, S., Beddor, P. S., & Coetzee, A. W. (2014). Gestural reduction, lexical frequency, and sound change: A study of post-vocalic /l/. *Laboratory Phonology*, *5*(1), 9–36.

Löfqvist, A. (2005). Lip kinematics in long and short stop and fricative consonants. *Journal of the Acoustical Society of America*, *117*(2), 858–878.

Löfqvist, A. (2006). Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese. *Journal of Acoustical Society of America*, *120*(5), 2872–2883.

Löfqvist, A. (2007). Tongue movement kinematics in long and short Japanese consonants. *Journal of Acoustical Society of America*, *122*(1), 512–518.

Löfqvist, A. (2017). Articulatory coordination in short and long consonants: An effect of rhythm class? In H. Kubozono (Ed.), The Phonetics and Phonology of Geminate Consonants (pp. 118–129). Oxford University Press.

Loporcaro, M. (2005). La sillabazione di muta cum liquida dal latino al romanzo [Muta cum liquida syllabification from Latin to Romance]. In S. Kiss, L. Mondin, & G. Salvi (Eds.), *Latin et langues romanes. Études linguistiques offertes à Jozsef Herman à l'occasione de son 80ème anniversaire* (pp. 19–43). Niemeyer.

Loporcaro, M. (2007). Facts, theory and dogmas in historical linguistics: Vowel quantity from Latin to Romance. In J. C. Salmons & S. Dubenion-Smith (Eds.), *Historical Linguistics 2005. Selected papers from the 17th International Conference on Historical Linguistics, Madison, Wisconsin, 31 July–5 August 2005* (pp. 311–336). John Benjamins.

Maddieson, I. (1984). Phonetic cues to syllabification. *UCLA Working Papers in Phonetics*, 59, 85–101.

Mairano, P., & De Iacovo, V. (2020). Gemination in northern versus central and southern varieties of Italian: A corpus-based investigation. *Language & Speech*, *63*(3), 608–634.

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, *14*(3), 380–407.

Marotta, G. (1985). *Modelli e misure ritmiche: La durata vocalica in italiano [Rhythmical models and measures: Vowel duration in Italian]*. Zanichelli.

McCrary, K. (2004). *Reassessing the role of the syllable in Italian phonology: An experimental study of consonant cluster syllabification, definite article allomorphy and segment duration* [Doctoral dissertation, University of California, Los Angeles].

McMillan, C. T., Corley, M., & Lickley, R. J. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, *24*(1), 44–66.

Meluzzi, C. (2016). A new sonority degree in the realization of dental affricates /ts dz/ in Italian. In M. J. Ball & N. Müller (Eds.), *Challenging sonority. Cross-linguistic evidence* (pp. 252–275). Equinox.

Moosmüller, S., & Brandstätter, J. (2014). Phonotactic information in the temporal organization of Standard Austrian German and the Viennese dialect. *Language Sciences*, *46*, 84–95.

Munhall, K., Fowler, C., Saltzman, E., & Hawkins, S. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, *20*, 225–239.

Öhman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America*, *41*, 310–320.

Pastätter, M., & Pouplier, M. (2014, May 5–8). *The temporal coordination of Polish onset and coda clusters containing sibilants* [*Proceedings*]. *10th International Seminar on Speech Production*, *Cologne, Germany*.

Payne, E. M. (2006). Non-durational indices in Italian geminate consonants. *Journal of the International Phonetic Association*, *36*(1), 83–95.

Pouplier, M. (2015). Between "whims of fashion" and "phonetic law": Performance constraints in speech production in the face of linguistic diversity. In M. Vayra, C. Avesani, & F. Tamburini (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destruttutrazione della strutura sonora del linguagguo / Language acquisition and language loss. Acquisition, change and disorders of the language sound structure* (Vol. 21, pp. 41–57). Officina.

