



**HAL**  
open science

# Articulatory consequences of lexical stress on post-tonic velar plosives in Italian

Bowei Shao, Anne Hermes, Philipp Buech, Maria Giavazzi

► **To cite this version:**

Bowei Shao, Anne Hermes, Philipp Buech, Maria Giavazzi. Articulatory consequences of lexical stress on post-tonic velar plosives in Italian. *Journal of Phonetics*, 2025, 108, pp.101371. 10.1016/j.wocn.2024.101371 . hal-04827047

**HAL Id: hal-04827047**

**<https://hal.science/hal-04827047v1>**

Submitted on 9 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Articulatory consequences of lexical stress on post-tonic velar plosives in Italian <sup>☆,☆☆</sup>

Bowei Shao <sup>a,b</sup> Anne Hermes <sup>b</sup> Philipp Buech <sup>b</sup> Maria Giavazzi <sup>a,\*</sup>

<sup>a</sup> Département d'études cognitives, École Normale Supérieure, Université Paris Sciences & Lettres, 29 rue d'Ulm, Paris 75005, France

<sup>b</sup> Laboratoire de Phonétique et Phonologie, UMR 7018, (CNRS & Université Sorbonne Nouvelle), 4 rue des Irlandais, Paris 75005, France

## ARTICLE INFO

### Article history:

Received 15 September 2023

Received in revised form 15 September 2024

Accepted 13 October 2024

Available online xxx

### Keywords:

Lexical stress

Articulation

Palatalization

Italian

$\mu$ -gesture

Articulatory phonology

## ABSTRACT

Lexically prominent positions are phonologically privileged: they are often phonetically strengthened and they are *loci* of contrast preservation. Cross-linguistically, stress-conditioned alternations target stress-adjacent consonants independently of syllabic boundaries. We argue that the phonetic bases of these processes can be found in the articulatory modulations induced by stress. They are anchored in the stressed vowel but have spill-over effects on adjacent consonants. In this study, we investigate the articulation of velar consonants in a palatalizing context. By comparing two conditions, with or without stress modulations, we aim to investigate potential articulatory underpinnings of a stress-conditioned phonological process, i.e., velar palatalization in Italian plural nouns and adjectives, which is largely blocked in post-tonic position. Using articulatory data (EMA), we show that lexical stress induces temporal and spatial modulations on post-tonic velar consonants. Temporal modulations surface with a delayed target achievement of the consonants' constriction gestures. Spatial modulations surface with a further back place of articulation in post-tonic velars. Both effects are due to the strengthening of the stressed vowel. We discuss the implications of our findings within the  $\mu$ -gesture proposal of Articulatory Phonology for the distribution of palatalization in Italian.

© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

This paper investigates articulatory modulation patterns in Italian associated with stressed vowels and their adjacent consonants, which may be phonetic precursors of stress-conditioned consonantal alternations. The paper presents an articulatory study on the velar consonants [k, g] in a coarticulatorily palatalizing context (i.e., followed by [i]). By comparing

these consonants in two conditions, with and without stress-modulation, we aim to investigate potential articulatory underpinnings of phonological “blocking” of velar palatalization in Italian, which is stress-conditioned. First, we introduce an interaction between lexical stress and supra-laryngeal articulation, focusing on a proposal accounting for the stress-induced modulations: the  $\mu$ -gesture proposal within Articulatory Phonology (AP). Second, we present briefly the phonological and acoustic aspects of stress-conditioning cross-linguistically, focusing on the phonetic bases of palatalization and its stress-conditioning in Italian.

### 1.1. Lexical stress and effects on supralaryngeal articulation

The articulatory characteristics of lexical stress have often been described in terms of sonority expansion or hyper-articulation (Beckman et al., 1992; de Jong et al., 1993; de Jong, 1995; de Jong, 2004). Both accounts of prominence marking consider the phonetic nature of stress as an expansion of certain key articulatory features over space and time

<sup>☆</sup> The results presented in this article are part of a bigger project devoted to investigating the role of phonetic substance in prosodic conditioning in Italian from both diachronic and synchronic perspectives (ANR-JCJC Dia-Syn-Phon, ANR-21-CE28-0008), with Maria Giavazzi as the principal investigator. This work is also supported by Ecole Universitaire de Recherche de l'Université PSL (Frontiers in Cognition, ANR-17-EURE-0017) and IdEx PSL (ANR-10-IDEX-0001). This work is also partially supported by a public grant overseen by IdEx Université Paris Cité (ANR-18-IDEX-0001) as part of Labex Empirical Foundations of Linguistics - EFL.

<sup>☆☆</sup> The experimental protocol for this study was submitted to the Comité d'Éthique de la Recherche de l'Université Paris Cité, and subsequently approved (N° 2022-106-SHAO-HERMES).

\* Corresponding author.

E-mail addresses: [bowei.shao@sorbonne-nouvelle.fr](mailto:bowei.shao@sorbonne-nouvelle.fr), [anne.hermes@sorbonne-nouvelle.fr](mailto:anne.hermes@sorbonne-nouvelle.fr), [philipp.buech@sorbonne-nouvelle.fr](mailto:philipp.buech@sorbonne-nouvelle.fr), [maria.giavazzi@ens.psl.eu](mailto:maria.giavazzi@ens.psl.eu)

(Iskarous and Pouplier, 2022). This expansion is viewed by Beckman and Edwards (1994) as being derived from relationships among many different content features specified for stressed vowels. The phonological content of lexical stress (i.e., the lower level in the prosodic hierarchy compared to phrasal prominence) involves a more open vocal tract, and hence a lower jaw and larger lip displacement in transitioning into the vowel. The duration difference is also an inherent part of the specification of the prominence contrast of lexical stress (Beckman and Edwards, 1994<sup>1</sup>). This expansion can also be seen as a hyperarticulation (de Jong, 1995) of stressed vowels. When stressed, vowels have a lower jaw, and anterior stop consonants have tighter closure. More generally, open sounds are more open, and closed sounds are more closed.

This expansionary modulation of the articulatory system needed for the realization of lexical stress is described within the Articulatory Phonology/Task Dynamics (AP/TD) framework in gestural grammar (Browman and Goldstein, 1992; Gafos and Benus, 2006; Goldstein et al., 2006). Stress-induced articulatory strengthening induces larger, longer, and faster lip opening and closing movements. This means that the articulatory movements of the stressed vowel are bigger in all ways – in distance, time, and speed (Cho, 2006). These modulations are modeled in terms of an abstract spatio-temporal gesture, the  $\mu$ -gesture, associated to the stress domain (Saltzman et al., 2008), as shown in Fig. 1.

Concretely, temporal differences between stressed and unstressed syllables are in this view the result of a temporal  $\mu_T$ -gesture being active during a stressed syllable. This temporal  $\mu_T$ -gesture operates at the level of gestural planning and modulates the stiffness of a syllable-planning oscillator (Iskarous and Pouplier, 2022) and slows the gestural activation of co-active gestures (Krivokapić, 2020). As a consequence, stressed syllables are longer than unstressed syllables. The spatial  $\mu_S$ -gesture has not received any definition yet (Iskarous and Pouplier, 2022). Recent empirical studies from Katsika (2016) and Katsika and Tsai (2021) have however shown that a spatial effect may be independent from a temporal one.

The stress domain is larger than the stressed syllable itself and extends into (at least) the following unstressed syllable to its right within a word (Turk and Sawusch, 1997). The  $\mu$ -gesture model also predicts that the stress domain may be larger than the stressed syllable itself; it is predicted to affect articulatory properties of pre-tonic as well as of post-tonic segments regardless of a syllabic boundary. Katsika and Tsai (2021) reported that in Greek, a  $\mu$ -conditioned vowel has longer, larger, and faster gestures than its unstressed counterparts. This stress-conditioned modulation has a direct impact on the surrounding sounds. Henceforth, we follow Katsika (2016) and Katsika and Tsai (2021) in referring to these impacts which extend beyond the stressed vowel as anticipatory and spill-over effects of the modulation. Katsika and Tsai (2021) also showed that the spill-over effect is stronger than the anticipatory one. The  $\mu$ -gesture approach is thus useful for the investigation of the phonetic bases of stress-

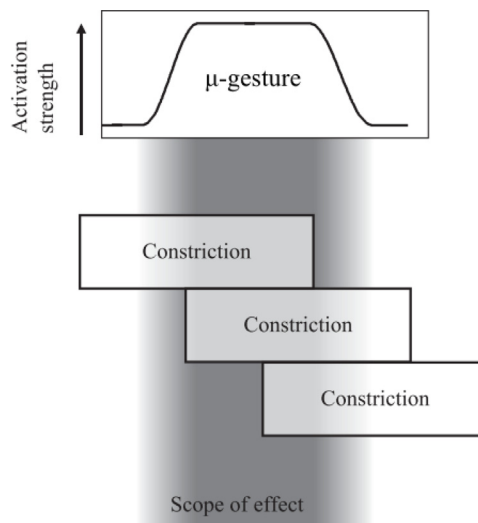


Fig. 1. A schematic representation of the  $\mu$ -gesture model showing the overlap of the  $\mu$ -gesture and constriction gestures. The arrow indicates the strength of activation of the gesture, and the shading represents the scope of its effect, with darker shading indicating stronger activation. Figure adapted from Krivokapić (2020).

conditioned phonological processes taking place in post-stress position, in the sense that it provides a theoretical tool for the interpretation of spatio-temporal articulatory effects across a syllabic boundary on both spatial and temporal aspects.

The stress modulation is a general enhancement of the segmental phonetic realization. Speakers use different strategies to enhance the realization of segmentally contrasting features. A primary mechanism for enhancing distinctions is to decrease coarticulatory overlap so that gestures for segments in stressed syllables blend less with each other or with segments in neighboring syllables (de Jong et al., 1993). This is viewed as a stress-induced resistance to coarticulation as argued by Cho (2004): coarticulatory reduction in prosodically strong positions gives rise to heightened phonetic clarity, which can be interpreted as enhancing linguistic contrast. This view can be interpreted within the  $\mu$ -gesture proposal as follows: stress-induced resistance to coarticulation shall be the strongest within the  $\mu$ -conditioned syllable, where segments are enhanced by the  $\mu$ -modulation; this resistance is less strong - but present - during the anticipatory and the spill-over effects of the modulation, the segments under the activation scope of the  $\mu$ -gesture are enhanced but to a lesser degree. Cho (2004) further states that stress-induced coarticulatory variation cannot be attributed to a duration factor. This means that the  $\mu_T$ -conditioning may not be able to fully explain a resistance to coarticulation, the  $\mu_S$ -conditioning in both anticipatory and spill-over effects of the  $\mu$ -gesture shall act as a resistance to coarticulation in the spatial domain.

1.2. Stress-conditioning

Lexically prominent positions are known to be privileged. They manifest positional maintenance of contrasts which are otherwise neutralized, and resistance to processes which apply elsewhere (Beckman, 1998). For instance, vocalic contrasts are preferentially realized in stressed positions and may be reduced in unstressed positions (see Crosswhite (1999) for a

<sup>1</sup> The experimental design of Beckman and Edwards (1994) put the targeted syllables in different phrasal prominence contexts, thus the lexical stress effects measured in this study are post-lexical effects.

review; e.g., English (Fourakis, 1991), Brazilian Portuguese (Nobre and Ingemann, 1983), Western Catalan (Marcet et al., 2022)). This phonological pattern is phonetically grounded, since syllable nuclei bearing lexical prominence are longer in duration, higher in *f0*, more peripheral and more intense than their unstressed counterparts (Gordon and Roettger, 2017). Thus, the reduction in unstressed vowels represents the phonetic implementation of a vowel that lacks the acoustic correlates of stress (Flemming, 2005). A similar behavior is observed for consonantal contrasts, which are preferentially maintained in lexically prominent positions. For instance, in Copala-Triquet the contrast between voiced and voiceless consonants is realized in pre-tonic position but neutralized elsewhere (Hollenbach, 1977). Another example comes from Finnish verbs, where the contrast between alveolar stops and fricatives before [i] is neutralized through assibilation, except when the stop is immediately post-tonic (Anttila, 2003; Suomi and Ylitalo, 2004). Giavazzi (2010) extends the contrast-based account of Flemming (2005) to the consonantal domain, proposing a unified account of these processes: Just as for the preservation of vocalic contrasts under stress, the preservation of consonantal contrasts in the domain of stress is a phonetic consequence of lexical prominence. Processes such as voicing neutralization (in Copala-Triquet) or assibilation (in Finnish) apply outside of the stress domain due to the absence of the phonetic correlates of stress that would typically result in hyper-articulation and/or lengthening in these positions.

