STOP DEVOICING AND PLACE OF ARTICULATION: A CROSS-LINGUISTIC STUDY USING LARGE-SCALE CORPORA

Anisia Popescu[1], Mathilde Hutin[1], Ioana Vasilescu[1], Lori Lamel[1], Martine Adda-Decker[1,2]

Université Paris Saclay, LISN, CNRS; Université Sorbonne-Nouvelle, CNRS, LPP, Paris
anisia.popescu@universite-paris-saclay.fr

ABSTRACT

Voicing in stops is relatively hard to maintain from an articulatory and aerodynamic standpoint. Consonants with posterior constrictions (velars) are more prone to devoicing than consonants with more anterior constrictions (labials, coronals). Aerodynamic accounts have linked this pattern to the compliant tissue surface available for passive vocal tract expansion. This study investigates whether place of articulation plays a role in stop devoicing in five Romance languages using large corpora and automatic alignment with pronunciation variants. Results show that voicing maintenance patterns do not support an aerodynamic hypothesis: In Romance languages velars do not devoice at higher rates than labials and coronals. This suggests other elements than articulation naturalness are at play.

Keywords: large corpora, voicing maintenance, automatic alignment, pronunciation variants, Romance languages

1. INTRODUCTION

Maintaining voicing requires sufficient transglottal airflow. In stops, air accumulates in the oral cavity, reducing the oral and subglottal pressure gradient which may lead to insufficient transglottal airflow and result in voicing loss. This is known as the Aerodynamic Voicing Constraint (AVC) [1, 2]. Certain places of articulation are more compatible with voicing than others: It has been shown that voiced stops are more marked the more posterior the constriction location is [3] (i.e., voiced labials and coronals are more frequent than voiced velars, making /g/ a frequent phonological gap [4, 5]). Initially this pattern was attributed to oral cavity volume [6]: Pressure builds up faster in a posterior constriction location leading to faster voicing loss in velar stops.

Several studies [7, 8] have since shown that the difference in cavity volume alone has a negligible effect on pressure differences, and that the pattern is more likely related to the compliant surface available for passive enlargement of the vocal tract: Stops with more anterior places of articulation have more soft tissue available for increasing the vocal tract volume (i.e., lowering the larynx, elevating the soft palate, advancing the tongue root, depressing the tongue body [9, 10]).

However, most studies investigating voicing alternations as a function of place of articulation are done in controlled, small-scale laboratory studies (e.g., [11, 12]). Large scale cross-linguistic studies are scarce. A recent corpus study investigating the voicing of Danish stops in intervocalic position [13] did not find the predicted pattern supporting an aerodynamic account: The highest rate of voicing was found for coronals, not velars. The present study aims to test whether velar stops are more prone to devoicing than their labial and coronal counterparts in the European variants of five Romance languages (French, Italian, Portuguese, Romanian and Spanish), using large corpora and automatic alignment with pronunciation variants.

2. METHODOLOGY

2.1. Corpora and devoiced stop detection

The corpus consists of more than 1000 hours of French, Italian, Portuguese, Romanian and Spanish journalistic speech from TV and radio shows.These corpora were acquired from the LDC and ELRA, or developed during international projects: IST ALERT for part of the data in French and Portuguese [14, 15]; IRST for part of the data in Italian [16, 17] and OSEO Quaero [18, 19] for all languages. Only European shows were kept in the final corpus. Manual transcription was available for all data, except for Romanian, which only had 7h of manually transcribed data; The reminder of the Romanian data was automatically transcribed using a speech-to-text system [20].

