Testing the Aerodynamic Voicing Constraint using large scale ASR with pronunciation variants: The case of five Romance languages

Anisia Popescu[(1)], Mathilde Hutin[(3)], Martine Adda-Decker[(1,2)], Lori Lamel[(1)], Ioana Vasilescu[(1)]
[(1)] LISN, CNRS, Université Paris Saclay, [(2)] LPP, CNRS, Université Sorbonne Nouvelle
[(3)] Université Catholique de Louvain

The Aerodynamic Voicing Constraint [1, 2] refers to the physical limitations speakers encounter when maintaining voicing in stops. When producing 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. Stops are not equal before the AVC. Voicing maintenance varies with place of articulation: Consonants with more posterior occlusions (velars) are less compatible with voicing than consonants with more fronted constrictions (labials and coronals). This has been attributed to differences in compliant surface availability for passive enlargement of the vocal tract [3]: for velars only the pharyngeal walls and part of the soft palate are compliant surfaces. Despite this, voiced velar consonants are present in phonological inventories across languages, an indication that certain articulatory strategies to circumvent the AVC are present [4]. The present study investigates whether non-canonical voicing patterns (voicing of phonemically voiceless and devoicing of phonemically voiced stops) vary as a function of place of articulation in intervocalic position, a position known to favoring voicing, in five Romance languages using large corpora and automatic alignment with pronunciation variants.

Corpora consisted of more than 1000 hours of French, Italian, Portuguese, Romanian and Spanish broadcast speech from TV and radio shows. Except for Romanian, which had automatically transcribed data, manual transcription was available for all data. An automatic speech recognition (ASR) system [5] was used to carry out the forced alignment, matching speech segments to their orthographic transcription using language specific acoustic models and pronunciation dictionaries. The latter were enriched with pronunciation variants for stop voicing (e.g., the Romanian *acut* /akut/ 'acute' has four possible pronunciations: [akut], [agut], [agud], [akud]) allowing the system to choose the variant which best matches the language-specific acoustic models [6]. Only intervocalic stop consonants were retained for the present analysis. Table 1 shows the total counts of phonemic intervocalic stops and the percentages of non-canonical productions identified by the system for each language.

Hypothesis: Based solely on aerodynamic constraints (AVC), we would expect phonemically voiceless velar stops to voice at lower rates than more fronted occlusion consonants (bilabials and coronals), and phonemically voiced velar stops to devoice at higher rates than their bilabial and coronal counterparts.

Statistical analysis: To test this hypothesis we ran two models (one for intervocalic voicing /p,t,k/ → [b,d,g], and one for intervocalic devoicing /b,d,g/ → [p,t,k]) using a binomial family, logit link function logistic regression with *place of articulation* (bilabial, coronal, velar), *duration, position in the word* (initial, medial, final), *language* (French, Italian, Portuguese, Romanian, Spanish) and *vowel height* (high, mid, low) as predictors. The duration variable was log-transformed to reach a normal-like distribution. Categorical variables were contrast coded using theoretically motived Helmert contrasts for the three level variables and deviation-coding was used for the five-level *language* variable.

Results show that for intervocalic *voicing* the AVC was upheld in Romanian, Portuguese and French, but not Italian and Spanish, where voicing patterns went against the predictions of the AVC (velars tend to voice at higher rates in intervocalic position than bilabials and coronals).