However, while acoustic correlates of stressed vowels have been extensively studied (Gordon and Roettger, 2017), acoustic properties of consonants in a lexically prominent position have received less attention. Consonants in these positions are the preferred target of a small class of frequent phonetic processes (González, 2003; Smith, 2002). For instance, consonants are often lengthened in pre-tonic and post-tonic position (as in English and Somali, respectively; Armstrong (1964); Turk and Sawusch (1997)). Also, consonants in these positions are described as having louder [sɪc] or affricate-like bursts, e.g., in Farsi, (Samareh, 1977) and Maori (Bauer, 2003). Giavazzi (2010) argues that, like the resistance to neutralization, these phonetic processes also result from the grammatical pressure to enhance the perceptual prominence of a metrically strong position.

1.3. Velar palatalization in Italian: A stress-conditioned phonological process

Velar palatalization is the process through which a velar stop, /k, g/, is fronted to a palatal or a palato-alveolar affricate (or fricative) (Kochetov, 2011). It applies more frequently before high and mid-high front vowels (e.g., /i, e/) than before other vowels. The palatalization process in modern Italian is the alternation of /k, g/ becoming [tʃ, dʒ], respectively, when they are directly followed by /i/.

In Romance languages, the origins of palatalization can be found in Late Latin: this process is known as the 2<sup>nd</sup> Romance palatalization, and it occurred after the 5<sup>th</sup> century CE (Rohlf et al., 1972; Recasens, 2020; Petrosino and Calabrese, 2021). It occurred before all front vowels, both root-internally and at the morpheme boundary, and independently of morpheme boundaries and stress position (Celata and Bertinotto, 2005). In contemporary Italian, palatalization of

velars takes place at the boundary between the root and inflectional (or derivational) suffixes in /-i/. In masculine nouns and adjectives, palatalization before the plural ending /-i/ is predominantly stress-conditioned. It occurs in the onset of the final syllable when stress occurs remotely in the antepenultimate position, while it is much rarer in the onset of the final syllable when stress occurs adjacently in the penultimate position (Giavazzi, 2008; Giavazzi, 2010), see Steddy (2015) and Flor (2023) for a discussion of palatalization in verbal domain, and more generally across the Italian lexicon.<sup>2</sup> In other words, when adding the plural suffix /-i/ to a trisyllabic word having /k/ as the onset of the final syllable, the palatalization of /k, g/ largely depends on its relative position to lexical stress. We refer to this phonological process as “phonological palatalization”, which is illustrated in (1) (different from “coarticulatory palatalization” presented in Section 1.4).

|     |       |        |             |                    |   |         |             |                    |
|-----|-------|--------|-------------|--------------------|---|---------|-------------|--------------------|
| (1) | FAR   | comico | [ko.mi.ko]  | “comedian m. sg.”  | → | comici  | [ko.mi.tʃi] | “comedians m. pl.” |
|     | CLOSE | caduco | [ka.ˈdu.ko] | “deciduous m. sg.” | → | caduchi | [ka.ˈdu.ki] | “deciduous m. pl.” |

The case of (1) FAR presents lexical stress far from the onset /k/ of the final syllable (i.e., there is another syllable in-between). In this case, adding the suffix /-i/ causes palatalization of the final onset /k/, and the underlying /k/ is realized as [tʃ] before [i]. This also happens when there are multiple syllables between lexical stress and final onset /k/. In (1) CLOSE, on the contrary, the final onset /k/ is close to the stressed syllable, i.e., directly post-tonic (no syllable occurs between the stressed syllable and the final onset /k/). Adding the suffix /-i/ does not cause palatalization, the final onset /k/ is realized as [k]. Thus in the CLOSE condition, the palatalization process is “blocked”.

The result of palatalization in the FAR context is that the contrast between /ki, gi/ and phonemic [tʃi, dʒi/ (as in *fradiciofradici*, [ˈfra.di.tʃo]-[ˈfra.di.tʃi], “soaked”) is neutralized. Conversely, the result of the blocking of palatalization in the CLOSE context, is that the contrast between /ki, gi/ and [tʃi, dʒi/ is maintained. This stress-conditioned distribution of a consonantal contrast thus makes velar palatalization in Italian comparable to the cases of assibilation in Finnish, of voicing neutralization in Copala Triquet, presented in Section 1.2.

It should be noted that the result of the palatalization of /ki/ differs across Italian varieties, with [tʃi] being the most frequent voiceless phonetic realization in Northern varieties, and [ʃi] the most frequent voiceless phonetic realization in Central and Southern varieties. This dialectal variability justifies our criteria of speakers recruitment (see Section 2).

1.4. Coarticulatory velar palatalization in articulation

The palatalization process has been studied in its acoustic and perceptual aspects (Guion, 1998; Kochetov, 1998; Kochetov, 2013). Ohala (1992) proposes that it is primarily perceptually-based. Guion (1998) reports that velar consonants before front vowels tend to be produced with greater frication at the release, which makes them both acoustically and perceptually similar to palato-alveolar consonants. The process also has articulatory bases (Kochetov, 2011): the articulatory motivation is seen as

<sup>2</sup> There are notable exceptions (e.g., *amico* – *amici*, [a.ˈmi.ko] – [a.ˈmi.tʃi], “friend” – “friends”).

the impossibility to co-produce a tongue body backing gesture for [k] and a tongue body fronting for [i]. Thus, a syllable such as [ki] can only be produced with an automatic coarticulatory adjustment. This gradient fronting of a [k] before [i] is considered as a “natural rule” (Hyman, 1975, p. 158). The velar palatalization, be it articulatorily-based or perceptually-based, is rooted in the coarticulation (and its acoustic output) between a velar plosive and a high vowel. The high vowel in nucleus position modifies acoustic and articulatory properties of preceding velar plosive in onset position, resulting in a acoustically more fricative-like and articulatorily fronted consonant. Henceforth, we refer to this “automatic” or “natural” palatalization as coarticulatory palatalization.

As presented earlier, the phonological palatalization in modern Italian is also stress-conditioned with post-tonic velar palatalization being “blocked”. This suggests that the lexical stress potentially modifies the coarticulation between the post-tonic velar plosive and the high vowel. The coarticulatory palatalization under stress-conditioning may result in a less palatalized realization of the velar plosive, compared to the same coarticulatory palatalization in a context without stress-conditioning. This less palatalized velar plosive, compared to more palatalized velar plosive without stress-conditioning, could be the perceptual and articulatory basis of the blocking of phonological palatalization.

The phonetic bases of stress-conditioned phonological palatalization have not received much attention (with the exception of Giavazzi (2010)); how lexical stress impacts post-tonic coarticulatory palatalization is largely unknown. Comparing the acoustic properties of velar stops in nonce words, such as, [ˈpi.ta.ki] (FAR) vs. [pi.ˈta.ki] (CLOSE), our recent studies (Shao et al., 2023b; Shao et al., 2023c) found that the release of [k, g] before [i] has a similar acoustic shape to [t̪, d̪] in terms of center of gravity. This confirms Guion (1998). We also found a significantly longer closure duration of [k] in the CLOSE context (see also Giavazzi (2010)). These acoustic properties of post-tonic plosives are likely caused by stress conditioning, which is the focus of the current study.

Within AP/TD, palatalization is analyzed in terms of gestural blending (Saltzman and Munhall, 1989; Browman and Goldstein, 1992; Kochetov, 1998; Bateman, 2007; Oh, 2022). Gestural blending occurs between two gestures on the same articulatory tier. In the case of palatalization, the targets of the process, /k, g/, and the trigger of the process, /i/, all recruit the tongue body. The consonantal and vocalic gestures inevitably overlap with each other, producing an outcome that differs from that of either individual gesture (Browman and Goldstein, 1989). This gestural blending is in turn under the impact of the  $\mu$ -gesture, which shall introduce a resistance to coarticulation (de Jong et al., 1993; Cho, 2004). That is to say, under the impact of the  $\mu$ -gesture, there shall be less gestural blending, similar to what has been reported in Fougeron and Keating (1997). Temporally, less gestural blending would result in longer velar consonantal gestures; spatially, it would result in less palatalized velar consonants.

### 1.5. Aims of this study

This study investigates stress-induced articulatory modulations on the coarticulatory palatalization. It focuses on spill-over effects of these modulations in the velar consonant

directly preceded by lexical stress. The study aims to describe how [k] and [g] differ in relation to lexical stress, in a context where both have a fronted tongue dorsum due to the coarticulatory palatalization triggered by the following [i]. In other words, we aim to investigate how the spill-over effects of lexical stress interact with the coarticulatory palatalization in the following unstressed syllable. To do so, we will focus on articulatory transitions between [V.C] (i.e., the vowel and the following velar plosive across a syllabic boundary), by comparing the FAR condition (e.g., [ˈpi.ta.ki]) to the CLOSE condition (e.g., [pi.ˈta.ki]). We aim to provide answers to the following questions:

1. How do stress-induced modulations affect stressed vowels and, more importantly, the [V.C] transition from the stressed vowel to the following consonant across a syllabic boundary?
2. What are the quantitative spill-over effects of lexical stress on both temporal and spatial domains?
3. How does stress-modulated coarticulatory palatalization relate to stress-conditioned phonological palatalization in modern Italian?

To answer these questions we present a direct investigation of the tongue dorsum movement transitioning from the stressed vowel to the following consonant. This is compared to the same vowel and consonant in an unstressed condition. Discussing our articulatory results in relation to acoustic findings presented in our work (Shao et al., 2023a; Shao et al., 2023b; Shao et al., 2023c) will allow us to better understand acoustic and perceptual consequences of these stress-induced modulations.

Considering phonological alternations described in Section 1.2 and Section 1.3, we hypothesize that the phonological processes targeting stress-adjacent consonants are phonetically grounded, and result from the grammatical pressure to enhance the perceptual prominence of the stress domain. The study presented here focuses on one such phonological process, i.e., velar palatalization in Italian, more concretely on its stress-conditioned “blocking”. The phonetic bases of this phonological process are the coarticulatory palatalization between a velar consonant and a high front vowel, and its “blocking” is argued to be rooted in the spatio-temporal modulation of the spill-over effects of lexical stress. That is to say, these spill-over effects could be the potential articulatory basis of the blocking of the phonological palatalization in post-tonic position in modern Italian.

More generally, this paper contributes to the study of the “prosodically driven phonetic details” (Cho et al., 2007) of Italian stress, and to the study of correlates of lexical stress more broadly. Further, it also shapes our understanding of the definition of the stress domain and suggests a paradigm which can be extended to similar phonological processes in other languages.

## 2. Method

### 2.1. Speakers

We recorded articulatory and acoustic data from 15 speakers from the northern regions of Italy based on their place of birth and primary school attendance. Seven were female and eight were male based on self-identification, the mean age

was 24.7 ( $\pm 4.3$ ) years. These speakers were born in and completed their primary education in their respective birth regions. See [Appendix A](#) for more information.

## 2.2. Stimuli construction and recording procedure

The structure of the target words was  $[C_1V_1.C_2V_2.C_3V_3]$ . The  $C_1$  and  $C_2$  positions were occupied by either /p/ or /t/. The  $V_2$  position was occupied by either /a/ or /e/. The nonce words were constructed to compare how the target consonants /k, g, tʃ, dʒ/ were produced in both FAR and CLOSE conditions, thus these four consonants were placed in  $C_3$  position. The  $V_3$  position was always occupied by the trigger of palatalization, i.e., the vowel /i/. Two examples are given in (2). The complete list of stimuli is presented in [Appendix A](#).

|     |       |         |             |
|-----|-------|---------|-------------|
| (2) | FAR   | pítachi | [ˈpi.ta.ki] |
|     | CLOSE | pitáchi | [pi.ˈta.ki] |

As indicated in (2), the target words were transcribed using standard Italian orthography. The stressed syllables were denoted by placing an accent mark on the vowel, following the common convention for marking stress in unfamiliar nouns in Italian. The two nonce words in (2) differ only by the stress position. In “pítachi” [ˈpi.ta.ki], the stress is on the first syllable,  $C_3$  [k] is thus FAR from stress. In “pitáchi” [pi.ˈta.ki], the stress is on the second syllable,  $C_3$  [k] is thus CLOSE to stress. During the recording session, the target words were presented on a computer screen to the speakers within a carrier phrase. The screen displayed a map of Italy with a nonce word (representing a village name), and below it was the carrier phrase “Pimpa parte da \_\_\_ la mattina presto” (English translation: “Pimpa leaves from \_\_\_ early in the morning.”). The speakers were told that the cartoon dog “Pimpa” was visiting small villages across Italy. The target words, representing the names of these villages, were displayed on the map in a randomized order. Since the nonce words were presented as proper nouns, there is no morphological boundary between  $C_3$  and  $V_3$ .

Since the nonce words were embedded within a carrier sentence, the stressed syllables of interest were not only stressed but also post-lexically accented in both the FAR and CLOSE conditions.<sup>3</sup> We consider that the effect of phrasal prominence was controlled for since it was kept constant across conditions, and only comparative measurements between the two conditions were carried out. Therefore, although it is impossible to isolate the embedded words from phrasal prominence, the observed effects could solely be explained as residual effects of lexical stress. We thus referred to the observed effects as “effects of lexical stress”.