An automatic speech recognition (ASR) system [21] was used to carry out the forced alignment, matching speech segments to their orthographic transcription using language specific acoustic models and pronunciation dictionaries. A total of 486,366 bilabial, 1,557,644 coronal and 283,328 velar stops were labeled by the system across the five languages. Table 1 shows the counts of phonemically voiced stops in the corpora for each of the five languages.

	b	d	g	Total
Fre	25,274	197,543	24,731	247,548
Ita	34,289	183,571	26,877	244,737
Por	26,689	119,566	25,125	171,380
Rom	100,518	398,219	87,399	586,136
Spa	299,596	658,745	119,196	1,077,537

Table 1: Counts of phonemically voiced stops per language: French (Fre), Italian (Ita), Portuguese (Por), Romanian (Rom), Spanish (Spa).

Following [21, 22], the dictionaries were enriched with equiprobable pronunciation variants for stop voicing (e.g., the Fr. *beaucoup* /boku/ 'a lot' had four possible pronunciations: [boku], [poku], [bogu], [pogu]) allowing the system to choose the variant which best matches the language-specific acoustic models. This method has proven reliable and useful to the investigation of voicing and devoicing in several recent works (Romanian [23], French [24], Spanish [25, 26, 27], Italian [28] or several Romance languages [29, 30, 31, 32]). The rates of devoiced variants detected by the system are given in Tab. 2 and illustrated in Fig. 1.

Language	bilabial	coronal	velar
French	6.69%	6.7%	7.01%
Italian	7.58%	6.11%	8.25%
Portuguese	8.23%	15.12%	10.03%
Romanian	6.81%	8.28%	8.43%
Spanish	7.19%	8.79%	8%

Table 2: Percentage of bilabial, coronal and velar devoiced stops per language.

2.2. Statistical analysis

Various phonetic and phonological factors have been shown to correlate with loss of voicing. The factors we are considering in this study are place of articulation (PoA: bilabial, coronal, velar), stop duration, language (French, Italian, Portuguese, Romanian, Spanish), position in the word (initial, medial, final), position in the syllable (onset, coda), phonotactic complexity (singleton, cluster) and voicing status of the previous and following segment (pause, voiceless, voiced). The data was

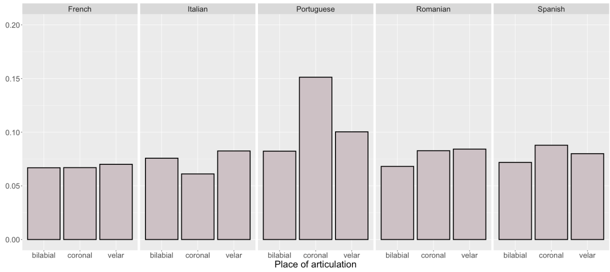

Figure 1: Percentages of devoiced stops detected in the corpora as a function of Language (from left to right: French, Italian, Portuguese, Romanian, Spanish) and Place of Articulation (from left to right: bilabial, coronal, velar).

modeled using a logistic regression and was largely inspired by the analysis in [13]. The duration variable was log-transformed in order to reach a near normal distribution. Categorical variables were contrast coded as follows: A sum contrast was coded for binary variables (phonotactic complexity, syllable position), theoretically motivated Helmert contrasts were coded for the three-level variables (see below) and deviation coding was used for the five-level Language variable (i.e., each language was compared with the overall mean of the Language variable). Each term and their interaction with place of articulation were added as factors to the model. Since speaker information (speaker ID) was not available for all data, random factors were not included for the present analysis. Below we describe the theoretically motivated Helmert contrast coding:

1. PLACE OF ARTICULATION
   - velar contrast ∼ -1/3 bilabials, -1/3 coronals, 2/3 velars
   - bilabials *vs* coronals: -1/2 bilabials, 1/2 coronals
2. POSITION IN WORD
   - final contrast ∼ -1/3 initial, -1/3 medial, 2/3 final
   - initial *vs* medial: -1/2 initial, 1/2 medial
3. VOICING ENVIRONMENT
   - voicing contrast ∼ -1/3 pause, -1/3 voiceless, 2/3 voiced
   - pause *vs* voiceless: -1/2 bilabials, 1/2 coronals