For intervocalic *devoicing* the AVC was disconfirmed for all five languages (velars do not devoice at higher rates than more fronted occlusion consonants).
Discussion: Results show that intervocalic stop voicing, a common phonological process, is sensitive to aerodynamic constraints *only* in Romanian, Portuguese and French. The aerodynamic hypothesis is not supported for Italian and Spanish, suggesting other elements than articulation naturalness are at play in these two languages. All studied languages behave similarly, and against the AVC in regards to intervocalic stop devoicing, a much rarer phenomenon.

	bilabial		coronal		velar	
Phoneme	b	p	d	t	g	k
French	17868 (4.6%)	63147 (9.1%)	105748 (4.5%)	83646 (7.7%)	8996 (3.6%)	63063 (9.5%)
Italian	17304 (6.6%)	54975 (10%)	116202 (4.7%)	82231 (11%)	8239 (5.6%)	69169 (15.3%)
Portuguese	12497 (6.59%)	28125 (8.2%)	67330 (13.2%)	55405 (12.9%)	11812 (9.6%)	43109 (9.69%)
Romanian	40938 (3.9%)	113007 (5.65%)	195629 (4.8%)	204176 (5.6%)	20829 (5.9%)	149215 (5.35%)
Spanish	153644 (3.7%)	123545 (7.54%)	347658 (4.5%)	142157 (8.18%)	52331 4.3%)	203569 (13.2%)

Table 1: Total counts of identified phonemic intervocalic stops per place of articulation and language. Percentages in brackets indicate counts of non-canonical productions (voiced /p,t,k/ and devoiced /b,d,g/).

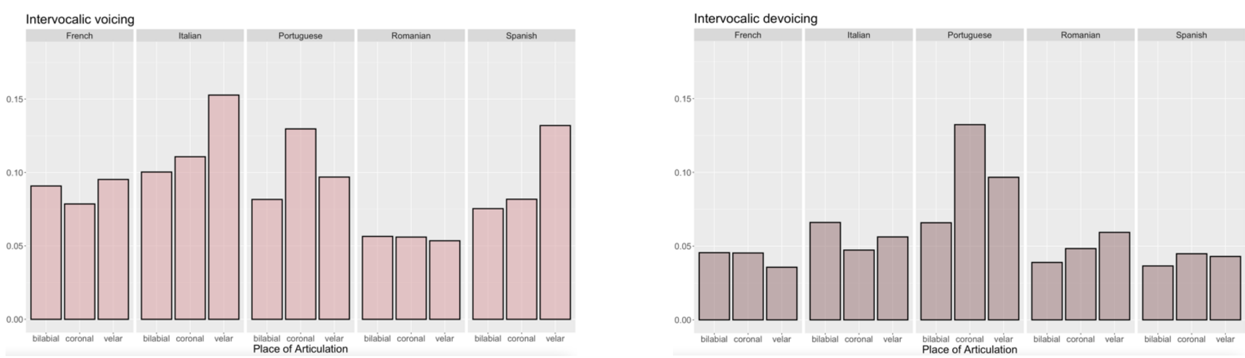

Figure 1: Intervocalic **voicing** (left) and **devoicing** (right) percentages per *Language* (French, Italian, Portuguese, Romanian, Spanish) and *Place of Articulation* (bilabial, coronal, velar) detected by the ASR systems

Selected references

[1] Ohala, J., The origin of sound patterns in vocal tract constraints, *The production of speech*, P. M. (Eds.), Ed. New York: Springer Verlag, 1983, pp. 189–216. [2] Ohala, J., Accommodation to the aerodynamic voicing constraint and its phonological relevance, *Proc. of the 17th International Congress of Phonetic Sciences*, Hong Kong, 2011, pp. 64–67. [3] Ohala J., & Riordan, C., Passive vocal tract enlargement during voiced stops, *Speech Communication papers*, J. Wolf and D. Klatt, Eds. New York: Acoustical. [4] Solé M.-J., Articulatory adjustments in initial voiced stops in Spanish, French and English, *Journal of Phonetics*, vol. 66, pp. 217–241, 2018. [5] L. Lamel & G. Adda, "On designing pronunciation lexicons for large vocabulary continuous speech recognition, *Proc. 4th International Conference on Spoken Language Processing*, Philadelphia, 1996. [6] M. Adda-Decker & L. Lamel, Pronunciation variants across system configuration, language and speaking style, *Speech Comm.*, vol. 29, 1999.