Pouplier, M., Lentz, T. O., Chitoran, I., & Hoole, P. (2020). The imitation of coarticulatory timing patterns in consonant clusters for phonotactically familiar and unfamiliar sequences. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1), 1–41.

R Core Team. (2018). *R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria)*. https://www.R-project.org/

Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, *102*, 544–561.

Ridouane, R., & Hallé, P. (2017). Word-initial geminates: From production to perception. In H. Kubozono (Ed.), *The phonetics and phonology of geminate consonants* (pp. 66–84). Oxford University Press.

Shaiman, S. (2001). Kinematics of compensatory vowel shortening: The effect of speaking rate and coda composition on intra- and inter-articulatory timing. *Journal of Phonetics*, *29*(1), 89–107.

Šimko, J., O'Dell, M., & Vainio, M. (2014). Emergent consonantal quantity contrast and context-dependence of gestural phasing. *Journal of Phonetics*, *44*, 130–151.

Smith, C. L. (1992). *The temporal organization of vowel and consonant gesture* [PhD Dissertation, Yale University].

Smith, C. L. (1995). Prosodic patterns in the coordination of vowel and consonant gestures. In B. Connell & A. Arvaniti (Eds.), *Phonology and Phonetic Evidence Papers in Laboratory Phonology IV* (pp. 205–222). Cambridge University Press.

Storme, B. (2019). Contrast enhancement as motivation for closed syllable laxing and open syllable tensing. *Phonology*, *36*(2), 303–340.

Strycharczuk, P., & Scobbie, J. (2015). *Velocity measures in ultrasound data. Gestural timing of post-vocalic /l/ in English [Proceedings]*. *18th International Congress on Phonetic Sciences*, Glasgow, UK, 10–14 August 2015. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html

Timpanaro, S. (1965). Muta cum liquida in poesia latina e nel latino volgare [Muta cum liquida in Latin poetry and Vulgar Latin]. *Rivista di cultura classica e medioevale*, *7*, 1075–1103.

Tomaschek, F., Tucker, B. V., Fasiolo, M., & Baayen, R. H. (2018). The consequences of lexical proficiency for articulation. *Linguistic Vanguard*, *4*(s2), Article 20170018.

Türk, H., Lippus, P., & Šimko, J. (2017). Context-dependent articulation of consonant gemination in Estonian. *Laboratory Phonology*, *8*(1), 1–26.

Vayra, M., Avesani, C., & Fowler, C. (1999). On the phonetic bases of vowel-consonant coordination in Italian: A study of stress and "compensatory shortening." In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (Eds.), *Proceedings of the XIV International Congress of Phonetic Sciences, San Francisco*, CA, USA, 1–7 August 1999 (Vol. 1, pp. 495–498).

Venneman, T. (2012). Structural complexity of consonant clusters: A phonologist's view. In P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, & B. Kühnert (Eds.), *Consonant clusters and structural complexity* (pp. 9–32). Mouton de Gruyter.

Vogel, I. (1982). *La sillaba come unità fonologica* [The syllable as a phonological unit]. Zanichelli.

Zeroual, C., Hoole, P., & Gafos, A. I. (2008). Spatio-temporal and kinematic study of Moroccan Arabic coronal geminate plosives. In R. Sock, S. Fuchs, & Y. Laprie (Eds.), *Proceedings of the 8th International Seminar on Speech Production, Strasbourg, France* (pp. 135–138). Institut national polytechnique de Lorraine.

Zmarich, C., Gili Fivela, B., Perrier, P., Savariaux, C., & Tisato, G. (2007). Consonanti scempie e geminate in italiano: Studio acustico e cinematico dell'articolazione linguale e bilabiale [Singleton and geminate consonants in Italian: Acoustic and kinematic study of the lingual and bilabial articulation]. In V. Giordani, V. Bruseghini, & P. Cosi (Eds.), *Atti del Convegno Nazionale AISV, Scienze vocali e del linguaggio. Metodologie di valutazione e risorse linguistiche* (pp. 151–163). EDK Editore.