2.3. Predictions

Based on previous literature, we make the following predictions for each of our model variables:

- PLACE OF ARTICULATION: higher rates of devoicing in velar than in labial and coronal consonants (velar > coronal > bilabial) [7]
- STOP DURATION: higher rates of devoicing in

longer stops [32]

- LANGUAGE: no specific predictions
- POSITION IN WORD: higher rates of devoicing in final position followed by initial and medial positions (final > initial > medial) [29]
- SYLLABLE POSITION: higher rates of devoicing in coda position (coda > onset) [33]
- PHONOTACTIC COMPLEXITY: higher rates of devoicing in clusters than in singletons (clusters > singletons)
- VOICING ENVIRONMENT: higher rates of devoicing in pause or voiceless contexts (pause > voiceless > voiced) [32]

3. RESULTS

Results will be presented in two stages. First we report on overall tendencies for all variables (i.e., all five languages pooled together). We then present language-specific devoicing differences related to place of articulation.

3.1. Overall trends

- PLACE OF ARTICULATION
  - velar contrast ∼ velars do not devoice at higher rates than bilabials and coronals (opposite trend: Est.: -0.654994, t ∼ -10.08, p<0.0001).
  - labials *vs* coronals: coronals devoice at higher rates than labials (1.116872, t ∼ 18.923, p<0.0001).
- STOP DURATION: higher rates of devoicing in longer stops (Est. -0.371034, t ∼ -44.202, p<0.0001).
- POSITION IN WORD: higher rates of devoicing in final position compared to initial and medial positions (Est.: 0.503506, t ∼ 22.936, p<0.0001) and higher rates of devoicing in intial than in medial position (Est. -0.399143, t ∼ -49.466, p<0.0001).
- POSITION IN SYLLABLE: higher rates of devoicing in coda position (Est.: 1.271676, t ∼ 69.280, p<0.0001).
- PHONOTACTIC COMPLEXITY: higher rates of devoicing in clusters than in singletons (Est.: 0.147639, t ∼ 16.994, p<0.0001).
- VOICING ENVIRONMENT: lower rates of devoicing if the previous or following segment is voiced (previous: Est ∼ -1.44, t ∼ -144.634, p<0.0001; following: Est.: -0.993, t ∼ -40.937, p<0.0001) and lower rates of devoicing if the previous or following segment is a pause when contrasted with a voiceless stop (previous: -0.461502, t ∼ -28.185, p<0.0001; following: p>0.1).

The reported patterns are given in Tab. 3 and illustrated in Fig. 2, which show the rates of devoiced stops detected by the system for each of our categorical variables (word position, syllable position, phonotactic complexity, voicing status in the previous and following segments) for all five languages.

Variable	Levels	%
Word position	initial	9%
	medial	5.9%
	final	35.9%
Syllable position	onset	7.22%
	coda	28.9%
Complexity	singleton	7.9%
	cluster	10.7%
Previous segment	pause	21.6%
	voiceless	19.5%
	voiced	6.8%
Following segment	pause	40.8%
	voiceless	28.8%
	voiced	7.5%

Table 3: Rates of devoiced stops detected in the corpora for each categorical variable in the model.

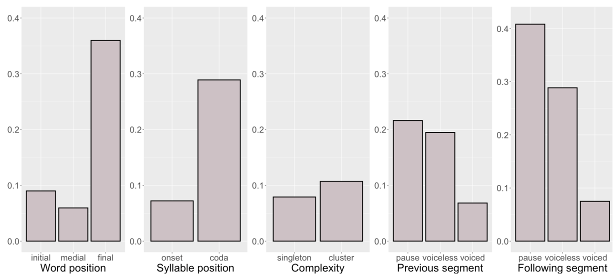

Figure 2: Rates of devoiced stops detected in the corpora for each categorical variable in the model.

