INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Lexical stress and velar palatalization in Italian: A spatio-temporal interaction

Bowei Shao[1,2], *Philipp Buech*[2], *Anne Hermes*[2], *Maria Giavazzi*[1]

[1] Département d'études cognitives, École Normale Supérieure, Université PSL, Paris, France
[2] Laboratoire de Phonétique et Phonologie, UMR7018, CNRS/Sorbonne Nouvelle, Paris, France
`{bowei.shao;maria.giavazzi}@ens.psl.eu`, `{philipp.buech;anne.hermes}@sorbonne-nouvelle.fr`

Abstract

Palatalization is the process whereby a velar stop is fronted to a palatal affricate or fricative. In Italian, it takes place at the boundary between the root and /i/ suffixes. In nouns and adjectives, palatalization occurs in words with antepenultimate stress ([ˈko.mi.t͡ʃi]), while it is much rarer in words with penultimate stress ([ka.ˈdu.ki]) Based on one acoustic and one articulatory study (EMA), we postulate that the resistance of post-tonic /k, g/ to palatalize is related to the stressed vowel directly preceding. In the acoustic domain, post-tonic consonants show longer closure duration. This increase in closure duration is directly related to a larger and longer tongue dorsum movement in the articulatory domain. We show an interaction between temporal (closure duration) and spatial (tongue dorsum displacement) aspects of lexical stress, which we interpret as the cause of resistance to palatalization in post-tonic velars. The findings are discussed within the μ-gesture framework.

Index Terms: palatalization, Italian, articulation, *μ*-gesture, lexical stress

1. Introduction

Palatalization is the process through which a velar stop, /k, g/, is fronted to a palatal/palato-alveolar affricate or fricative. It applies more frequently before high and mid-high front vowels (e.g., /i, e/) than before other vowels. In Romance languages, the origins of this phonological process are to be found in Late Latin: this process is known as the 2nd Romance palatalization, and it occurred after the 5th century CE [1, 2]. It occurred before all front vowels, both root internally and at the morpheme boundary, and independently of morpheme boundaries and stress position [3].

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 words with antepenultimate stress, while it is much rarer in words with penultimate stress [4, 5]. This is illustrated in (1).

(1) FAR [ˈko.mi.ko]–[ˈko.mi.**t͡ʃ**i] 'comedian'-'comedians'
POST [ka.ˈdu.ko]–[ka.ˈdu.**k**i] 'caducous'-'caducous *pl.*'

The case of (FAR) presents a stressed syllable FAR from the position of palatalization, while in (POST) the /k/ is directly POST-tonic. Note that in FAR, the underlying /k/ is realized as [t͡ʃ] before /i/, in POST it is realized as [k].

Whereas the perceptual and articulatory underpinnings of palatalization before high front vowels have been studied rather extensively [6, 7], the phonetic bases of the stress-conditioned process are not well understood (cf. [5]).

Lexically prominent positions such as the stressed domain (I) resist the application of phonological processes applying elsewhere, and they manifest the positional maintenance of contrasts otherwise neutralized [8]. For instance, vocalic contrasts are preferentially realized in stressed positions, but they may be reduced in unstressed positions (e.g., in English, Brazilian Portuguese, Western Catalan [9, 10, 11]). Similar patterns are observed for consonantal contrasts (e.g., in Copala Trique, and Finnish [12, 13]). However, prominent positions are also (II) the preferred target for a small class of frequent phonetic processes: e.g., consonants are often lengthened in pre-tonic and post-tonic positions (as in English and Somali [14, 15]), consonants in these two positions are described as having louder (e.g., Farsi [16, 17]) or affricate-like bursts (e.g., Maori [18]).

The two seemingly contradictory linguistic behaviors, (I) and (II), may have a common phonetic basis: they may be the articulatory and perceptual by-product of lexical prominence. A syllable nucleus bearing lexical prominence is known to be longer in duration, higher in f0, more peripheral and more intense [19]. While f0 and intensity are less related to oral cavity activities, the hyper-articulated vowel qualities related to lexical prominence shall manifest on the lingual articulation of the nucleus and have effects on adjacent segments [20, 21].

