# CROSS-DIALECTAL COMPARISON OF /st/ CLUSTERS IN SPANISH: IMPLICATIONS FOR SOUND CHANGE

Amber Galvano[a] & Nicholas Henriksen[b]

[a]University of California, Berkeley & [b]University of Michigan
amber_galvano@berkeley.edu, nhenriks@umich.edu

## ABSTRACT

This study examines the variation present in phonological /st/ sequences for speakers of three regional varieties of Spanish: North-Central Peninsular Spanish (NCPS), Buenos Aires Spanish (BAS), and Western Andalusian Spanish (WAS). The three varieties are intended to reflect related stages of sound change in phonological /st/ clusters whereby speakers may shift from /s/-retention (NCPS) to aspiration of /s/ (BAS) to post-aspiration (WAS). We acoustically inspected the spontaneous speech of 60 speakers and coded each /st/ production for one of eight possible phonetic variants. While the results confirm expected patterns of variation based on a speaker's regional origin, they also show that long VOT is a pervasive characteristic in the realization of /st/ clusters in all varieties. We contextualize the findings within proposals that consider the phonetic pathway that gives rise to both pre- and post-aspiration of phonological /st/ in Spanish.

**Keywords:** sound change, actuation, propagation, Spanish, post-aspiration

## 1. INTRODUCTION

Spanish dialects are often classified as either 'retention' or 'lenition' varieties regarding the treatment of syllable-final /s/: while retention varieties retain the alveolar fricative, lenition varieties weaken the /s/, typically through aspiration or deletion [1]. Regarding Andalusian Spanish (spoken in southern Spain), for most speakers of this variety, pre-pausal /s/ is deleted, word-final intervocalic /s/ is aspirated or optionally deleted, and preconsonantal /s/ is aspirated [2]. Notably, in Andalusia's three westernmost provinces, pre-aspiration in /s/ + voiceless stop sequences has become highly variable such that speakers are shifting to post-aspiration in which glottal frication follows the stop closure [3, 4, 5, 6]. In other words, lexemes that were historically produced with coda retention (e.g., pasta [pasta] 'pasta') or pre-aspiration ([pahta]) now surface as post-aspirated stops, sometimes transcribed as [C$^h$], as in [pat$^h$a] [5, 7].

Articulatory Phonology [8] explains the shift from pre- to post-aspiration through a single realignment of the glottal-opening and oral-closure gestures [4, 9]. One question that remains unanswered, however, concerns how actuation would have unfolded among language users of WAS; in other words, how would variation present in the ambient language have instigated sound change among WAS pre-aspirating language users? Apparent-time and comparative studies converge on the idea that the sound change likely emerged in the coronal context (i.e., /st/-words) prior to undergoing transmission to the labial and velar contexts. Ruch and Peters [5] propose a three-step pathway from pre- to post-aspiration: (i) the high intraoral air pressure from the lengthened stop closure produced in pre-aspirated [hp ht hk] would result in unusually long VOT produced as [hp$^h$ ht$^h$ hk$^h$] [7]; (ii) the long VOT of such productions, especially [ht$^h$], would become more perceptually salient for listeners [10, 11, 12]; and (iii) actuation would materialize when WAS speakers imitate (and subsequently lengthen) the perceptually salient VOT of /st/-words.

With this background, the present study is guided by three goals. The first goal is to examine speaker data from a more informal speaking style than is commonly used in research on WAS post-aspiration. While most previous studies of the sound change elicit data in a controlled setting [3, 4, 7, 9, 13], one study that analyzes WAS spontaneous data reports a wider range of acoustic variants produced in /st/-words than straightforward aspiration after the closure [14]. Access to a wider range of variants, likely more prevalent in informal speaking contexts [15, 16], will aid in understanding the envelope of variation present in /st/-words. This range of data can provide new insight into the phonetic pathway along which speakers progress throughout the change from [st] → [ht] → [t$^h$].

The second goal is to investigate the linguistic and social factors that condition the attested /st/ variation in Spanish, in order to elucidate how propagation of the sound change unfolds. Previous work

suggests that factors such as the surrounding vocalic context and word position have a significant effect on WAS post-aspiration realization, while the effect of speaker sex is less certain [7, 17].

A third goal is to test the validity of Ruch and Peters' [5] proposal by comparing data produced by speakers of three regional varieties of Spanish (NCPS, BAS, and WAS), which likely reflect three stages of the sound change itself. The NCPS data are intended to reflect the ambient-language variation that is present prior to the emergence of pre-aspiration, and the BAS data are intended to reflect the variation present among pre-aspirating speakers prior to the emergence of post-aspiration (see [18]).