3.2. Language-specific patterns

When modeling the velar and labial *vs* coronal contrasts per language (comparing each language with the overall mean of the LANGUAGE variable), we find that the velar contrast (velars compared to bilabials and coronals) is not significant for Italian, Spanish and French (p>0.1). Romanian and Portuguese pattern with the overall direction (velars tend to devoice less than bilabials and coronals), the effect being highest for Portuguese (Por: Est.: -0.106281, t ∼ -4.842, p<0.0001; Rom: Est.: -0.037202, t ∼ -2.204, p<0.05). When looking at contrasts between bilabials and coronals, we find

that there is no significant difference in devoicing rates for Romanian and French ($p>0.1$). Portuguese has significantly higher rates of coronal than bilabial devoicing (Est.: 0.540493, $t \sim 25.003$, $p<0.0001$). Spanish and Italian both show patterns that support aerodynamic accounts of voicing maintenance: Bilabials tend to devoice more than coronals in both languages, when compared to the overall mean (Ita: Est.: -0.417066, $t \sim -20.304$, $p<0.0001$; Spa: Est.: -0.025293 $t \sim -2.090$, $p<0.05$).
To sum up, velars do not devoice at higher rates than bilabial and coronal consonants: No significant effect was found for Italian, Spanish and French and the opposite trends (i.e., velars devoice less than bilabials and coronals) was found for Portuguese and Romanian. The labial *vs* coronal contrast seems to be more language-specific: In Italian and Spanish, bilabials devoice at higher rates than coronals, Portuguese shows the opposite trend (coronals > bilabials) and Romanian and French show no difference between bialabial and coronal devoicing rates when compared to the overall mean.

4. DISCUSSION

The present study investigates whether there are any differences in stop voicing maintenance based on place of articulation in the European varieties of five Romance languages. More than 1000 hours of journalistic speech were automatically aligned using pronunciation variants, which allow the system to automatically detect devoiced stops based on language-specific automatic speech recognition (ASR) systems. By analyzing the effect of place of articulation along with other phonetic and phonological factors, we find that our predictions based on aerodynamic accounts of voicing maintenance are not confirmed: Stops with more posterior constriction locations (velars) are not more prone to voicing loss. Instead, in line with previous studies [13], we find that coronals tend to devoice at higher rates (a trend that is mainly driven by the high coronal devoicing in Portuguese). Several explanations can be put forward. First, coronals tend to be more frequent than bilabials and velars, and articulatory reduction phenomena (e.g., glottis spreading) tends to occur in more frequent environments. This is also the case for our data (there are twice as many coronal stops than velars and bilabials). Second, results might be influenced by the high rate of coronal devoicing found for Portuguese. Portuguese has been shown to behave differently than other Romance languages when it comes to voicing patterns [11, 34]. Another explanation could be that in languages which require actual vocal fold vibration to cue the voicing-contrast, such as Romance languages (with the possible exception of Portuguese), speakers might adapt their articulatory strategies to circumvent the AVC and maintain the voicing contrast, as suggested by [35]. In other words, the articulatory goal of vocal fold vibration for voiced stops [36] is achieved by compensating for the aerodynamic constraints.

Finally, there is also the issue of methodology. The present study relies on language dictionaries with pronunciation variants to detect devoiced stops. When selecting the appropriate variant the ASR system relies on acoustic features, which aren't accessible to us. Our analysis of the variant selection is mostly phonological. To thoroughly investigate the relationship between place of articulation and voicing maintenance, as well as language-specific differences within a single language family, large-scale acoustic studies of the data are planned. Acoustic analyses (e.g., voicing probability throughout the stop) would give us a clearer picture of the stop devoicing phenomenon. Furthermore, the present analysis should be extended to languages that, similarly to Romance languages, need vocal fold vibration during the stop's closure (Polish, Bulgarian, Catalan, Dutch) as well as to languages that do not rely on voicing during the stop's closure (German, English, Mandarin).