The modulation of the articulatory system needed for the realization of lexical prominence is described within Articulatory Phonology, in terms of an abstract spatio-temporal μ-gesture [21]. A μ-conditioned vowel has longer, larger, and faster gestures than its unstressed counterparts [22]. This stress-conditioned modulation is predicted to have a direct impact on its surrounding segments. Furthermore, given that the modulation gesture has a symmetrical shape [23], it is predicted to affect articulatory properties of pre-tonic segments as well as post-tonic segments (such as the second [k] in [ka.'du.ki], regardless of a syllabic boundary.

The aim of this study is to explore the acoustic and articulatory bases of stress-conditioned palatalization in Italian, specifically in relation to the spatio-temporal modulation μ-gesture. We postulate that the resistance of POST-tonic /k, g/ to palatalize is related to the μ-conditioned articulation of the stressed vowel directly preceding them. In POST-tonic position, the μ-conditioned vowel shall have larger and longer movements, which shall lead to a later target achievement of the following gestures. The delayed target achievement shall interact with the following consonantal gestures, since the μ-conditioned stressed vocalic gesture and the velar closing gesture of [k, g] both recruit the tongue dorsum.

2. Method

First, an acoustic recording was conducted on 18 native Italian speakers (mean age=27.2; 9 female, 9 male). Target words were

10.21437/Interspeech.2023-1456

trisyllabic nonce words, structured /C_1V_1.C_2V_2.C_3V_3/, differing solely by the position of stress on the first or the second syllable (e.g., /ˈpi.ta.ki/, /pi.ˈta.ki/). The nonce words were designed to compare how the target consonants /k, g, t͡ʃ, d͡ʒ/ (in C_3 position) were produced in both FAR and POST contexts. V_3 was always /i/. C_1V_1.C_2V_2 sequences were /pita/, /fesa/, /pufa/ /tipa/ and /suta/. For example, in the nonce words /ˈpi.ta.ki/ and /pi.ˈta.ki/, C_3 /k/ was in FAR and POST contexts respectively. The nonce words were written in their orthographic forms with lexical stress marked by an accent (e.g., pítachi, pitáchi). They were embedded in the carrier phrase "*Dimmi __ di nuovo*" (Engl. "Say ____ again.") and randomized. The acoustic recordings were conducted in a double-walled sound booth on a TASCAM DR-100MKIII Linear PCM Recorder. Sound files were recorded at a sampling rate of 44.1 kHz. Target words were repeated three times, yielding a list of 3328 nonce words.

Second, an articulatory recording was conducted on five speakers (mean age=26.6; 3 female, 2 male). Articulatory and acoustic data were collected simultaneously using the Electromagnetic Articulograph AG 501 (Carstens Medizinelektronik Gmbh). Sensors were placed on upper and lower lips, tongue tip, tongue blade, and tongue dorsum, with additional sensors behind the left and right ear for head correction. The articulatory signal was recorded with a sample rate of 250 Hz and filtered using a Butterwort lowpass filter with a cut-off frequency of 25 Hz and order 5 afterwards. The acoustic signal was recorded at a sample rate of 44.1 kHz and 16-bit resolution. The structure of the target words was the same as in the acoustic study, except that the V_2 position was occupied by both /a/ and /e/, and C_1, C_2 alternated between /p, t/. The nonce words were embedded in the carrier phrase "*Pimpa parte da __ la mattina presto*" (Engl. "Pimpa leaves from __ early in the morning.") and randomized. Target words were repeated three times, yielding a list of 1256 nonce words.

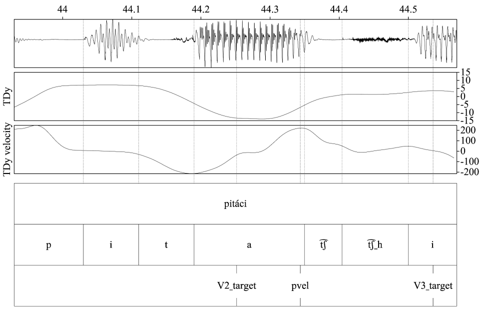

Figure 1: *An example of a [pitát͡ʃi] sequence with waveform (top), tongue dorsum trajectory in y-dimension (TDy, mid), velocity in y-dimension (TDy velocity, bottom). Dashed lines mark the acoustic boundaries and gestural landmarks as indicated in the annotation.*

The utterances from the acoustic recording and from the EMA recording (A, B) were pre-segmented using Audacity v3.1.3 [24]. They were then extracted and automatically annotated using the Montreal Forced Aligner v2.0.6 [25]. For the acoustic recordings, stop bursts were automatically identified by a custom script written in Python v.3.10.6 [26] utilizing a modified version of the algorithm described in [25]. The acoustic data collected during the EMA recording was used as reference to annotate the articulatory data, the alignments (see Figure 1) were manually inspected and corrected. The articulatory targets of V_2 and V_3 as well as the peak velocity were identified automatically using a custom Python script. The EMA data was then converted using ema2wav [27] and the landmarks were manually corrected in Praat [28].