The entire list of target words was repeated four times, which resulted in a total of 3840 utterances (64 targets  $\times$  4 repetitions  $\times$  15 speakers), among which 3635 utterances were usable.<sup>4</sup> In this study, we only analyze utterances containing the consonants [k, g] ( $n=1809$ ).

## 2.3. EMA data acquisition

Articulatory and acoustic data were collected simultaneously using the Electromagnetic Articulograph (EMA) AG501 (Carstens Medizinelektronik GmbH). Speakers were placed in the center of the electromagnetic field generated by EMA and they wore a head-mounted microphone for acoustic recording ([Fig. 2](#) left). EMA sensors were placed on upper and lower lips, tongue tip, tongue mid, and tongue dorsum, with additional sensors behind the left and right ear for head correction ([Fig. 2](#) right). The articulatory signal was recorded with a sample rate of 1250 Hz and filtered using a Butterworth lowpass filter with a cut-off frequency of 25 Hz and order 5 afterwards. The acoustic signal was recorded at a sample rate of 48 kHz and 32-bit float.

The utterances were pre-segmented using Audacity v3.1.3 ([Audacity Team, 2022](#)), then extracted and automatically annotated using the Montreal Forced Aligner v2.0.6 ([McAuliffe et al., 2017](#)) and manually inspected in Praat v6.3.10 ([Boersma and Weenink, 2023](#)). Gestural landmarks were automatically detected based on the acoustic annotations and using a custom script written in Python v3.10 ([Van Rossum and Drake, 2009](#)). Within a window specified by the acoustic segment boundaries of the targets [k, g] (with various temporal extensions for each speaker, the mean extension is  $\pm 400$  ms), the landmarks were detected using a 20% peak velocity threshold of the vertical tongue dorsum trajectory. The annotations and gestural landmarks were then manually inspected using ema2wav ([Buech et al., 2022b](#)), as shown in [Fig. 3](#).

## 2.4. Data extraction and statistical analysis

We have conducted the measurements and the statistical models to account for both spatial and temporal aspects of the stress effects, by systematically comparing the FAR condition to the CLOSE condition. The gestural landmarks and the acoustic annotations are combined to provide insights into the stress-induced temporal interaction between articulation and acoustics. The modeling of the tongue dorsum movement trajectories provides direct access to the spatio-temporal differences between FAR and CLOSE conditions. The tongue dorsum position analyses investigate the stress-induced spatial differences. These three analyses are therefore complementary.

### 2.4.1. Dynamic data extraction and generalized additive mixed modeling

A dynamic analysis on the tongue dorsum sensor is conducted without the specific constriction gestures of each consonant or vowel. As illustrated in [Fig. 3](#), the trajectories of the tongue dorsum sensor from the acoustic onset of  $V_2$  to the acoustic offset of  $V_3$  were extracted and modeled. These trajectories will show *how fast* and *how far* the tongue dorsum moves from  $V_2$  to  $C_3$  [i].

For this dynamic analysis, we compared the trajectories of the tongue dorsum movement from  $V_2$  to  $V_3$  in the high-low dimension. The trajectories were measured in 30 steps over the time-course of the movement from the acoustic onset of  $V_2$  to the acoustic offset of  $V_3$ . The EMA data was normalized for each speaker according to the highest y-position measured

<sup>3</sup> We thank an anonymous reviewer for bringing this point to our attention.

<sup>4</sup> 205 utterances (5.3%) were eliminated: 64 due to a detached sensor, 141 due to stuttering, hesitation, incorrect placement of stress, or error in production.

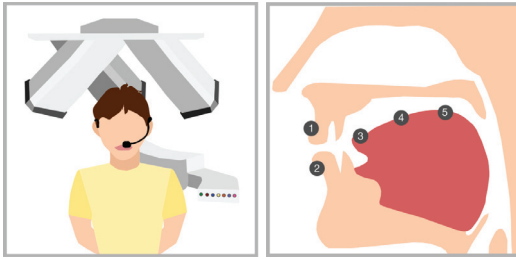


Fig. 2. (Left) Illustration of the Electromagnetic Articulograph (EMA) AG 501 with the speaker placed inside the electromagnetic field. The speaker wears a head-mounted microphone. (Right) Illustration of the positions of the sensors. Additional sensors behind the left and right ear are not shown here. This figure is a courtesy of Andrés Felipe Lara.

with the tongue dorsum sensor (Buech et al., 2022a). Generalized additive mixed models (GAMMs) were constructed using the `mgcv` package (1.8-40) in R (4.2.2) (Wood, 2011; Wieling et al., 2016; Wieling, 2018; R Core Team, 2023), and visualized using `Tidyverse` (1.3.2) and `Tidymv` (3.3.2) (Wickham et al., 2019; Coretta et al., 2023). The GAMMs were performed on each type of  $C_3$  with `Stress Position` ( $V_1$ ,  $V_2$ ) as fixed effect and smoothing parameters, with `by-Stress Position` factor smooths for `Speaker` as random effect. The `bam` function was used because of the large size of the data-set. The left-skewed distribution of the position data justified the use of scaled- $t$  family. The structure of the models is as follows:

```
bam(TD position ~ Stress Position + s(Time-
step, by = Stress Position) + s(Timestep,
Speaker, by = Stress Position, bs = "fs", m = 1),
data = data, scat(theta = NULL, link = "iden-
tity", min.df = 3))
```

This dynamic analysis was applied to show how the tongue dorsum moves through different locations and heights over relative time. The acoustic duration difference is thus normalized.

To complement and reinforce the dynamic analysis, the window method and corresponding statistical modeling is applied to the same data set, as presented below.

#### 2.4.2. Window method data extraction and parametric statistical modeling

To measure the absolute tongue dorsum position in both high-low (vertical) and front-back (horizontal) dimensions during the sequences  $[V_2.C_3]$ , we applied the window method (Roessig, 2021; Roessig et al., 2022), where a part of the acoustic segment is used as a time window to measure the averaged tongue dorsum position therein. The window method is a complementary analysis to the dynamic analysis, since it provides crucial and comparable information on *where* the tongue dorsum is located during the articulation of the vowels and consonants. For  $V_2$  [a, e], the mean position values of the tongue dorsum movement was calculated over the mid-50% to capture the representative tongue dorsum position in the vowel. For  $C_3$ , the mean position of a 10 ms window around the acoustic burst (i.e., 5 ms before the burst and 5 ms after it) of the consonant was used as an indicator to capture the constriction location of the consonant. The extracted mean position data were then standardized for each speaker utilizing z-scores. Note that in the horizontal domain, lower values indicate more fronted tongue positions; in the vertical domain, lower values indicate lowered tongue positions. A linear mixed effects model (`lme4`, 1.1-33) (Bates et al., 2015) was run for each  $V_2$  and  $C_3$  context using the same model structure as presented in Section 2.4.3.

#### 2.4.3. Gestural landmark and relating articulation to acoustics

The gestural target of the plosives [k, g] was achieved during their acoustic closure, which is shown in Fig. 3 and

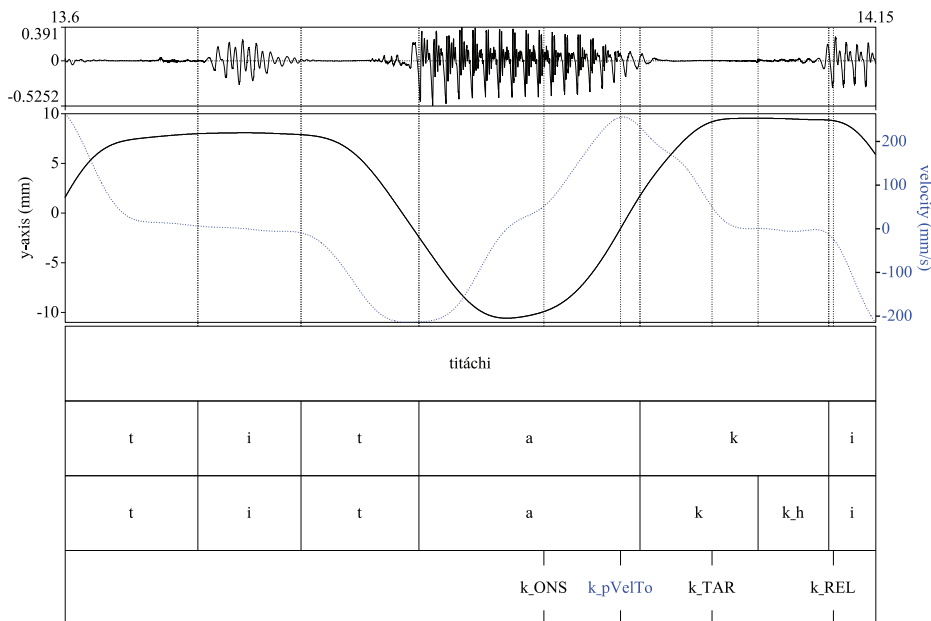


Fig. 3. An example from speaker 3 showing the target word [ti'takɪ], generated with `ema2wav` (Buech et al., 2022b). The upper panel shows the acoustic signal, the mid panel shows the movement of tongue dorsum sensor in the high-low (y) dimension accompanied with velocity (blue dotted line), and the lower panel shows the annotations with gestural landmarks. `k_ONS` refers to the gestural onset of [k], `k_pVelTo` refers to the peak velocity to achieve the gestural target of [k], `k_TAR` refers to the gestural target of [k], `k_REL` refers to the gestural release of [k]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

schematised in Fig. 4. The gestural target landmark thus divided the acoustic closure of the plosive into a first part from the acoustic offset of  $V_2$  (or, the acoustic onset of  $C_3$ ) to the gestural target, and a second part from the gestural target to the acoustic burst of the plosives. The first interval indicated to what extent the gestural target was delayed with respect to its acoustic onset, referred to as “Target delaying (ms)”. The second interval indicated how long it takes for the constriction gesture to be released acoustically, referred to as “Target to burst (ms)”.

These two time lag measurements were chosen to directly relate the acoustic closure duration to the articulation of plosives. In the  $[V_2.C_3]$  sequences, the full constriction of  $C_3$  [k, g] (indicated by the gestural target landmark) should be found after the vowel  $V_2$  and before the acoustic burst of the plosive  $C_3$ . Had this full articulatory constriction been delayed, the acoustic burst should also have been delayed, which would in turn result in a longer acoustic closure duration.

A linear mixed effects model (lme4, 1.1-33) (Bates et al., 2015) was run for each  $V_2$  and  $C_3$  context, in order to compare two stress conditions (i.e., FAR and CLOSE) in each  $V_2$ ,  $C_3$  context. In the model, we entered *Stress Position* as a fixed effect. As random effects, we had *by-Speaker* random slopes and intercepts for the effect of *Stress Position*. *P*-values were obtained by likelihood ratio tests of the full model with *Stress Position* against the model without the effect. The structure of the full models was as follows: `lmer(mean TD position ~ Stress Position + (1+Stress Position|Speaker), data = data, REML = F)`

We also conducted analyses of gestural landmark based mass-spring parameters, such as gestural duration (with acceleration and deceleration phases), displacement, and stiffness. They are modeled utilizing the same model structure as presented above, the results can be found in Appendix C.

#### 2.4.4. Reproducibility

To enhance reproducibility and offer comprehensive result documentation, both the data tables and the R markdown document have been made publicly available through the Open Science Framework (<https://osf.io/2we4z/>).

### 3. Results

The results are presented in the following order: we first present how the tongue movement unfolds over time with a dynamic analysis modeled with GAMMs in Section 3.1, with the aim of showing *how far* and *how fast* the tongue dorsum travels from the stressed/unstressed  $V_2$  to the following  $C_3V_3$ , which directly visualizes the target delaying pattern. In Section 3.2, we present the parametric analysis of positional data using the window method. This is to show the tongue dorsum’s absolute positional difference between FAR and CLOSE conditions in  $C_3$  and  $V_3$  without involving temporal aspects. In Section 3.3, we present a combined analysis of gestural landmarks and acoustic landmarks, showing a clear pattern of post-stress gestural target delaying and its correlation to acoustics.

#### 3.1. Dynamic analysis of the tongue dorsum trajectories of $[V_2.C_3]$

The GAMM estimates of the tongue dorsum movement on high-low (y) dimension are presented in Fig. 5. In the time-normalized data, the temporal relation between the acoustic onset of  $V_2$ , i.e., the 0 on the normalized time, and the subsequent movements are easily observable.