## 2. METHODS

### 2.1. Participants and data collection

We collected speech data from 60 speakers (age range 18-25): 19 NCPS speakers from Salamanca, Spain (9 female & 10 male); 20 BAS speakers from Buenos Aires, Argentina (10 female & 10 male); and 21 WAS speakers from Jerez, Spain (8 female & 13 male). All speakers participated in twenty-minute sociolinguistic interviews conducted in their home cities by a native speaker of their regional variety. We collected the data using a Marantz Pmd620 digital recorder and a Shure WH20 head-mounted microphone.

### 2.2. Analysis

The inclusion criterion was that a phonemic /st/-token should be produced between vowels, either word-medially or across word boundaries, which resulted in a total of 6,334 tokens. Using the acoustic characterization from [14] as a baseline, we coded each instance of /st/ per the following classification: (i) /s/-retention [st]; (ii) /s/-retention with long VOT [st$^{h}$]; (iii) /s/-deletion [t] or pre-aspiration [ht]; (iv) pre-aspiration with long VOT [ht$^{h}$]; (v) post-aspiration [t$^{h}$]; (vi) dento-alveolar affricate [t$^{s}$]; (vii) alveolar fricative [s]; and (viii) post-alveolar affricate [tʃ]. Based on the proposal by [5] that lengthened VOT should be present among speakers for which the post-aspiration process is not instigated, we used 29 ms as the threshold for establishing tokens with "long" VOT (per [19]), which we transcribed as [t$^{h}$] throughout the study. This decision is further supported by VOT findings of singleton /t/ in Spanish showing that speakers produce mean values between 15 and 20 ms [3]. Figures 1 through 5 show examples of pre-aspiration [ht], post-aspiration [t$^{h}$], pre- and post-aspiration [ht$^{h}$], the dento-alveolar affricate [t$^{s}$], and the alveolar fricative [s], respectively. To ensure reliability in the coding, both authors initially labeled 10% of the data. For the instances in which they did not agree, the two authors established updated labeling criteria. Following this, the first author proceeded with the labeling of the remaining files and consulted with the second author in cases of ambiguity.

For statistical analysis, we divided the data by DIALECT and, due to low token counts for some of the variants (see Table 1), we limited the comparisons to five. For BAS, we compared the variants [st] vs. [t, ht]; [st] vs. [ht$^{h}$]; and [t, ht] vs. [ht$^{h}$]. For NCPS we compared [st] vs. [st$^{h}$], and for WAS we compared [t$^{h}$] vs. [t$^{s}$] (see Table 2). For each of the five comparisons, we fitted a mixed logistic regression model using the package "lme4" [20] within the statistical software R, version 4.2.2 [21]. We included SEX (male, female), WORD POSITION (word-medial, cross-word), PRECEDING VOWEL, and FOLLOWING VOWEL (/a e i o u/) as fixed main effects. All models included random-intercept effects for SPEAKER to account for within-speaker dependencies. We also included by-SPEAKER random slopes for WORD POSITION, PRECEDING VOWEL, and FOLLOWING VOWEL [22]. We computed likelihood-ratio test (LRT) statistics for all fixed effects. Here, we report the significant fixed effects based on an $\alpha = 0.05$ significance level.

## 3. RESULTS

Table 1 presents the frequency of use for each variant based on dialect. The most common variants per dialect were as follows: /s/-retention for NCPS (72.32%), pre-aspiration (or no aspiration) for BAS (62.06%), and post-aspiration for WAS (68.56%). Regarding the second-most common variant, for WAS it was the dento-alveolar affricate (12.68%), while for both NCPS and BAS approximately 25% of the data contained long VOT above 29 ms (i.e., [st$^{h}$] and [ht$^{h}$] were second-most common, respectively). Note also that productions of [s] and [tʃ] are nearly exclusive to WAS speakers.