For the acoustic analyses, we measured the closure duration and total duration of the plosives and affricates in C_3 position. We calculated the *closure ratio* by taking the ratio between *closure duration* and *total duration* of C_3. Linear mixed-effects models were run for stops and affricates separately using lme4 (1.1-30) [29]. The model structure was the same for both datasets: Closure ratio was taken as the dependent variable, and C_3 (/k, g/; /t͡ʃ, d͡ʒ/), stress (POST, FAR) and their interaction as fixed effects. By-speaker random slopes and intercepts for C_3 and random intercepts for the source (acoustics from A and B) were added as random effects. Post-hoc tests were conducted using lmerTest (2.0-33) [30] with default settings.

For the articulatory analysis, we compared the trajectories of the vocalic movement from V_2 to V_3 as a whole and applied the following mass-spring parameters on the tongue dorsum movement (high-low, y) [31, 32]: *peak velocity* as the maximum speed for V_3 formation, *displacement* from V_2-target to V_3-target, *onset-to-target duration* (i.e., gestural activation interval) of V_3. The trajectory was measured in 20 steps for tongue dorsum over the time-course of the movement from V_2-target to V_3-target (i.e., gestural activation interval of V_3). The EMA data was normalized for each speaker according to the highest y-position measured [33]. The onset-to-target movement of V_3 was modeled with generalized additive mixed models (GAMM) using the mgcv package (1.8-40) [34] in R (4.2.2) [35], and visualized using Tidyverse (1.3.2) and Tidymv (3.3.2) [36, 37]. Four GAMMs were performed on each type of C_3 with stress position (V_1, V_2) as fixed effect and smoothing parameters, with by-words factor smooths for speakers as random effect. The stiffness of the trajectories was calculated by taking the ratio between peak velocity and displacement, following [22, 38]. We report here only on sequences wherein V_2 is [a], since both acoustic and EMA recordings contain these nonce words.

3. Results

3.1. Closure ratio in C_3 (acoustics)

Table 1: *Fixed effects of linear mixed-effects analyses with post-hoc tests conducted on closure ratio of C_3, grouped by voicing and manner of articulation. *=<0.5, **<0.01, ***<0.001.*

	Estimate	**S.E.**	***p***
/k/ POST (Intercept)	0.569	0.017	***
/g/ POST	0.111	0.031	**
/k/ FAR	-0.038	0.011	***
/g/ FAR	0.041	0.016	**
/t͡ʃ/ POST (Intercept)	0.414	0.041	**
/d͡ʒ/ POST	0.137	0.017	***
/t͡ʃ/ FAR	-0.014	0.01	0.152
/d͡ʒ/ FAR	0.014	0.014	0.312
Satterthwaite post-hoc (POST vs. FAR)			
/k/	0.038	0.011	***
/g/	-0.003	0.011	0.8
/t͡ʃ/	-0.014	0.01	0.152
/d͡ʒ/	-0.0001	0.017	0.99

The linear mixed-effects analysis on closure ratio in C_3 is reported in Table 1. Recall that higher closure ratio implies longer closure duration in relation to total duration.

We observed that the closure ratio in [k, t͡ʃ] is much higher compared to their voiced counterparts, implying that voiceless C_3 have longer closure duration than voiced C_3. As can also be seen in the estimates, the closure ratio is significantly higher when C_3 is in POST-tonic position. However, as the post-hoc tests showed, the overall effect of stress is mainly driven by [k]. This significant difference in [k] suggests that the stressed V_2 has a strong impact on the closure duration of the following [k].