As presented in Fig. 5, panels [aki, agi], the tongue dorsum is lowered in [a] and raises to form the constriction of [ki, gi]. The stressed  $V_2$  has a much lower tongue dorsum than the unstressed  $V_2$ . In other words, the displacement from [a] to [ki, gi] is much larger when [a] is stressed (see also Table C.1). Furthermore, the targets of [ki, gi] are achieved at around time point 8 when [a] is unstressed, and at around time point 15 when [a] is stressed. This difference (i.e., the delayed articulatory target) is caused by two main factors. The first factor is the longer duration of the stressed vowel. The stressed vowel being more than two times as long (Shao et al., 2023b) inevitably delays all subsequent articulatory gestures. The second factor is the larger displacement of the stressed vowel. When the displacement is larger, the tongue has to raise from a lower point to achieve the targets of [ki, gi], which additionally contributes to the delaying of the targets. The two figures of estimated differences of [aki, agi] show that the tongue dorsum movements in [ki, gi] have the same high-low dimension values in both stress conditions, as can be inferred

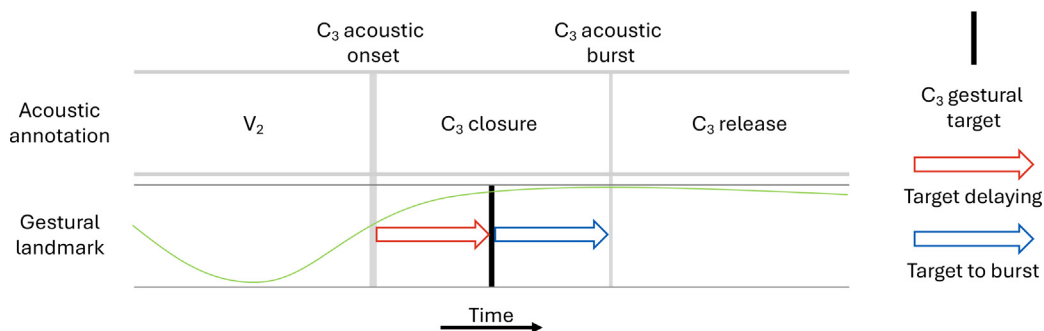
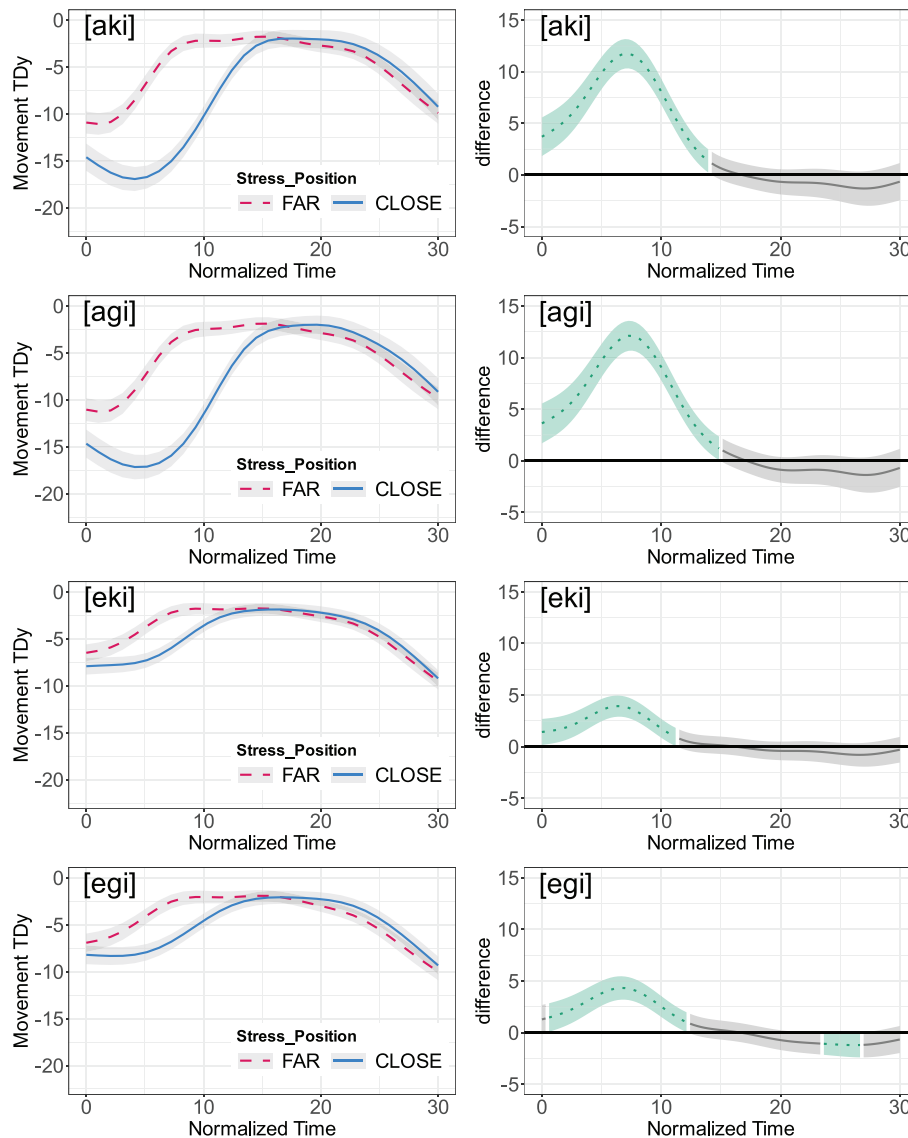


Fig. 4. Schematic representation of “Target delaying (ms)” and “Target to burst (ms)” intervals. The target delaying time lag is calculated as  $C_3$  gestural target relative to  $C_3$  acoustic onset; the target to burst time lag is calculated as the  $C_3$  acoustic burst time relative to  $C_3$  gestural target time.





**Fig. 5.** GAMM fitted tongue dorsum trajectories in  $[V_2C_3i]$  sequences on the high-low (y) dimension (in mm), by normalized time (left), with estimated differences (right). Time point 0 corresponds to the acoustic onset of  $V_2$ , and time point 30 corresponds to the acoustic offset of  $V_3$ . The y-dimension was normalized according to the highest position measured. Estimated differences are presented with 95% confidence intervals by time (normalized). The green dotted lines represent the significant difference between stress in FAR vs. stress in CLOSE contexts.

from the non-significant portions in the estimated differences, approximately from time point 14 and onward.

In Fig. 5, see panels [eki, egi], we observe a slightly different pattern when  $V_2$  is [e] compared to when  $V_2$  is [a]. The difference between the tongue dorsum height of [e] when the vowel is stressed or unstressed is smaller than for [a].<sup>5</sup> In other words, the stressed [e] does not have a larger displacement in tongue dorsum, unlike the stressed [a] (see also Table C.1). However, stressed [e] is still much longer than unstressed [e], the target achievement of the following [k, g] is still delayed in

the same way as in [aki, agi]. From the two figures of estimated differences of [eki, egi], we see that the tongue dorsum movements of [ki, gi] do not have a significant difference once they reached their highest points in both stress conditions, as can be inferred from the non-significant portions in the estimated differences, approximately from time point 12 onward.<sup>6</sup>

In sum, the most important systematic pattern observed in tongue dorsum trajectory is the delayed target achievement of  $[C_3i]$  in the CLOSE condition. The stressed  $V_2$  [a] induces more delay, and the stressed  $V_2$  [e] induces less delay. This result corresponds to Beckman and Edwards (1994)'s obser-

<sup>5</sup> Some speakers produced a mid-open vowel [e] when stressed. There is a neutralization of the phonemic contrast between [e] and [ɛ] in Northern Italian, in unstressed syllables the mid-vowel is always mid-close. In Milanese the two front mid vowels are in complementary distribution, with the mid-open one in closed and the mid-close one in open stressed syllables (Krämer, 2009). Since the variability in terms of vowel dorsum height (see Section 3.2) is very small and there is no observable tongue dorsum height pattern based on the mid-front vowel quality, we decided to not report the mid-front vowels separately.

<sup>6</sup> We note that there is a small significant difference in [egi], however this significant difference is very small, with the largest difference estimated at around 1 mm. This difference is just around the standard deviation of the prescribed accuracy of the EMA machine AG501 used in this study (Stella et al., 2013; Savariaux et al., 2017), thus we will not further explore it in our analysis.

vation of lexical stress, which includes a more open vocal tract during the stressed vowel, and hence larger tongue dorsum displacement in moving away from the stressed vowel and in going into the following consonant.

### 3.2. Parametric analysis of the averaged absolute tongue dorsum positions of $[V_2]$ and $[C_3]$

In this section, we analyse the tongue dorsum positions averaged across a certain time window. This spatial difference could be crucial, because it gives information about the tongue dorsum position during  $V_2$  and the place of articulation of  $C_3$ . We conducted a parametric analysis with the window method (Roessig, 2021; Roessig et al., 2022) and with mixed effect models constructed for  $V_2$  and  $C_3$  ([i]) separately.

We first report the averaged absolute tongue dorsum position over the middle 50% of  $V_2$ . As shown in Fig. 6 and Table 1, there is a difference between stressed vowels and unstressed vowels in terms of tongue dorsum position. However, the two vowels [a, e] behave differently.

The vowel [a] shows differences in both high-low and front-back dimensions in relation to stress: on the high-low dimension, the tongue dorsum of stressed [a] is much lower ([aki]:  $p < 0.001$ , [agi]:  $p < 0.001$ ); in the front-back dimension, it is fronted ([aki]:  $p < 0.001$ , [agi]:  $p < 0.001$ ).

Unlike [a], [e] has a lowered-retracted tongue dorsum when it is stressed. On the high-low dimension, the tongue dorsum of stressed [e] is lower ([eki]:  $p < 0.001$ , [egi]:  $p < 0.001$ ); on the front-back dimension it is slightly retracted ([eki]:  $p < 0.05$ , [egi]:  $p < 0.01$ ). The lowering could be related to the opening of the stressed vowel, but the more retracted tongue dorsum position is more difficult to interpret. A possible explanation is that it is not related to the articulatory profile of the stressed vowel *per se*, but rather related to the fronted unstressed [e]. In our stimuli, the vowel in the first and the third syllable is always [i]. When the second syllable  $[C_2e]$  is unstressed, the stressed [i] in the first syllable may lead to a slightly fronted and higher tongue position during the following unstressed [e], as can be observed in Fig. 6. From a different angle, it may also be plausible to assume that the stressed [e], being hyperarticulated and surrounded by two [i], uses the lower-retracted tongue dorsum position as a stress-related dissimilation strategy.

Fig. 7 and Table 2 present the tongue dorsum position obtained around the acoustic burst of  $C_3$  [k, g] and the corresponding statistical modeling results, which indicate that a stressed vowel induces a more retracted tongue dorsum position in the following velar plosive. We see that on the x-axis, the tongue dorsum is more retracted when the preceding vowel is stressed ([aki]:  $p < 0.01$ , [eki]:  $p < 0.001$ , [egi]:  $p < 0.001$ ), except for [agi] ( $p = 0.06$ ). This indicates that the constriction location at the point of the acoustic release is affected by stress: in the *close* condition, the velar plosives have a more retracted place of articulation, and in the *far* condition, the place of articulation is more fronted. Even in the case of [agi] where this difference is not statistically significant, we still see a retracted tongue dorsum position in Fig. 7. The overall pattern is consistent across conditions.

Another explanation to this difference may come from the stress-induced articulation of  $V_3$ . Similar to  $V_2$  [e] in *far* condition, where the [e] is articulated with a fronted tongue body pre-

sumably due to the stressed [i] in  $V_1$  position. This may have also happened from the stressed  $V_2$  to the unstressed  $V_3$ , which would have resulted in a fronted (for [a]) or retracted (for [e]) tongue dorsum in  $V_3$  and in turn a fronted or retracted constriction location in *close*  $C_3$ . This is however not the case as shown in appendix D, where the window method is applied to  $V_3$  and does not reveal a consistent pattern. In *close* condition, the  $V_3$  have a lowered tongue dorsum, but there is no consistent fronted or retracted tongue dorsum similar to what is reported here. The retracted tongue dorsum in  $C_3$  can thus not be attributed to the articulation of  $V_3$ .

By comparing Fig. 6 to Fig. 7, we can infer that the two  $V_2$  [a, e] have different degrees of impact on  $C_3$ . Stressed [e], by having a more retracted tongue dorsum itself, induces more tongue dorsum retraction in the following consonant ([eki]:  $\beta = 0.40$ ,  $p < 0.001$ , [egi]:  $\beta = 0.36$ ,  $p < 0.001$ ). Stressed [a] induces less tongue retraction in the following consonant ([aki]:  $\beta = 0.27$ ,  $p < 0.01$ , [agi]:  $\beta = 0.24$ ,  $p = 0.06$ ), but considering that stressed [a] is articulated with a fronted tongue body, this retraction observed in the following consonant is a strong evidence of a stress-related articulatory pattern, and not a vowel-quality related articulatory pattern.

In sum, the tongue dorsum position showed that stressed vowels led to a further back tongue dorsum position independently of the vowel quality, and the impact is more important in the  $[e.C_3]$  sequences. These results indicate that stress has a significant impact on the place of articulation of the following consonants.