Table 2 presents the results from the five logistic regressions (one for NCPS, three for BAS, and one for WAS). In the NCPS model, the single significant predictor was FOLLOWING VOWEL, with /i u/ favoring [st$^{h}$] the most (predicted probabilities of 32% and 42%) and /a/ the least (predicted probability of 8%). For the BAS comparison of [st] vs. [t, ht], the lenited variants were favored word-medially as opposed to across word boundaries, although the estimated probabilities were very similar (99% to 98%,

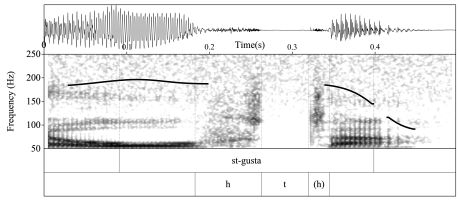


**Figure 1:** Sample spectrogram for pre-aspiration [ht] from a BAS speaker. VOT = 25 ms

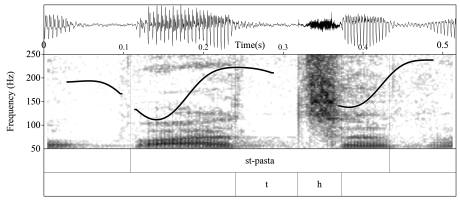


**Figure 2:** Sample spectrogram for post-aspiration [$t^h$] from a WAS speaker. VOT = 54 ms

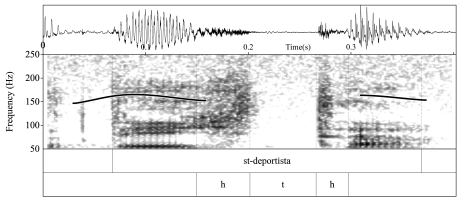


**Figure 3:** Sample spectrogram for pre- and post-aspiration [$ht^h$] from a BAS speaker. VOT = 31 ms

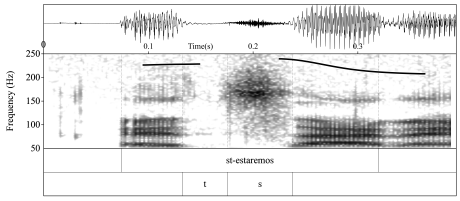


**Figure 4:** Sample spectrogram for [$t^s$] from a WAS speaker. [s] duration = 62 ms

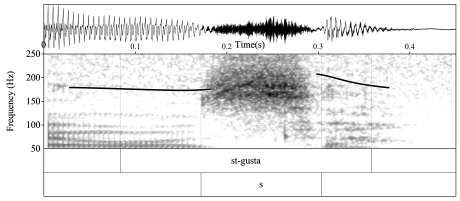


**Figure 5:** Sample spectrogram for the alveolar fricative [s] from a WAS speaker. [s] duration = 132 ms

respectively). For BAS [st] vs. [$ht^h$], the lenited variant was also favored word-medially (96%) more so than in the between-word context (72%). For the [t, ht] vs. [$ht^h$] comparison in the same dialect, following /i u/ favored [$ht^h$] (14% and 18%) compared to following /a/ (4%). Finally, in the WAS model, a following /i u/ favored [$t^h$] with a predicted probability of 98-100%, while for following /a/ probability was

| Production type | NCPS% | BAS% | WAS% |
|---|---|---|---|
| [st] | **72.32** | 5.62 | 0.05 |
| [$st^h$] | 23.22 | 0.61 | 0.35 |
| [t, ht] | 1.06 | **62.06** | 5.53 |
| [$ht^h$] | 0.16 | 26.96 | 2.41 |
| [$t^h$] | 0.05 | 3.01 | **68.56** |
| [$t^s$] | 0.00 | 0.00 | 12.68 |
| [s] | 3.18 | 1.67 | 8.15 |
| [tʃ] | 0.00 | 0.08 | 2.26 |

**Table 1:** Frequency of token types by dialect

92%. Male speakers also favored [$t^h$] (compared to [$t^s$]) with a probability of 97%, while female speakers did so at 92%.

| Model: Comparison | Predictor | *df* & LRT-value |
|---|---|---|
| NCPS: [st] vs. [$st^h$] | FV*** | 4, 30.12 |
| BAS: [st] vs. [t, ht] | WP* | 1, 4.62 |
| BAS: [st] vs. [$ht^h$] | WP* | 1, 4.76 |
| BAS: [t, ht] vs. [$ht^h$] | FV*** | 4, 29.96 |
| WAS: [$t^h$] vs. [$t^s$] | FV** | 4, 16.62 |
| WAS: [$t^h$] vs. [$t^s$] | SEX** | 1, 8.70 |

**Table 2:** Degrees of freedom and Likelihood-ratio test (LRT) value for sig. predictors in the five models. FV = following vowel; WP = word position; *** = $p < .001$; ** = $p < .01$; * = $p < .05$