3.2. Tongue dorsum vertical position from V_2 to V_3 (articulation)

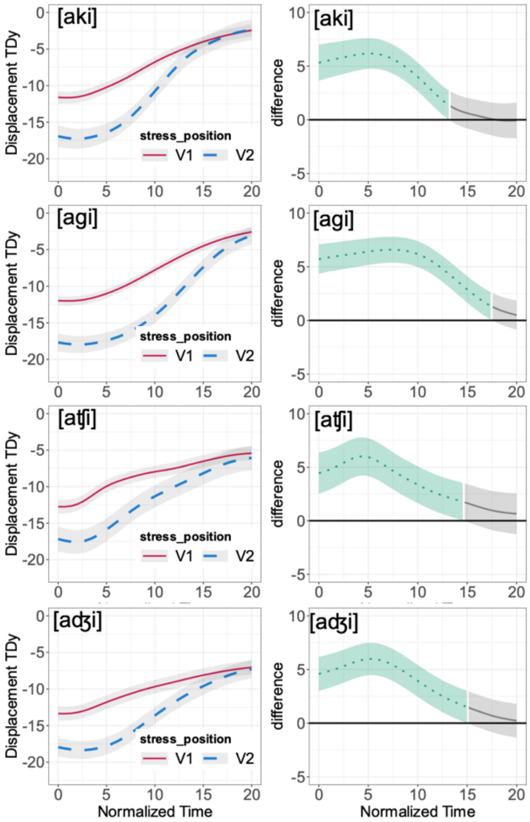

Figure 3: *GAMM fitted Tongue Dorsum displacement in [aC₃i] sequences in y-dimension by normalized time, with estimated differences. The y-dimension was normalized according to the highest y-position measured. Estimated differences are presented with 95% confidence intervals by time (normalized). The green dotted lines represent the significant difference in stress-on-V_1 (C_3 FAR) vs. stress-on-V_2 (C_3 POST) contexts.*

One of the objectives of this study is to quantify the vertical tongue dorsum movement (low-high) from V_2 to V_3 (i.e., onset-to-target of [i]) according to stress position. The difference in tongue trajectories between two stress conditions is shown in Figure 3 (solid red lines show the trajectories when stress is on V_2 and blue dashed lines show when stress is on V_3). When stress is on V_2, the target of V_2 is much lower (blue dashed line) compared to when V_2 is unstressed (red solid line) This difference is significant and confirmed by GAMM estimated differences (green dotted lines in Figure 3, right column).

We can also observe from Figure 3 that the target of V_3 [i] is stable and unaffected by stress position. This indicates that V_3 is achieved regardless of the position of stress. In other words, the position of lexical stress does not have an impact on the lingual target achievement of V_3, but on *how* the target of V_3 is achieved. This point is of great interest in this study since palatalization is triggered precisely by V_3 [i]. A stable V_3 [i] target confirms our hypothesis: the non-palatalization of [k, g] in POST-tonic position (i.e., when stress is on V_2) is not related to the articulatory specificities of the trigger itself (i.e., V_3 [i]) but to how the trigger is achieved.

3.3. Tempo-spatial analyses of tongue dorsum movement (articulation)

To shed further light on how the trigger V_3 [i] is achieved and what parameters are modified in two stress positions (i.e., stress on V_1/V_2), we analyzed the mass-spring parameters presented in Figure 4.

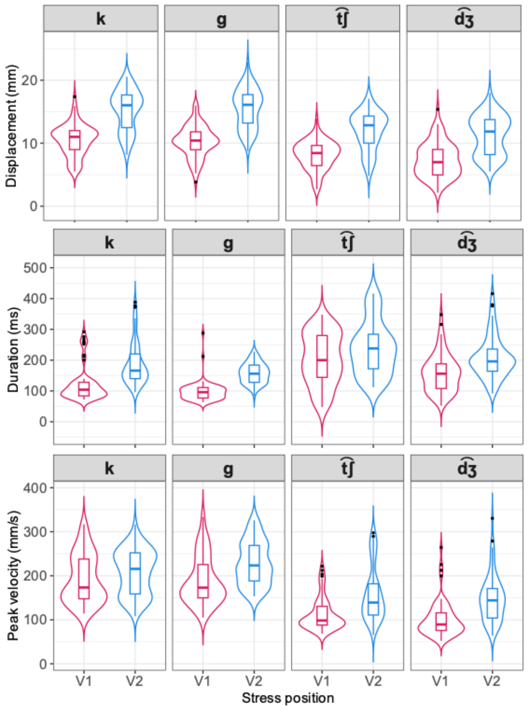

Figure 4: *Violin-box plots showing displacement, peak velocity, and duration of onset-to-target gesture of V_3 in Tongue Dorsum y-dimension, according to C_3 and the position of stress. Stress on V_1/V_2 corresponding to the relative position of [k, g, t͡ʃ, d͡ʒ] as FAR from or POST stress.*