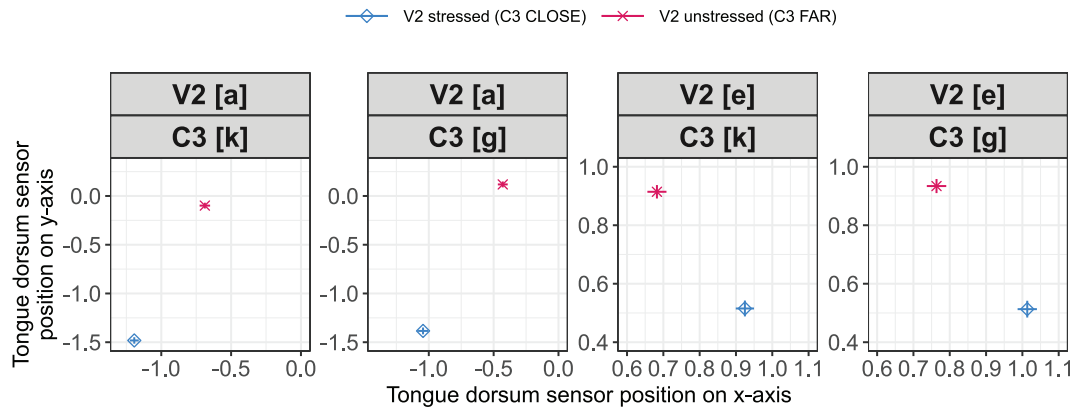
### 3.3. Articulatory target delaying in relation to acoustics

This section investigates the articulatory basis of the acoustic pattern that velar plosives [k, g] have a longer closure duration in the *close* condition (see Shao et al. (2023b) and its replication for the current dataset in Appendix B). This is done by directly measuring acoustic and articulatory landmarks.

In our data, 91% of  $C_3$  (1646 tokens out of 1809 tokens in total) were correctly labeled with gestural landmarks by the Python script presented in Section 2.4.3. Among which 83% of the tokens (1369 tokens out of 1646 tokens in total) had their gestural target located within the acoustic closure of the  $C_3$  (Svensson Lundmark, 2023).<sup>7</sup> The two subdivisions of the closure duration of a velar plosive - target delaying time lag and target to burst time lag - were calculated based on gestural landmarks and acoustic annotations as illustrated in Fig. 4.

Statistical modeling results of the two time lags are presented in Table 3. Clearly, the target delaying time lag shows a stress-conditioned pattern: in the *close* condition, the target delaying time lag is longer than in the *far* condition for all vowels and consonants ([aki]:  $p < 0.01$ , [agi]:  $p < 0.001$ , [eki]:  $p < 0.01$ , [egi]:  $p < 0.05$ ). However, the target to burst time lag does not show a clear stress-conditioned pattern ([aki]:  $p = 0.25$ , [agi]:  $p = 0.79$ , [eki]:  $p = 0.01$ , [egi]:  $p < 0.001$ ). This result may be directly related to the longer closure duration observed in velar consonants in the *close* condition: the longer closure duration in the *close* condition could be caused by a delayed articulatory target, which in turn delayed the acoustic burst of the consonants.

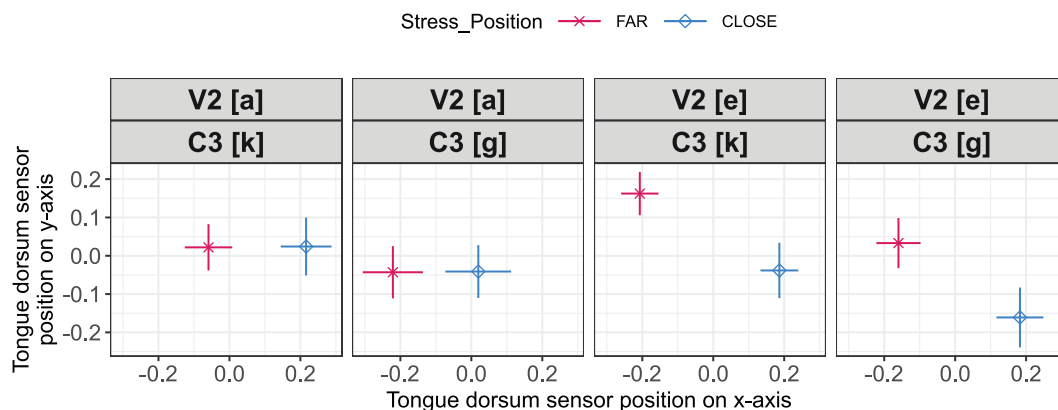
<sup>7</sup> The rest of the data, which is 440 tokens, is not included in the analysis.



**Fig. 6.** Z-scored mean tongue dorsum positions in V<sub>2</sub>, obtained over the mid-50% of the acoustic duration using the window method, presented with standard errors. The red crosses and error bars correspond to when the stress is on the first syllable [C<sub>1</sub>V<sub>1</sub>] (i.e., when [k, g] consonants are in the FAR condition); the blue diamonds and error bars correspond to when the stress is on the second syllable [C<sub>2</sub>V<sub>2</sub>] (i.e., when [k, g] consonants are in the CLOSE condition). The y-axis represents the high-low dimension and the x-axis represents the front-back (left to right) dimension. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**  
Parametric analysis of z-scored mean positions of tongue dorsum sensor during V<sub>2</sub> in different stress conditions. The mean position values are obtained over the mid-50% of the acoustic duration of V<sub>2</sub>. The likelihood ratio tests of linear mixed effects models test the main effect of the predictor STRESS (C<sub>3</sub> at FAR (baseline) and CLOSE positions respectively).  $\chi^2$  values (with one degree of freedom) are reported as effect size estimates.  $\beta$ -coefficients, standard errors (SE) and  $p$ -values of the models are also reported.

|          | Dimension | $\chi^2$ | $\beta$ | SE   | $p$    |
|----------|-----------|----------|---------|------|--------|
| in [aki] | y         | 50.85    | -1.39   | 0.08 | <0.001 |
|          | x         | 11.70    | -0.51   | 0.05 | <0.001 |
| in [agi] | y         | 63.43    | -1.50   | 0.05 | <0.001 |
|          | x         | 18.90    | -0.63   | 0.10 | <0.001 |
| in [eki] | y         | 11.90    | -0.40   | 0.10 | <0.001 |
|          | x         | 5.44     | 0.24    | 0.10 | <0.05  |
| in [egi] | y         | 12.70    | -0.43   | 0.10 | <0.001 |
|          | x         | 7.40     | 0.24    | 0.08 | <0.01  |



**Fig. 7.** Z-scored mean tongue dorsum positions in C<sub>3</sub>, obtained over a 10 ms window around the acoustic burst of the consonants, presented with standard errors. The red crosses and error bars correspond to when the stress is on the first syllable [C<sub>1</sub>V<sub>1</sub>] (i.e., when [k, g] consonants are in the FAR condition); the blue diamonds and error bars correspond to when the stress is on the second syllable [C<sub>2</sub>V<sub>2</sub>] (i.e., when [k, g] consonants are in the CLOSE condition). The y-axis represents the high-low dimension and the x-axis represents the front-back (left to right) dimension. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

One noticeable result is the target to burst time lag in [egi], which is significantly longer by about 3 ms in the CLOSE condition ([egi]:  $\beta = 5.65$ ,  $p < 0.001$ ). This may be caused by the statistical modeling, which is constructed with only random intercepts for speakers and does not contain random slopes due to convergence issue, certain vari-

ations among speakers may not have been appropriately addressed.

In sum, these two time lag measurements establish a plausible link between the articulatory gestural target and the acoustic closure duration. The gestural target in the CLOSE condition is delayed by the preceding stressed vowel.

**Table 2**

Parametric analysis of z-scored mean positions of tongue dorsum sensor during the burst of C<sub>3</sub> in different stress conditions. The mean position values are obtained over a 10 ms window around the acoustic burst of C<sub>3</sub>. The likelihood ratio tests of linear mixed effects models test the main effect of the predictor STRESS (C<sub>3</sub> at FAR (baseline) and CLOSE positions respectively).  $\chi^2$  values (with one degree of freedom) are reported as effect size estimates.  $\beta$ -coefficients, standard errors (SE) and  $p$ -values of the models are also reported.

|          | Dimension | $\chi^2$ | $\beta$ | SE    | $p$    |
|----------|-----------|----------|---------|-------|--------|
| in [aki] | y         | 0        | -0.0004 | -0.09 | =0.99  |
|          | x         | 8.54     | 0.27    | 0.09  | <0.01  |
| in [agi] | y         | 0.004    | 0.006   | 0.10  | =0.95  |
|          | x         | 3.68     | 0.24    | 0.12  | =0.06  |
| in [eki] | y         | 4.98     | -0.20   | 0.09  | <0.05  |
|          | x         | 28.72    | 0.40    | 0.07  | <0.001 |
| in [egi] | y         | 3.79     | -0.20   | 0.10  | =0.05  |
|          | x         | 16.11    | 0.36    | 0.09  | <0.001 |

**Table 3**

Parametric analysis with positive values of the articulatory target delaying of C<sub>3</sub> with respect to the acoustic onset of C<sub>3</sub> in different stress conditions. The delaying is measured as the time lag (ms) between the acoustic offset of V<sub>2</sub> to the articulatory target of C<sub>3</sub>. The time lag between the articulatory target to the acoustic burst of C<sub>3</sub> is also measured. The likelihood ratio tests of linear mixed effects models test the main effect of the predictor STRESS (C<sub>3</sub> at FAR (baseline) and CLOSE positions respectively).  $\chi^2$  values (with one degree of freedom) are reported as effect size estimates.  $\beta$ -coefficients, standard errors (SE) and  $p$ -values of the models are also reported. The model of target to burst duration in [egi] context contains only random intercept due to convergence issue and is marked with #.

|                      |           | Mean (sd)     |               | $\chi^2$ | $\beta$ | SE   | $p$    |
|----------------------|-----------|---------------|---------------|----------|---------|------|--------|
|                      |           | FAR           | CLOSE         |          |         |      |        |
| Target delaying (ms) | in [aki]  | 43.95 (17.33) | 50.62 (18.27) | 8.51     | 10.74   | 3.21 | <0.01  |
|                      | in [agi]# | 32.61 (14.23) | 40.04 (13.67) | 35.98    | 7.20    | 1.17 | <0.001 |
|                      | in [eki]  | 30.11 (20.90) | 34.09 (21.85) | 9.97     | 5.49    | 1.52 | <0.01  |
|                      | in [egi]  | 23.61 (16.82) | 27.37 (17.26) | 4.80     | 2.69    | 1.22 | <0.05  |
| Target to burst (ms) | in [aki]  | 18.95 (24.39) | 20.80 (18.40) | 1.30     | -3.51   | 3.03 | =0.25  |
|                      | in [agi]  | 24.66 (18.37) | 25.61 (16.44) | 0.07     | 0.47    | 1.77 | =0.79  |
|                      | in [eki]  | 31.73 (28.82) | 37.10 (29.50) | 2.35     | 3.74    | 2.30 | =0.13  |
|                      | in [egi]# | 33.39 (23.21) | 36.66 (22.18) | 12.38    | 5.65    | 1.59 | <0.001 |

**4. Discussion**

In this study, we investigated the relationship between lexical stress and velar palatalization in Italian based on dynamic and parametric analyses of articulatory data. We aimed to understand whether and how the realization of lexical stress impacts an adjacent velar consonant in a context of coarticulatory palatalization (i.e., followed by [i]). Our findings show that lexical stress affects the articulation of not only the stressed vowel, but also of the following consonants [k, g]. These effects have both temporal and spatial components. Temporally, the articulatory target of the velar consonants is delayed when they are directly preceded by a stressed vowel, which leads to a longer acoustic closure duration. Spatially, the place of articulation of the velar consonants is less fronted by coarticulatory palatalization when they are directly preceded by a stressed vowel.

In the following sections, we will discuss our findings in two directions: the spatio-temporal spill-over effects of stress (Section 4.1); then, we will discuss how these spill-over effects may condition the application of the phonological palatalization in Italian (Section 4.2); we then tentatively extend our discussion into a general pattern of stress-conditioning (Section 4.3).

*4.1. The spill-over effects of lexical stress*

Lexical stress is argued to strengthen its surrounding sounds, especially post-tonic consonants. Several languages are found to have a lengthening of consonants in post-tonic position (e.g., Gualavía Zapotec, Popoloca, Somali;

Armstrong (1964); Jones and Knudson (1977); Veerman-Leichsenring (1984)). Dimitrova and Turk (2012) showed that in English accent-induced lengthening is not only found on the primary-stressed syllable, but it also spills over onto a following unstressed syllable. Katsika and Tsai (2021) reported on a similar phenomenon in Greek, from an articulatory perspective. Consonantal gestures had their largest displacement on the syllable that immediately followed the stressed one, suggesting that the spill-over effect of the stress domain is larger than the anticipatory effect. In our previous study (Shao et al., 2023b) and in this one, we found that post-tonic velar consonants [k, g] are strengthened, both in the acoustic and in the articulatory domains. In the acoustic domain, closure durations of post-tonic plosives is longer, and their acoustic burst (for [k] only) is higher in intensity. In the articulatory domain, post-tonic articulatory gestures have a temporally delayed target (which is responsible for the lengthened closure duration) and are spatially less palatalized.