## 4. DISCUSSION

Regarding this study's first goal, the spontaneous corpus yielded a richer set of variants than has been attested in the previous work on WAS post-aspiration [3, 5, 7, 9]. The more informal speech elicited in the spontaneous setting likely yielded a wider array of variants compared to methods involving controlled speech [15, 23]. The broader range of variants reflects the ways in which the sound change continues to unfold, as some WAS speakers show presumably more innovative productions than post-aspiration, such as the dento-alveolar affricate, the pre-palatal affricate, and the alveolar fricative (collectively accounting for 23.09% of the WAS corpus). Whereas Ruch [24] proposed that the shift from [$t^h$] to [$t^s$] reflects the most advanced phase of the sound change (since the [$t^s$] variant had not been described in traditional dialectological studies), the presence of the [s] variant in the WAS data suggests that some speakers may now be transitioning to a straightforward fricative. This would suggest a change such as: [$t^h$] → [$t^s$] → [s]. This proposal is supported by the fact that [s] tokens were notably more common among the WAS speakers than the two other groups

(cf. [25]).

In terms of the most frequent variant for each dialect, the NCPS speakers favored /s/-retention without long VOT, the BAS speakers favored /s/-weakening through pre-aspiration or coda deletion, and the WAS speakers favored post-aspiration of /st/. Notably, variants containing long VOT (i.e., values higher than 29 ms) surfaced in approximately 25% of each of the NCPS and BAS corpora. That speakers from "non-post-aspirating" varieties produced /st/ with long VOT suggests that lengthened VOT was likely present in the ambient language prior to the actuation of both pre- and post-aspiration. From this finding, we might infer that, prior to the actuation of WAS post-aspiration, language users were already attuned to the idea that the release of phonological /st/ is susceptible to lengthening in a fashion unlike phonological /sp/ or /sk/ (see [9]). WAS language users may have capitalized on this implicit knowledge by exaggerating the longer VOT of /st/, when post-aspiration was initially advancing in /st/-words.

Regarding the second goal of this study, the two most common significant predictors in the regression models were WORD POSITION and FOLLOWING VOWEL. Regarding the effect of WORD POSITION, in the BAS model, speakers favored lenited variants word-medially, but /s/-retention across word boundaries. This finding corroborates previous research on Spanish /s/-lenition showing that speakers are more likely to lenite preconsonontal /s/ word-medially (where there is no allomorphy) than across word boundaries (where word-final /s/ realization varies depending on the linguistic environment, e.g., [26, 27]). Concerning FOLLOWING VOWEL, this predictor was significant in three of the five models, and in each instance speakers favored /st/ variants containing long VOT (i.e., [st$^h$], [ht$^h$], or [t$^h$]) in the context of the following high vowels /i u/. In contrast, the variants lacking long VOT were most favored in the context of the following low vowel [a]. One possible explanation for this finding is that the transition of the vocal tract between [h] and [i, u] requires less articulatory effort than in the context of the low vowel [a] (see [28], p. 325-326). This is further supported by [29], which showed that VOT in English was longer before sonorants and high vowels than before mid and low vowels.

SEX emerged as a significant predictor in the WAS model, with female speakers favoring the presumably more advanced variant [t$^s$] in comparison to male speakers [25]. Labov's Principle II states that women (i.e., referring to all speakers assigned female at birth) are most often innovators in changes from below [30]. Our study aligns with Principle II in showing the female speakers favor the innovative [t$^s$] variant of the post-aspiration process.

Regarding this study's third goal, the findings corroborate the proposal from [5] that pre-aspirating language users are exposed to a notable proportion of /st/ tokens involving a long VOT (this was confirmed in the BAS data). Yet, in addition, the results for the NCPS speakers demonstrate that longer VOT is also present in /s/-retention varieties, which suggests that this acoustic characteristic was already present in the input prior to Andalusian speakers historically shifting to pre-aspiration. We might therefore infer that the higher intraoral air pressure of pre-aspirated [ht] (per [5]) was likely not the single source of lengthened VOT that could have triggered actuation of post-aspiration. It is nonetheless important to note the presence of pre-aspiration and/or lengthened VOT in /st/-words does not necessarily result in the development of post-aspiration in Spanish (see [31]).