For the *displacement* (i.e., how far the articulator travels on low-high dimension), we see a clear pattern based on stress position: the displacement is systematically larger when stress is on V_2. This behavior is also reported in Greek stressed vowels [22], confirming that lexical stress has an impact on the articulatory displacement of the vocalic movement. This pattern also confirms the GAMM estimations reported in Figure 3, which show that when V_2 [a] is stressed, the tongue dorsum lowers further. Consequently, it must travel from a lower position to reach the V_3 [i] target.

For the onset-to-target *duration* (i.e., how long the articulator travels in time), we find a longer onset-to-target duration for V_3, as can be seen in Figure 4, in line with the knowledge that the stressed vowel is longer in its acoustic duration [19, 39]. In [k, g] contexts, the onset-to-target duration of V_3 is much longer when V_2 is a stressed vowel. This means that the tongue dorsum takes more time to reach the target of V_3 [i]. In [t͡ʃ, d͡ʒ] contexts, the onset-to-target duration is also modulated by stress, but the difference is smaller as compared to velar plosives, especially for [t͡ʃ].

Table 2: *Correlation (Pearson) between peak velocity and onset-to-target displacement of* V_3 *in Tongue Dorsum y-dimension, according to stress position. Stress on* V_1/V_2 *corresponding to the relative position of [k, g, t͡ʃ, d͡ʒ] as FAR from or POST stress.*

	r **(Stress on V_1)**	***r*** **(Stress on V_2)**
[aki]	**0.71**	**0.74**
[agi]	**0.70**	**0.79**
[at͡ʃi]	0.23	0.48
[ad͡ʒ]	0.49	0.57

Table 3: *Stiffness measured as the ratio between peak velocity and onset-to-target displacement of* V_3, *according to stress position. Stress on* V_1/V_2 *corresponding to the relative position of [k, g, t͡ʃ, d͡ʒ] as FAR from or POST stress.*

Stress position and $V_2C_3V_3$	**Mean stiffness**	**s.d.**
V_1 [aki]	**18**	3
V_2 [aki]	**14**	2
V_1 [agi]	**19**	4
V_2 [agi]	**15**	2
V_1 [at͡ʃi]	15	7
V_2 [at͡ʃi]	14	4
V_1 [ad͡ʒi]	15	6
V_2 [ad͡ʒi]	13	4

For the *peak velocity* (i.e., how fast the articulator travels), we see that generally velar plosives have higher peak velocity than palatals. This is because the tongue dorsum is involved in a velar but not in a palatal consonant. When lexical stress is on V_2 (i.e., when the displacement is larger), peak velocity is also higher. This correlation is further confirmed by Pearson's correlation tests, as shown in Table 2. We can see that peak velocity is positively correlated to displacement in [k, g] contexts, but the correlation is much weaker in [t͡ʃ, d͡ʒ] contexts.

For *stiffness* (i.e., the relative speed of a movement in relation to displacement), we can see from Table 3 that when stress is on V_1, stiffness is higher; when stress is on V_2, stiffness is much lower. This difference is even more pronounced when C_3 is velar. Based on stiffness and the correlation reported in Table 2, we can conclude that a POST stress V_3 [i] target is achieved later than a FAR from stress one.

4. Discussion and conclusions

In this study, we investigated the relationship between lexical stress and velar palatalization in Italian based on both acoustic and articulatory data. We aimed to understand how a stressed vowel "blocks" the following velar consonant from palatalizing.

In the acoustic domain, we found that when [k] was directly preceded by a stressed vowel, its closure duration was longer compared to when it was preceded by an unstressed vowel. This difference shows that the stressed vowel had an impact on the acoustic shape of the following velar plosive [k].

In the articulatory domain, this difference in closure duration can be seen as being caused by a delayed target achievement in V_3. In summary, lexical stress on V_2 has an impact on the articulation of not only the stressed vowel, but also the following segments [k, g] and [i]. The release phase of a stressed V_2 has longer, larger, faster, and stiffer gestures than their unstressed counterparts, confirming observations based on Greek [22]. The displacement of the tongue dorsum from this stressed V_2 [a] to the following [i] (i.e., onset-to-target of [i]) was larger and the time needed to reach the target of [i] was longer. Although velocity increases in POST-stress, this increase is not large enough to compensate for the longer duration in this position, as reflected by the decreased stiffness. The V_3 [i] target was therefore delayed compared to when the stress was FAR from V_3.