Our articulatory data provides new evidence in line with the literature discussed above, that the stress domain stretches beyond the syllable and strengthens the following consonantal gesture in both time and space. Articulatory Phonology allows to model these results, since it regards lexical stress as associated with an abstract spatio-temporal modulation gesture, the  $\mu$ -gesture (Saltzman et al., 2008), which has its own activation scope stretching beyond the stressed syllable. The temporal  $\mu_T$ -gesture modulates the stiffness of a syllable-planning oscillator and slows the gestural activation of co-active gestures (Saltzman et al., 2008; Krivokapić, 2020; Iskarous and Pouplier, 2022). This prediction is clearly confirmed in our data:

the delayed target of post-tonic consonants is consistently observed, as well as a significantly lower stiffness (see Appendix C) in post-tonic consonants' formation gesture. That is to say, the spill-over effect is observed to modulate both the target and the stiffness of post-tonic consonants, although it is possible that all dynamical parameters are interactively influential on kinematic realizations with different degrees of effect (Cho, 2006).

Moreover, the  $\mu$ -gesture proposal argues that spatial ( $\mu_S$ -gesture) and temporal ( $\mu_T$ -gesture) modulations can be dissociated. Evidence for the separability of these two components comes on the one hand from the cross-linguistic typology of stress correlates (Gordon and Roettger, 2017). Lexical prominence may be realized by durational enhancement only (e.g., in Finnish; Sluijter and Van Heuven (1996)), or as a combination of durational and sonority enhancement (e.g., in Italian; Farnetani and Kori (1986)). On the other hand, evidence for the separability of durational and spatial components of lexical prominence comes from vowel-specific effects. The realization of lexical prominence on different vowels is observed to have different articulatory correlates based on the vowels' phonological nature (Erickson, 2002).

We can therefore discuss the spatial articulatory differences between the two stressed vowels [a, e] in our data and how they may affect the following consonants differently. Spatially, stressed [a] has a lowered and slightly fronted tongue dorsum; stressed [e] has a slightly lowered and retracted tongue dorsum. These behaviors are in line with the general findings of lexical prominence (Engstrand, 1988; de Jong, 1995; Harrington et al., 2000; Erickson, 2002). The spill-over effect of lexical prominence modulates the spatial characteristics of post-tonic consonants: both stressed vowels introduce a less palatalized place of articulation. Since the stressed [e] has a slightly retracted tongue dorsum, it introduces a stronger modulation in the spatial domain in this particular case.

Cross-linguistically, our results on stressed [a] confirm earlier findings (Erickson, 2002; Katsika and Tsai, 2021). But our results on [e] point to a different direction compared to Erickson (2002), in which the emphasized mid-front vowels of English were shown to have a fronted tongue dorsum. This language-specific behavior may reflect a difference in the vocalic system of the two languages. In English, unstressed vowels are reduced (Lindblom, 1963; Crosswhite, 1999), in the sense that they are centralized. Their stressed counterparts are thus realized with a fronted tongue dorsum. In Italian, unstressed vowels are characterized much less by reduction compared to English (Burzio, 2007), and much more by their shorter duration (Bertinetto, 1980; Albano Leoni et al., 1995; Eriksson et al., 2016). This difference may explain why stressed [e] is not fronted in Italian.

Our results also contribute to the phonological discussion of lexical stress. The spill-over effects of lexical stress manifest in the articulation of the following consonants, across a syllabic boundary, since they are taking up the onset position. Most phonological theories hold that the syllable is the stress-bearing unit (Hayes, 1995), with the rhyme as its rhythmical sub-unit. Research on stress-related phonological processes has also adopted the syllable as the stress domain (see Beckman (1998), Smith (2002), Bye and de Lacy (2008),

amongst others). Consonantal stress-conditioned processes such as the one discussed here however occur stress-adjacently, independently of syllabic boundaries (Giavazzi, 2010). Our present findings clearly show that the spill-over effects of stress go beyond the syllabic boundary in both time and space. These results thus suggest that the domain of stress may not be equivalent with the syllable, as it seems to extend from the onset of the stressed vowel to at least the onset of the post-tonic syllable (Turk and Sawusch, 1997; Ryan, 2019; Steriade, 2011; Steriade, 2023). Future research has to further investigate the exact time-course of the modulations, both in their formation and in their release phases.

#### 4.2. Stress-induced modulations on coarticulatory palatalization and its potential influence on the distribution of phonological palatalization in Italian

Velar palatalization process is typically considered as a "natural" phonological process and has clear phonetic bases, which are both articulatory and acoustic-perceptual. From an articulatory point of view, palatalization is grounded in coarticulatory adjustments, i.e., fronting, of a velar stop which is followed by a front vowel (Ohalá, 1992; Kochetov, 2011). This fronting has perceptual consequences, as it increases the perceptual similarity of a [ki] and [gi] with, respectively, [tʃi] and [dʒi] (Guion, 1998). In turn, this affects the phonological grammar, giving rise to the frequent neutralization of these contrasts cross-linguistically before front vowels, through velar palatalization (Guion, 1998; Wilson, 2006).

Our study shows that the spill-over effect of lexical stress on the articulation of an adjacent velar consonant leads to a decrease of the coarticulatory overlap between the velar consonants and the vowel [i] (Cho, 2004). That is to say, the gestures of the plosives and the high-front vowel are more blended in the FAR condition and less blended in the CLOSE condition. This resistance to coarticulation (or, resistance to blending) has two outcomes: one is more related to the perceptual recovery of the plosive category, i.e., the longer closure duration due to the delayed target (see Table B.1 and Shao et al. (2023b)); the other is more related to the place of articulation of the plosives, i.e., the less palatalized place of articulation for post-tonic [k, g].

We can thus interpret this resistance as an enhancement, which enhances the phonological contrast (together with other acoustic enhancements, such as a louder burst observed in post-tonic [k]) between [ki] and [tʃi], [gi] and [dʒi], and may result in the maintenance of these contrasts (Flemming, 2005; Giavazzi, 2010). Building on these results, we argue here that the stress-conditioned distribution of phonological palatalization in Italian could be grounded in the articulatory modulations affecting coarticulatory palatalization in post-tonic contexts. As this context is modified by the spill-over effect of lexical stress on the following consonant, both the perceptual and articulatory bases of phonological palatalization are modified: the "natural" effect of a following [i] on the velar stop is counter-balanced by the effect of stress on its closure duration and place of articulation. The perceptual distance with respect to the corresponding palato-alveolar affricates is therefore large enough to be maintained in this prosodic context (Flemming, 2005; Giavazzi, 2010). This is what we refer to

as the phonetic bases of the “blocking” of phonological palatalization. The “blocking” of phonological palatalization in post-tonic position may have arisen from a combined effect of both perception and articulation rooted in the stress-induced spill-over modulations: it is arguably a consequence of the articulatory correlates of lexical prominence. This spatio-temporal spill-over effect of lexical stress therefore may have a phonological consequence: the preservation of a phonological contrast through the “blocking” of neutralization.

We note here that, in our interpretation, the perceptual recovery of the plosive category is in part based on the longer closure duration due to the delayed target in *CLOSE* condition. This longer closure duration caused by preceding lexical stress is in turn an acoustic detail which is used as a perceptual cue. When the closure duration is longer, the perceptual salience of the release of [k] is decreased, favoring the perceptual recovery of a plosive category rather than an affricate category. This hypothesis is currently being investigated with a perceptual study, which specifically questions the perceptual distinctiveness of the [ki]-[tʃi] and [gi]-[dʒi] in *FAR* and *CLOSE* contexts (preliminary data in [Giavazzi \(2010\)](#) and [Mitchell \(2022\)](#)).

#### 4.3. General discussion on stress conditioning

There are two main cross-linguistic strategies to implement metrical prominence in a stress domain: increasing the perceptual salience of the stressed vowel (i.e., by increasing its intensity or by hyperarticulating it), or increasing the duration of the stress domain (i.e., by lengthening some or all of its segments). Languages differ in the exact combination of these strategies they use ([Giavazzi, 2010](#); [Gordon and Roettger, 2017](#)). In some languages, e.g., Italian, ([Famietani and Kori, 1986](#); [Payne, 2005](#)), segments within a stress domain have a greater duration than segments in a stress-less domain, and stressed vowels have a greater intensity and are hyperarticulated compared to stress-less vowels. In other languages, e.g., Finnish ([Suomi and Ylitalo, 2004](#)), metrical prominence is exclusively expressed by the increased duration of segments within the stress domain. As we discussed earlier, these two cross-linguistic behaviors could be interpreted within the  $\mu$ -gesture proposal, as arising from differences in the activation strength of the  $\mu_T$  and the  $\mu_S$ -gestures. What follows is the prediction, put forth by [Giavazzi \(2010\)](#) that in a given language, the stress-conditioned processes of the sort discussed above will strictly depend on the precise correlates of lexical prominence (or, the language-specific activation strength of  $\mu_T$  and  $\mu_S$ -gestures) in that language.

In languages of the first group, in which the  $\mu_S$ -gesture is active, the articulation of the vowel is modulated spatially, yielding hyperarticulation. This modulation also affects adjacent segments ([Beckman et al., 1992](#); [de Jong, 1995](#); [Harrington et al., 2000](#); [de Jong, 2004](#); [Giavazzi, 2010](#)).

Our findings allow us to predict that other stress-conditioned processes may be similarly grounded in specific articulatory effects of lexical stress in that language. For instance, in languages of the second group, e.g., Finnish ([Anttila, 2003](#); [Suomi and Ylitalo, 2004](#)), only the  $\mu_T$ -gesture should be active. In this language, the alveolar plosives and fricatives follow the same stress-conditioned distribution as in Italian: a *FAR* from stress plosive undergoes an assibilation process, whereas a *CLOSE* to stress plosive does not. We predict that the longer closure duration of post-tonic plosives in Finnish ([Suomi and Ylitalo, 2004](#); [Giavazzi, 2010](#)) is a result of a temporal stress-induced articulatory modulation (i.e., a  $\mu_T$ -gesture modulation), but not of a spatial one (i.e., no effect on the place of articulation). This temporal modulation may contribute to the perceptual recovery of a plosive sound, leading to the “blocking” of assibilation in post-tonic position.

## 5. Conclusion

We find that lexical stress in Italian induces both temporal and spatial modulations which target the stressed vowel and affect the post-tonic consonant. Temporally, the post-tonic consonants have a delayed articulatory target compared to far from stress ones, which in turn delays the acoustic burst of the consonants. The acoustic result of target delaying is the lengthening of the closure duration in post-tonic consonants. Spatially, the post-tonic consonants are less palatalized compared to far from stress ones. These two effects can be seen as a stress-induced resistance to coarticulatory palatalization, which in turn may be the phonetic underpinnings of the “blocking” of the phonological palatalization in post-tonic condition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

We would like to thank the Italian speakers who participated in the EMA experiment and Sejin Oh for assistance with data collection. Our thanks also go to Sharon Peperkamp, Cécile Fougeron, Ioana Chitoran, Doris Mücke, and Louis Goldstein for their valuable input, suggestions, and discussions. We also extend our gratitude to the three anonymous reviewers and the editor Professor Taehong Cho, for their significant contributions.

## Appendix A. Speakers and speech materials

This appendix contains the information on speakers and the list of nonce word.

**Table A.1**  
Information on speakers.

| N° | Sex | Age              | Birth region        | Second languages |
|----|-----|------------------|---------------------|------------------|
| 1  | F   | 27               | Emilia-Romagna      | en, fr           |
| 2  | F   | 30               | Veneto              | en, fr, es       |
| 3  | M   | 24               | Veneto              | en, fr           |
| 4  | F   | 26               | Piemonte            | en, fr, es       |
| 5  | M   | 24               | Veneto              | en, fr           |
| 6  | M   | 21               | Valle d'Aosta       | en, fr           |
| 7  | M   | 27               | Piemonte            | en, fr           |
| 8  | F   | 22               | Lombardia           | en, fr           |
| 9  | F   | 37               | Lombardia           | en, fr           |
| 10 | M   | 24               | Emilia-Romagna      | en, es           |
| 11 | F   | 22               | Piemonte            | en, fr           |
| 12 | M   | 22               | Emilia-Romagna      | en, fr           |
| 13 | F   | 20               | Trentino-Alto Adige | pt, en, fr       |
| 14 | M   | 23               | Piemonte            | en, fr           |
| 15 | M   | 22               | Piemonte            | en, fr, de, es   |
| –  | –   | $\bar{x} = 24.7$ | –                   | –                |

**Table A.2**  
List of stimuli in orthographic form.

| CiCaCi  | CiCáCi  | CiCeCi  | CiCéCi   |
|---------|---------|---------|----------|
| pítaci  | pitáci  | píteci  | pitéci   |
| títaci  | titáci  | títeci  | titéci   |
| pítagi  | pitági  | pítegi  | pitégi   |
| títagi  | titági  | títegi  | titégi   |
| pítachi | pitáchi | pítechi | pitéchi  |
| títachi | titáchi | títechi | titéchi  |
| pítaghi | pitághi | píteghi | pitéghi  |
| títaghi | titághi | títeghi | titéghi  |
| pípaci  | pipáci  | pipeci  | pipéci   |
| típaci  | tipáci  | típeci  | tipéci   |
| pípagi  | pipági  | pípegi  | pipégi   |
| típagi  | tipági  | típegi  | tipégi   |
| pípachi | pipáchi | pipechi | pipéchi  |
| típachi | tipáchi | típechi | tipéchi  |
| pípaghi | pipághi | pípeghi | pipégghi |
| típaghi | tipághi | típeghi | tipégghi |

## Appendix B. Additional results on acoustic analysis

This appendix contains the results of acoustic analyses. In Table B.1, we present the results of linear mixed effects models conducted on the closure duration,

release (positive voice-onset-time) duration, and total duration of all C<sub>3</sub> [k, g]. In Table B.2, we present the results of linear mixed effects models conducted on the root-mean-square (rms) amplitude around the burst of all C<sub>3</sub> [k, g].