5. REFERENCES

[1] J. Ohala, "The origin of sound patterns in vocal tract constraints," in *The production of speech*, P. M. (Eds.), Ed. New York: Springer Verlag, 1983, pp. 189–216.

[2] ——, "Accommodation to the aerodynamic voicing constraint and its phonological relevance," in *Proc. of the 17th International Congress of Phonetic Sciences*, Hong Kong, 2011, pp. 64–67.

[3] T. Gamkrelidze, "On the correlation of stops and fricatives in a phonological system," *Lingua*, vol. 35, 1975.

[4] D. Nikolaev, "A typology of consonant inventory gaps," *Linguistic Typology*, vol. 26, 2022.

[5] D. Sherman, "Stop and fircative systems: A discussion of paradigmatic gaps and the question of language sampling," *Working Papers on Language Universals*, vol. 17, pp. 1–31, 1975.

[6] A. Martinet, *Economie des changements phonétiques*. Berne: Ed. A. Francke, 1955.

[7] J. Ohala and C. Riordan, "Passive vocal tract enlargement during voiced stops," in *Speech Communication papers*, J. Wolf and D. Klatt, Eds. New York: ASA, 1979, pp. 82–92.

[8] H. Javkin, “Towards a phonetic explanation for universal preferences in implosives and ejectives,” *Berk. Ling. Soc.*, vol. 3, pp. 559–565, 1977.
[9] J. Westburry, “Enlargement of the supraglottal cavity and its relation to stop consonant voicing,” *JASA*, vol. 73, pp. 1322–1336, 1983.
[10] M. Rhotenberg and J. Mahshle, *The breadth-stream dynamics of simple-released-plosive production*. Basel: Karger, 1968.
[11] D. Pape and L. Jesus, “Stop and Fricative Devoicing in European Portuguese, Italian and German,” *Lang Speech*, pp. 224–246, 2015.
[12] P. Keating, “Physiological effects on stop consonant voicing,” *UCLA Working Papers in Phonetics*, vol. 59, pp. 18–28, 1984.
[13] R. Pugaard-Rode, C. Horslund, and H. Jorgensen, “The rarity of intervocalic voicing of stops in danish spontaneous speech,” *Laboratory Phonology*, vol. 13, 2022.
[14] T. F. da Silva, J.-L. Gauvain, and L. Lamel, “Lattice-based unsupervised acoustic model training,” in *Proc. of ICASSP*, 2011, pp. 4656–4659.
[15] ——, “Interpolation of acoustic models for speech recognition,” in *Interspeech*, 2013, pp. 3347–3351.
[16] J. Després, L. Lamel, J.-L. Gauvain, B. Vieru-Dimulescu, C. Woehrling, V. B. Le, and I. Oparin, “The vocapia research asr systems for evalita 2011,” in *Lecture Notes in Computer Science*, C. Bosco and A. Mazzei, Eds. Berlin Heidelberg: Springer, 2013, vol. 7689, pp. 286–294.
[17] F. Marcello, D. Giordani, and P. Coletti, “Development and Evaluation of an Italian Broadcast News Corpus,” in *Proc. of LREC 2000*, 2000, pp. 921–924.
[18] L. Lamel, S. Courcinous, J. Despres, J.-L. Gauvain, Y. Josse, K. Kilgour, F. Kraft, L. V. Bac, H. Ney, M. Nussbaum-Thom, I. Oparin, T. Schlippe, R. Schlüter, T. Schultz, T. F. D. Silva, S. Stüker, M. Sundermeyer, B. Vieru, N. T. Vu, A. Waibel, and C. Woehrling, “Speech Recognition for Machine Translation in Quaero,” in *Proc. of IWSLT 2011, San Francisco, USA*, 2011.
[19] L. Lamel, “Multilingual speech processing activities in Quaero: application to multimedia search in unstructured data,” in *International Conference on Human Language Technologies - The Baltic Perspective*. Tartu, Estonia: IOS Press, 2012.