These acoustic and articulatory results can be interpreted within the μ-gesture model. When considering lexical stress as activated by a symmetrical spatio-temporal μ-gesture, the tongue dorsum gesture of the onset of V_3 [i] overlaps largely with the release of μ-conditioned [a]. In other words, the onset-to-target phase of V_3 [i] is within the activation interval of a μ-conditioned [a]. Inevitably, V_3 [i] is also shaped temporarily and spatially by the μ-gesture. The direct consequence is a longer and larger tongue dorsum gesture, delaying the target achievement of [i].

When the syllable C_3V_3 is [ki, gi], the place of articulation of [k, g] is fronted by [i] and both recruit the tongue dorsum. In other words, delaying the target achievement of [i] also delays the target achievement of [k, g]. This delayed target achievement of [k, g] is, in our interpretation of both acoustic and articulatory data, the reason why [k] has a longer closure duration when the preceding vowel is stressed. The higher closure ratio favors the perceptual recovery of a plosive category rather than an affricate category, as it decreases the perceptual saliency of the fricated release, thus resulting in the "blocking" of palatalization.

This study also contributes to the empirical understanding of the μ-gesture [21], by examining the activation scope of a μ-conditioned vowel. Previous studies on the μ-gesture focused on the formation/initiation of the μ-gesture (i.e., its left edge). They found that stressed syllables involve longer, larger, and faster gestures than their unstressed counterparts [22]. We find a similar pattern on the release of a μ-conditioned vowel (i.e., the right edge of the μ-gesture). Our findings thus confirm the symmetrical shape of the μ-gesture, and extend the effect of the μ-gesture beyond the stressed CV syllable, on the following CV syllable.

Future work is needed to fully understand the articulatory/perceptual basis of the "blocking" of palatalization. A perceptual study would confirm if a longer closure duration directly led to the recovery of a plosive (instead of a palatal affricate) category.

5. Acknowledgements

We would like to thank Sam Mitchell for his contribution in the study's earlier stage. This study was funded by ANR-JCJC DIA-SYN-PHON funding attributed to Maria Giavazzi (ANR-21-CE28-0008). This study was partially supported by EUR Frontiers in Cognition and the Idex PSL (ANR-17-EURE-0017), and Labex EFL (ANR-10-LABX-0083).

6. References

[1] R. Petrosino and A. Calabrese, "Palatalization in Romance," in *Manual of Romance Phonetics and Phonology*, C. Gabriel, R. Gess, and T. Meisenburg, Eds. Berlin, Boston, Massachusetts: De Gruyter, 2022, pp. 173–214.

[2] G. Lepschy, G. Rohlfs, S. Persichino, T. Franceschi, and M. C. Fancelli, "Grammatica storica della lingua italiana e dei Suoi Dialetti," *The Modern Language Review*, vol. 67, no. 3, p. 660, 1972.

[3] C. Celata and P. M. Bertinetto, "Lexical Access in Italian: Words with and without Palatalization," *Lingue e linguaggio*, vol. 2, pp. 293–318, 2005.

[4] M. Giavazzi, "On the application of velar palatalization in Italian," in *17th Manchester Phonology Meeting* , 2009.

[5] M. Giavazzi, "The phonetics of metrical prominence and its consequences on segmental phonology," thesis, Massachusetts Institute of Technology, 2010.

[6] J. Ohala, "What's cognitive, what's not, in sound change," in *Diachrony within synchrony: Language History and Cognition: Papers from the International Symposium at the University of Duisburg*, 26-28 March 1990, M. D. Morrissey and Kellermann Günter, Eds. Frankfurt am Main: Peter Lang, 1992, pp. 309–355.

[7] S. Guion, "The role of perception in the sound change of velar palatalization," *Phonetica*, vol. 55, no. 1-2, pp. 18–52, 1998.

[8] J. N. Beckman, "Positional faithfulness," thesis, University of Massachusetts Amherst, 1998.