**Table B.1**

Parametric analysis of the acoustic closure duration of C<sub>3</sub> in different stress conditions. Each consonant is measured in its closure duration, positive voice-onset-time duration, and total duration. The likelihood ratio tests of linear mixed effects models test the main effect of the predictor STRESS (C<sub>3</sub> at far (baseline) and close positions respectively).  $\chi^2$  values (with one degree of freedom) are reported as effect size estimates.  $\beta$ -coefficients, standard errors (SE) and  $p$ -values of the models are also reported. The models marked with # are reduced models containing only random intercepts.

| Acoustic durations (ms) |           | Mean (sd)      |                | $\chi^2$ | $\beta$ | SE   | $p$    |
|-------------------------|-----------|----------------|----------------|----------|---------|------|--------|
|                         |           | FAR            | CLOSE          |          |         |      |        |
| Closure duration        | in [aki]  | 62.58 (17.41)  | 71.93 (18.31)  | 21.39    | 9.41    | 1.34 | <0.001 |
|                         | in [agi]  | 56.52 (15.14)  | 65.65 (13.73)  | 16.20    | 9.03    | 1.70 | <0.001 |
|                         | in [eki]  | 61.98 (18.35)  | 71.36 (20.56)  | 18.11    | 9.27    | 1.53 | <0.001 |
|                         | in [egi]# | 55.31 (16.14)  | 63.97 (16.37)  | 32.23    | 8.61    | 1.49 | <0.001 |
| VOT duration            | in [aki]  | 68.88 (17.08)  | 68.62 (19.49)  | 0.01     | -0.18   | 1.52 | =0.91  |
|                         | in [agi]# | 22.90 (14.74)  | 25.41 (14.21)  | 2.04     | 1.92    | 1.35 | =0.15  |
|                         | in [eki]  | 64.17 (19.42)  | 68.17 (21.32)  | 3.49     | 3.93    | 1.98 | =0.06  |
|                         | in [egi]# | 21.99 (14.59)  | 23.03 (14.74)  | 0.61     | 1.03    | 1.32 | =0.44  |
| Total duration          | in [aki]  | 131.46 (23.24) | 140.55 (25.68) | 10.61    | 9.42    | 2.39 | <0.01  |
|                         | in [agi]  | 79.42 (19.45)  | 91.06 (18.77)  | 13.43    | 10.80   | 2.33 | <0.001 |
|                         | in [eki]  | 126.15 (23.40) | 139.52 (25.33) | 18.16    | 13.11   | 2.18 | <0.001 |
|                         | in [egi]  | 77.30 (21.14)  | 87.00 (22.40)  | 14.27    | 9.60    | 1.97 | <0.001 |

**Table B.2**

Parametric analysis of the root-mean-square (rms) amplitude of the burst in C<sub>3</sub> in different stress conditions. The rms amplitude is measured as the mean of a 10 ms window around the acoustic burst of each C<sub>3</sub> (5 ms before the burst and 5 ms after it). The likelihood ratio tests of linear mixed effects models test the main effect of the predictor STRESS (C<sub>3</sub> at far (baseline) and close positions respectively).  $\chi^2$  values (with one degree of freedom) are reported as effect size estimates.  $\beta$ -coefficients, standard errors (SE) and  $p$ -values of the models are also reported.

|          | Mean (sd)   |             | $\chi^2$ | $\beta$ | SE   | $p$   |
|----------|-------------|-------------|----------|---------|------|-------|
|          | FAR         | CLOSE       |          |         |      |       |
| in [aki] | 1.53 (1.01) | 1.80 (1.04) | 11.69    | 0.26    | 0.11 | <0.01 |
| in [agi] | 5.61 (3.34) | 4.46 (2.76) | 8.34     | -1.29   | 0.39 | <0.01 |
| in [eki] | 1.19 (0.77) | 1.44 (0.93) | 11.24    | 0.24    | 0.07 | <0.05 |
| in [egi] | 5.00 (2.92) | 5.02 (3.08) | 0.02     | 0.06    | 0.41 | =0.88 |

**Appendix C. Additional results of analysis based on gestural landmarks**

In this appendix, we present the statistical analyses conducted on the measurements and mass-spring

parameters based on gestural landmarks of C<sub>3</sub> [k, g]. In Table C.1, we present the results of linear mixed effect models conducted on commonly used measurements and mass-spring parameters, based on the gestural landmarks of C<sub>3</sub>.

**Table C.1**

Parametric analysis of the measurements and mass-spring parameters of C<sub>3</sub> in different stress conditions based on gestural landmarks. The gestural landmarks are obtained as described in Section 2.4. The likelihood ratio tests of linear mixed effects models test the main effect of the predictor STRESS (C<sub>3</sub> at far (baseline) and close positions respectively).  $\chi^2$  values (with one degree of freedom) are reported as effect size estimates.  $\beta$ -coefficients, standard errors (SE) and  $p$ -values of the models are also reported. The model of stiffness in [agi] context contains only random intercept due to convergence issue and is marked with #.

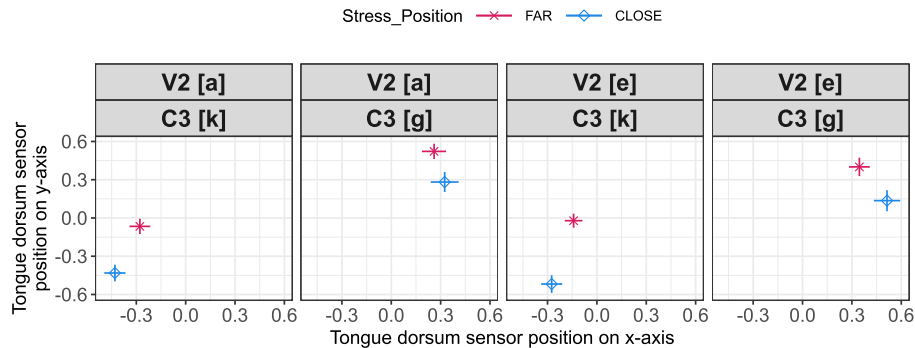
| Measurements & Parameters   |           | Mean (sd)      |                | $\chi^2$ | $\beta$ | SE    | $p$    |
|---|-----------|----------------|----------------|----------|---------|-------|--------|
|   |           | FAR            | CLOSE          |          |         |       |        |
| Onset-to-target duration (ms)   | in [aki]  | 89.64 (23.93)  | 108.26 (20.92) | 11.98    | 19.16   | 4.48  | <0.001 |
|   | in [agi]  | 92.80 (18.98)  | 110.21 (17.38) | 20.09    | 16.57   | 2.61  | <0.001 |
|   | in [eki]  | 81.21 (23.93)  | 95.57 (23.83)  | 9.81     | 14.57   | 3.93  | <0.01  |
|   | in [egi]  | 81.18 (21.56)  | 95.32 (25.49)  | 8.53     | 17.02   | 5.05  | <0.01  |
| Onset-to-target displacement (mm)   | in [aki]  | 9.47 (3.84)    | 13.94 (4.42)   | 22.18    | 4.41    | 0.62  | <0.001 |
|   | in [agi]  | 10.52 (3.62)   | 14.86 (2.79)   | 22.21    | 4.63    | 0.63  | <0.001 |
|   | in [eki]  | 5.86 (2.78)    | 5.68 (2.43)    | 0.03     | -0.13   | 0.74  | =0.86  |
|   | in [egi]  | 6.13 (2.98)    | 5.60 (2.22)    | 0.07     | 0.26    | 0.99  | =0.80  |
| Peak velocity (mm/s)  | in [aki]  | 177.24 (59.79) | 187.68 (49.32) | 0.46     | 6.25    | 9.10  | =0.50  |
|   | in [agi]  | 202.22 (60.57) | 210.90 (45.58) | 2.98     | 10.89   | 6.04  | =0.08  |
|   | in [eki]  | 123.21 (49.47) | 98.70 (39.96)  | 4.85     | -23.30  | 9.73  | <0.05  |
|   | in [egi]  | 133.95 (50.83) | 99.37 (38.81)  | 4.22     | -24.49  | 10.99 | <0.05  |
| Ratio peak velocity/displacement<br>(Munhall stiffness (Munhall et al. (1985))) | in [aki]  | 18.28 (5.06)   | 13.50 (3.33)   | 24.86    | -4.89   | 0.61  | <0.001 |
|   | in [agi]# | 18.92 (4.55)   | 14.36 (2.47)   | 41.73    | -7.31   | 1.10  | <0.001 |
|   | in [eki]  | 21.06 (5.15)   | 18.23 (4.43)   | 7.58     | -2.94   | 0.94  | <0.01  |
|   | in [egi]  | 21.14 (4.76)   | 18.31 (3.85)   | 9.47     | -2.97   | 0.82  | <0.01  |
| Acceleration phase (ms)<br>(i.e., duration from onset to peak velocity)         | in [aki]  | 45.37 (15.91)  | 66.08 (16.68)  | 26.65    | 20.43   | 2.40  | <0.001 |
|   | in [agi]  | 47.7 (14.28)   | 64.95 (11.85)  | 12.23    | 14.21   | 3.27  | <0.001 |
|   | in [eki]  | 40.70 (13.57)  | 52.52 (15.79)  | 16.32    | 12.43   | 2.29  | <0.001 |
|   | in [egi]  | 41.69 (12.60)  | 53.91 (15.96)  | 18.37    | 12.50   | 2.10  | <0.001 |
| Deceleration phase (ms)<br>(i.e., duration from peak velocity to target)        | in [aki]  | 44.28 (17.15)  | 42.18 (14.33)  | 0.49     | -2.09   | 2.94  | =0.48  |
|   | in [agi]  | 45.11 (13.67)  | 45.26 (14.00)  | 0        | 0.01    | 1.70  | =0.99  |
|   | in [eki]  | 40.51 (17.95)  | 43.05 (16.90)  | 0.86     | 2.14    | 2.28  | =0.35  |
|   | in [egi]  | 39.49 (15.22)  | 41.41 (16.95)  | 1.08     | 3.66    | 3.49  | =0.30  |
| Target-to-release plateau duration (ms)   | in [aki]  | 102.7 (65.14)  | 95.25 (58.15)  | 1.52     | -5.56   | 4.41  | =0.21  |
|   | in [agi]  | 74.77 (49.98)  | 76.27 (53.58)  | 0.002    | -0.20   | 4.45  | =0.96  |
|   | in [eki]  | 119.48 (67.66) | 117.84 (66.43) | 0.03     | 1.04    | 6.19  | =0.87  |
|   | in [egi]  | 102.10 (63.25) | 94.70 (62.15)  | 0.58     | -3.19   | 4.19  | =0.45  |



**Appendix D. Additional results of parametric analysis of the averaged absolute tongue dorsum positions of V<sub>3</sub> [i]**

In this appendix, we present the parametric analysis of the averaged absolute tongue dorsum positions of V<sub>3</sub>, obtained

using the window method (Roessig, 2021; Roessig et al., 2022). This appendix shows that the V<sub>3</sub> [i] does not have a x-axis (front-back) difference with respect to its relative position to lexical stress.



**Fig. D.1.** Z-scored mean tongue dorsum positions in V<sub>3</sub> [i], obtained over the the mid-50% of the acoustic duration using the window method, presented with standard errors. The red crosses and error bars correspond to when the stress is on the first syllable [C<sub>1</sub>V<sub>1</sub>] (i.e., when [k, g] consonants are in the FAR condition); the blue diamonds and error bars correspond to when the stress is on the second syllable [C<sub>2</sub>V<sub>2</sub>] (i.e., when [k, g] consonants are in the CLOSE condition). The y-axis represents the high-low dimension and the x-axis represents the front-back (left to right) dimension.