[20] I. Vasilescu, B. Vieru, and L. Lamel, “Exploring pronunciation variants for Romanian speech-to-text transcription,” in *Proc. 4th International Workshop on Spoken Language Technologies for Under-resourced Languages*, St. Petersburg, 2014, pp. 203–206.
[21] L. Lamel and G. Adda, “On designing pronunciation lexicons for large vocabulary continuous speech recognition,” in *Proc. 4th International Conference on Spoken Language Processing*, Philadelphia, 1996.
[22] M. Adda-Decker and L. Lamel, “Pronunciation variants across system configuration, language and speaking style,” *Speech Comm.*, vol. 29, 1999.
[23] M. Hutin, A. Jatteau, I. Vasilescu, L. Lamel, and M. Adda-Decker, “Ongoing Phonologization of Word-Final Voicing Alternations in Two Romance Languages: Romanian and French,” in *Interspeech*, 2020, pp. 4138–4142.
[24] A. Jatteau, I. Vasilescu, L. Lamel, M. Adda-Decker, and N. Audibert, “’Gra[f]e!’ Word-Final Devoicing of Obstruents in Standard French: An Acoustic Study Based on Large Corpora,” in *Interspeech*, 2019, pp. 1726–1730.
[25] N. Ryant and M. Liberman, “Large-scale analysis of spanish /s/-lenition using audiobooks,” in *Proceedings of the 22d International Congress on Acoustics*, 2016.
[26] I. Vasilescu, N. Hernandez, B. Vieru, and L. Lamel, “Exploring Temporal Reduction in Dialectal Spanish: A Large-scale Study of Lenition of Voiced Stops and Coda-s,” in *Interspeech*, 2018, pp. 2728–2732.
[27] J. Hualde and P. Prieto, “Lenition of intervocalic alveolar fricatives in catalan and spanish,” *Phonetica*, vol. 71, pp. 109–127, 2014.
[28] J. Hualde and M. Nadeu, “Lenition and phonemic overlap in Rome Italian,” *Phonetica*, vol. 68, pp. 215–242, 2011.
[29] I. Vasilescu, Y. Wu, A. Jatteau, M. Adda-Decker, and L. Lamel, “Alternances de voisement et processus de lénition et de fortition: une étude automatisée de grands corpus en cinq langues romanes,” *Traitement Automatique des Langues*, vol. 61, no. 1, 2020.
[30] M. Hutin, Y. Wu, A. Jatteau, I. Vasilescu, L. Lamel, and M. Adda-Decker, “Synchronic Fortition in Five Romance Languages? A Large Corpus-Based Study of Word-Initial Devoicing,” in *Interspeech*, 2021, pp. 996–1000.
[31] M. Hutin, M. Adda-Decker, L. Lamel, and I. Vasilescu, “When phonetics meets morphology: Intervocalic voicing within and across words in romance languages,” *Interspeech*, 2022.
[32] Y. Wu, M. Hutin, I. Vasilescu, L. Lamel, and M. Adda-Decker, “Extracting Linguistic Knowledge from Speech: A Study of Stop Realization in 5 Romance Languages,” in *LREC 2022*, Marseille, France, Jun. 2022.
[33] P. Ségéral and T. Scheer, “Positional factors in lenition and fortition,” *Lenition & fortition, ed. J. Brandao de Carvalho, T. Scheer, & P. Ségéral*, pp. 131–172, 2008.
[34] M. Hutin, M. Adda-Decker, L. Lamel, and I. Vasilescu, “When phonetics meets morphology: Intervocalic voicing within and across words in romance languages,” in *Interspeech*, 2022.
[35] M.-J. Solé, “Articulatory adjustments in initial voiced stops in spanish, french and english,” *Journal of Phonetics*, vol. 66, pp. 217–241, 2018.
[36] C. Browman and L. Goldstein, “Articulatory gestures as phonological units,” *Phonology*, vol. 6, pp. 201–251, 1989.