[9] M. Fourakis, "Tempo, stress, and vowel reduction in American English," The *Journal of the Acoustical Society of America*, vol. 90, no. 4, pp. 1816–1827, 1991.

[10] A. Marcet, M. Fernández-López, A. Baciero, A. Sesé, and M. Perea, "What are the letters E and É in a language with vowel reduction? the case of Catalan," *Applied Psycholinguistics*, vol. 43, no. 1, pp. 193–210, 2021.

[11] M. A. Nobre and F. Ingemann, "Oral vowel reduction in Brazilian Portuguese," *Kansas Working Papers in Linguistics*, 1983.

[12] B. Hollenbach, "Phonetic vs. phonemic correspondence in two Trique dialects," in *Studies in Otomanguean Phonolog*, W. Merrifield, Ed. Dallas: SIL and University of Texas at Arlington, pp. 35–67.

[13] K. Suomi and R. Ylitalo, "On durational correlates of word stress in Finnish," *Journal of Phonetics*, vol. 32, no. 1, pp. 35–63, 2004.

[14] L. E. Armstrong, *The phonetic structure of Somali*. Farnborough, Hampshire: Gregg, 1969.

[15] A. E. Turk and J. R. Sawusch, "*The domain of accentual lengthening in American English*," Journal of Phonetics, vol. 25, no. 1, pp. 25–41, 1997.

[16] González Carolina, "The effect of stress and foot structure on consonantal processes," thesis, USC, 2003.

[17] Samāra Yādallāh, *The arrangement of segmental phonemes in Farsi*. 1977.

[18] W. Bauer, *Maori*. London u.a.: Routledge, 1993.

[19] M. Gordon and T. Roettger, "Acoustic correlates of word stress: A cross-linguistic survey," *Linguistics Vanguard*, vol. 3, no. 1, 2017.

[20] D. Byrd and E. Saltzman, "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," *Journal of Phonetics*, vol. 31, no. 2, pp. 149–180, 2003.

[21] Saltzman, E. et al. (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.

[22] A. Katsika and K. Tsai, "The supralaryngeal articulation of stress and accent in greek," Journal of Phonetics, vol. 88, p. 101085, 2021.

[23] D. Byrd and J. Krivokapić, "Cracking prosody in articulatory phonology," *Annual Review of Linguistics*, vol. 7, no. 1, pp. 31–53, 2021.

[24] Audacity Team. *Audacity(R): Free Audio Editor and Recorder.* (2022). Accessed: December 25th 2021. [Online]. Available: https://audacityteam.org/.

[25] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal forced Aligner: Trainable text-speech alignment using Kaldi," *Interspeech 2017*, 2017.

[26] G. Van Rossum and F. L. Drake Jr, *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[27] P. Buech, S. Roessig, L. Pagel, D. Muecke, and A. Hermes, "Ema2wav: Doing articulation by Praat," *Interspeech 2022*, 2022.

[28] P. Boersma and D. Weenink, "Praat: doing Phonetics by Computer," v. 6.1.16, 2022, https://www.fon.hum.uva.nl/praat/

[29] D. Bates, M. Machler, B. Bolker, and S. Walker, "Fitting linear ¨ mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[30] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[31] C. P. Browman and L. M. Goldstein, "Towards an articulatory phonology," *Phonology Yearbook*, vol. 3, pp. 219–252, 1986.

[32] C. P. Browman and L. Goldstein, "Gestural specification using dynamically-defined articulatory structures," *Journal of Phonetics*, vol. 18, no. 3, pp. 299–320, 1990.

[33] P. Buech, R. Ridouane, and A. Hermes, "Pharyngealization in Amazigh: Acoustic and articulatory marking over time," *Interspeech 2022*, 2022.

[34] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *Journal of the Royal Statistical Society (B)*, vol. 73, no. 1, pp. 3–36, 2011.

[35] R Core Team,R: *A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[36] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. Franc¸ois, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Muller, ¨ J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019.

[37] S. Coretta, Tidy Model Visualisation for Generalised Additive Models, 2022, r package version 3.3.2.

[38] K. G. Munhall, D. J. Ostry, and A. Parush, "Characteristics of velocity profiles of speech movements.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 11, no. 4, pp. 457–474, 1985.

[39] F. Albano Leoni, F. Cutugno, and R. Savy, "The vowel system of italian connected speech," in *Proceedings of the International Congress of Phonetic Sciences*, 1995, pp. 396–399.