**Table D.1**

Parametric analysis of z-scored mean positions of tongue dorsum sensor during V<sub>3</sub> in different stress conditions. The mean position values are obtained over the mid-50% of the acoustic duration of V<sub>3</sub>. The likelihood ratio tests of linear mixed effects models test the main effect of the predictor STRESS (C<sub>3</sub> at far (baseline) and close positions respectively).  $\chi^2$  values (with one degree of freedom) are reported as effect size estimates.  $\beta$ -coefficients, standard errors (SE) and  $p$ -values of the models are also reported. The models contain only random intercept due to convergence issue are marked with #.

|          | Dimension | $\chi^2$ | $\beta$ | SE   | $p$    |
|----------|-----------|----------|---------|------|--------|
| in [aki] | y         | 10.81    | -0.35   | 0.10 | <0.01  |
|          | x         | 1.84     | -0.27   | 0.19 | =0.17  |
| in [agi] | y#        | 5.17     | -0.27   | 0.11 | <0.05  |
|          | x         | 0.48     | 0.19    | 0.27 | =0.48  |
| in [eki] | y         | 13.05    | -0.49   | 0.11 | <0.001 |
|          | x         | 0.19     | -0.08   | 0.18 | =0.66  |
| in [egi] | y         | 1.98     | -0.22   | 0.15 | =0.15  |
|          | x#        | 2.90     | 0.26    | 0.15 | =0.09  |

**References**

Albano Leoni, F., Cutugno, F., & Savy, R. (1995). The vowel system of Italian connected speech. In *Proceedings of the international congress of phonetic sciences* (pp. 396–399).

Anttila, A. (2003). Finnish assibilation. In M. Kadowaki, & S. Kawahara (Eds.), *The Proceedings of NELS 33* (pp. 13–24). Amherst, MA: GLSA.

Armstrong, L. E. (1964). *The phonetic structure of Somali*. Gregg International Publishers Limited.

Audacity Team (2022). Audacity(R): Free Audio Editor and Recorder. URL: <https://audacityteam.org/>.

Bateman, N. (2007). *A crosslinguistic investigation of palatalization* Ph.D. thesis. San Diego: University of California.

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

Bauer, W. (2003). *Maori*. Routledge.

Beckman, J. N. (1998). *Positional faithfulness*. University of Massachusetts Amherst.

Beckman, M. E., & Edwards, J. (1994). *Articulatory evidence for differentiating stress categories. Phonological structure and phonetic form: Papers in laboratory phonology III*.

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Segment, gesture, prosody chapter 3* (pp. 68–89). Cambridge University Press.

Bertinetto, P. M. (1980). The perception of stress by Italian speakers. *Journal of Phonetics*, 8, 385–395.

Boersma, P., & Weenink, D. (2023). Praat: doing phonetics by computer. URL: <http://www.praat.org>.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6, 201–251.

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

Buech, P., Ridouane, R., & Hermes, A. (2022a). Pharyngealization in Amazigh: Acoustic and articulatory marking over time. In *Interspeech* (pp. 3448–3452).

Buech, P., Roessig, S., Pagel, L., Muecke, D., & Hermes, A. (2022b). *ema2wav: doing articulation by Praat*. In *Interspeech* (pp. 1352–1356).

Burzio, L. (2007). *Phonology and phonetics of English stress and vowel reduction. Language Sciences*, 29, 154–176.

Bye, P., & de Lacy, P. (2008). Metrical influences on fortition and lenition. In J.B. de Carvalho, T. Scheer, & P.Ségéral (Eds.), *Lenition and Fortition* (pp. 173–206). De Gruyter Mouton volume 99 of *Studies in Generative Grammar [SGG]*.

Celata, C., & Bertinetto, P. M. (2005). Lexical access in Italian: words with and without palatalization. *Lingue e linguaggio*, 4, 293–318.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 32, 141–176.

Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. *Laboratory Phonology*, 8, 519–548.

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35, 210–243.

- Coretta, S., van Rij, J., & Wieling, M. (2023). tidyvm: Tidy Model Visualisation for Generalised Additive Models. URL: <https://stefanocoretta.github.io/tidyvm/> r package version 3.3.2.
- Crosswhite, K. (1999). Vowel reduction in Optimality Theory. Ph.D. thesis UCLA.
- Dimitrova, S., & Turk, A. (2012). Patterns of accentual lengthening in English four-syllable words. *Journal of Phonetics*, 40, 403–418.
- Engstrand, O. (1988). Articulatory correlates of stress and speaking rate in Swedish vcv utterances. *The Journal of the Acoustical Society of America*, 83, 1863–1875.
- Erickson, D. (2002). Articulation of extreme formant patterns for emphasized vowels. *Phonetica*, 59, 134–149.
- Eriksson, A., Bertinetto, P.M., Heldner, M., Nodari, R., & Lenoci, G. (2016). The acoustics of lexical stress in Italian as a function of stress level and speaking style. In *Interspeech* (pp. 1059–1063).
- Farnetani, E., & Kori, S. (1986). Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication*, 5, 17–34.
- Flemming, E. (2005). *A phonetically-based model of phonological vowel reduction*. Ms. MIT.
- Flor, E. (2023). *Base-derivative faithfulness in Italian palatalization*. Ms. MIT.
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, 101, 3728–3740.
- Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical Society of America*, 90, 1816–1827.
- Gafos, A. I., & Benus, S. (2006). Dynamics of phonological cognition. *Cognitive Science*, 30, 905–943.
- Giavazzi, M. (2008). On the application of velar palatalization in Italian. In Proceedings of the 27th West Coast Conference on Formal Linguistics Poster Session, UCLA Working Papers in Linguistics (pp. 20–29).
- Giavazzi, M. (2010). The phonetics of metrical prominence and its consequences on segmental phonology. In *Ph.D. thesis Massachusetts Institute of Technology*.
- Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 215–249). Cambridge University Press.
- González, C. (2003). *The effect of stress and foot structure on consonantal processes*. University of Southern California.
- Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, 3, 1–11.
- Guion, S. G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica*, 55, 18–52.
- Harrington, J., Fletcher, J., & Beckman, M. E. (2000). Manner and place conflicts in the articulation of accent in Australian English. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon chapter 3* (pp. 40–51). Cambridge University Press.
- Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.
- Hollenbach, B. E. (1977). Phonetic vs. phonemic correspondence in two Trique dialects. In W. R. Merrifield (Ed.), *Studies in Otomanguan phonology chapter 3* (pp. 35–67). The Summer Institute of Linguistics and The University of Texas at Arlington.
- Hyman, L. M. (1975). *Phonology: theory and analysis*. Rinehart and Winston: Holt.
- Iskarous, K., & Poupplier, M. (2022). Advancements of phonetics in the 21st century: A critical appraisal of time and space in articulatory phonology. *Journal of Phonetics*, 95, 101195.
- Jones, T. E., & Knudson, L. M. (1977). Guelavía Zapotec phonemes. In W. R. Merrifield (Ed.), *Studies in Otomanguan phonology chapter 8* (pp. 163–180). The Summer Institute of Linguistics and The University of Texas at Arlington.
- de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America*, 97, 491–504.
- de Jong, K. J. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32, 493–516.
- de Jong, K. J., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech*, 36, 197–212.
- Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics*, 55, 149–181.
- Katsika, A., & Tsai, K. (2021). The supralaryngeal articulation of stress and accent in Greek. *Journal of Phonetics*, 88, 101085.
- Kochetov, A. (1998). Labial palatalization: A gestural account of phonetic implementation. *The Canadian Linguistics Association Annual Proceedings*, 3850.
- Kochetov, A. (2011). Palatalization. In M. van Oostendorp, C.J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology chapter 11*. (pp. 1–25). Wiley Online Library volume 3.
- Kochetov, A. (2013). *Production, Perception, and Phonotactic Patterns: A Case of Contrastive Palatalization*. Routledge.
- Krämer, M. (2009). *The phonology of Italian*. Oxford University Press.
- Krivokapić, J. (2020). *Prosody in articulatory phonology. Prosodic theory and practice*. Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35, 1773–1781.
- Marcet, A., Fernández-López, M., Baciero, A., Sesé, A., & Perea, M. (2022). What are the letters e and é in a language with vowel reduction? the case of Catalan. *Applied Psycholinguistics*, 43, 193–210.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kald. In *Interspeech* (pp. 498–502).
- Mitchell, S. (2022). Investigating the perceptual basis for stress-conditioned palatalization in Italian. CogMaster MA thesis, ms. DEC, ENS-PSL.
- Munhall, K. G., Ostry, D. J., & Parush, A. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 457.
- Nobre, M. A., & Ingemann, F. (1983). Oral vowel reduction in Brazilian Portuguese. *Kansas Working Papers. Linguistics*, 8.
- Oh, S. (2022). *Incomplete neutralization in articulatory phonology* Ph.D. thesis. City University of New York.
- Ohala, J. (1992). What's cognitive, what's not, in sound change. In M. D. Morrissey & K. Günter (Eds.), *Diachrony within synchrony: Language History and Cognition: Papers from the International Symposium at the University of Duisburg* (pp. 309–355). Peter Lang.
- Payne, E. M. (2005). Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association*, 35, 153–181.
- Petrosino, R., & Calabrese, A. (2021). Palatalization in Romance. *Manual of Romance Phonetics and Phonology*, 27, 173–214.
- R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria. URL: <https://www.R-project.org/>.
- Recasens, D. (2020). Palatalizations in the Romance Languages. In *In Oxford Research Encyclopedia of Linguistics*.
- Roessig, S. (2021). *Categoriality and continuity in prosodic prominence* (Volume 10). Language Science Press.
- Roessig, S., Winter, B., & Mücke, D. (2022). Tracing the phonetic space of prosodic focus marking. *Frontiers in Artificial Intelligence*, 5, 842546.
- Rohlf, G., Persichino, S., Franceschi, T., & Caciagli Fancelli, M. (1972). Grammatica storica della lingua Italiana e dei suoi dialetti. *The Modern Language Review*, 67, 660–663.
- Ryan, K.M. (2019). *Prosodic Weight: Categories and Continua* volume Oxford Studies in Phonology and Phonetics. Oxford University Press.
- Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody (speech prosody 2008), Campinas, Brazil* (pp. 175–184).
- Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333–382.
- Samareh, Y. (1977). *The arrangement of segmental phonemes in Farsi*. University of Tehran, Faculty of Letters.
- Savariaux, C., Badin, P., Samson, A., & Gerber, S. (2017). A comparative study of the precision of Carstens and Northern Digital Instruments electromagnetic articulographs. *Journal of Speech, Language, and Hearing Research*, 60, 322–340.
- Shao, B., Buech, P., Hermes, A., & Giavazzi, M. (2023a). Lexical stress and velar palatalization in Italian: A spatio-temporal interaction. In *Interspeech* (pp. 1833–1837). Dublin, Ireland.
- Shao, B., Buech, P., Hermes, A., & Giavazzi, M. (2023b). Stress conditioned phonological process: A case study of Italian palatalization. In Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023) (pp. 2189–2193). Prague, Czech Republic.
- Shao, B., Hermes, A., Buech, P., & Giavazzi, M. (2023c). Velar palatalization in Italian: Lexical stress induces resistance to sound change. In *HISPhonCog2023*. Seoul, Korea.
- Sluijter, A.M., & Van Heuven, V.J. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* (pp. 630–633). IEEE volume 2.
- Smith, J. (2002). *Phonological Augmentation in Prominent Positions*. Ph.D. thesis University of Massachusetts, Amherst.
- Steddy, S. (2015). Palatalisation across the Italian lexicon. Ms., Rutgers Optimality Archive (ROA).
- Stella, M., Stella, A., Sigona, F., Bernardini, P., Grimaldi, M., & Fivela, B.G. (2013). Electromagnetic articulography with AG500 and AG501. In *Interspeech* (pp. 1316–1320).
- Steriadi, D. (2011). Rhyming evidence for intervals. In *Colloquium presentation at the UCLA Department of Linguistics, March*. volume 4.
- Steriadi, D. (2023). *Resyllabification revisited: arguments for v-to-v intervals as units of rhythm*. Seoul, South Korea: HISPhonCog.
- Suomi, K., & Ylitalo, R. (2004). On durational correlates of word stress in Finnish. *Journal of Phonetics*, 32, 35–63.
- Svensson Lundmark, M. (2023). Rapid movements at segment boundaries. *The Journal of the Acoustical Society of America*, 153, 1452–1467.
- Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics*, 25, 25–41.
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Veerman-Leichsenring, A. (1984). *El Popoloca de los Reyes Metzontla* volume 4 of *América Indígena*. Número Spécial. Paris: Association d'Ethnolinguistique Américainienne.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4, 1686. <https://doi.org/10.21105/joss.01686>.

- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between I1 and I2 speakers of english. *Journal of Phonetics*, 70, 86–116.
- Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., & Baayen, R. H. (2016). Investigating dialectal differences using articuography. *Journal of Phonetics*, 59, 122–143.
- Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, 30, 945–982.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73, 3–36.