Finally, regarding limitations of this study, future work might analyze VOT in more gradient terms (i.e., by extracting raw VOT values) for the sub-set of "long" VOT tokens identified here as /st/ productions with values higher than 29 ms (per [19]). A second limitation of this study is that, within each of the dialects, the frequency counts for numerous variants were too low for reliable statistical analysis. It is difficult to make generalizations about rarely occurring variants [32] and would thus be advisable for future studies to elicit such tokens in a more controlled setting [15]. Researchers interested in the replicability of our findings might also pursue automatic methods of data coding in addition to the impressionistic approach adopted here (which yielded eight variants)

## 5. CONCLUSION

This study analyzed the production of phonological /st/ clusters in spontaneous Spanish. The cross-dialectal results provided implications for the study of sound change, in particular, regarding the importance of analyzing data from as-yet unaffected speakers of the change in question (see [18]), in this case the NCPS and BAS participant groups. The results for the WAS speakers additionally unveiled innovative variants resulting from the change from pre- to post-aspiration, such as dento-alveolar affricates and alveolar fricatives. We proposed that, for some WAS speakers, the change may be shifting as follows: [t$^h$] $\rightarrow$ [t$^s$] $\rightarrow$ [s] (cf. [25]).

## 6. REFERENCES


[1] Carvalho, A. M. 2006. Spanish (s) aspiration as a prestige marker on the Uruguayan-Brazilian border. *Spanish in Context* 3(1), 85–114.

[2] Catalán, D. 1989. *El español: orígenes de su diversidad*. Thomson-Paraninfo.

[3] Henriksen, N., Galvano, A., Fischer, M. 2023. Sound change in western andalusian spanish: Investigation into the actuation and propagation of post-aspiration. *Journal of Phonetics* 98, 101238.

[4] Parrell, B. 2012. The role of gestural phasing in Western Andalusian Spanish aspiration. *Journal of Phonetics* 20, 37–45.

[5] Ruch, H., Peters, S. 2016. On the origin of post-aspirated stops: Production and perception of /s/ + voiceless stop sequences in Andalusian Spanish. *Laboratory Phonology* 7, 1–36.

[6] Torreira, F. 2006. Coarticulation between aspirated-s and voiceless stops in spanish: An interdialectal comparison. *Selected proc. 9th Hispanic Linguistics Symposium*. Cascadilla Press, 113–120.

[7] Ruch, H., Harrington, J. 2014. Synchronic and diachronic factors in the change from pre-aspiration to post-aspiration in Andalusian Spanish. *Journal of Phonetics* 45, 12–25.

[8] Browman, C. P., Goldstein, L. M. 1986. Towards an Articulatory Phonology. *Phonology* 3, 219–252.

[9] Torreira, F. 2012. Investigating the nature of aspirated stops in Western Andalusian Spanish. *Laboratory Phonology* 42, 49–63.

[10] Harrington, J. 2010. *Acoustic phonetics* volume 2, 81–129. Wiley-Blackwell.

[11] Ohala, J. 1983. *The origin of sound patterns in vocal tract constraints*, 189–216. Springer.

[12] Stevens, K. 1980. Discussion during symposium on phonetic universals in phonological systems and their explanation. *Proceedings of the 9th ICPhS* volume 2, 181–194.

[13] Gilbert, M. B. 2022. *An Experimental and Formal Investigation of Sevillian Spanish Metathesis*. PhD thesis New York University.

[14] Ruch, H. 2008. Affrication of /st/ in Western Andalusian Spanish: Variation and change from a sociophonetic point of view. Master's thesis Ludwig-Mazimilians-Universität München.

[15] Xu, Y. 2010. In defense of lab speech. *Journal of phonetics* 38(3), 329–336.

[16] Eckert, P., Rickford, J. R. 2001. *Style and Sociolinguistic Variation*. Cambridge University Press.

[17] Torreira, F. 2007. Pre- and postaspirated stops in Andalusian Spanish. *Segmental and Prosodic Issues in Romance Phonology*, 67–82.

[18] Carignan, C., Coretta, S., Frahm, J., Harrington, J., Hoole, P., Joseph, A., Kunay, E., Voit, D. 2021. Planting the seed for sound change: Evidence from real-time MRI of velum kinematics in German. *Language* 97(2), 333–364.

[19] Quilis, A. 1981. *Fonética acústica de la lengua española*. Gredos Madrid.

[20] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1–48.

[21] R Core Team, 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[22] Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* 68(3), 255–278.

[23] Henriksen, N. 2013. Style, prosodic variation, and the social meaning of intonation. *Journal of the International Phonetic Association* 43(2), 153–193.

[24] Ruch, H. 2010. Affrication of /st/-clusters in western andalusian spanish: variation and change from a sociophonetic point of view. *Proceedings of the Workshop "Sociophonetics, at the crossroads of speech variation, processing and communication"*.

[25] Vida-Castro, M. 2022. On competing indexicalities in southern peninsular spanish. a sociophonetic and perceptual analysis of affricate [ts] through time. *Language Variation and Change*, 1–27.

[26] File-Muriel, R. J., Brown, E. K. 2011. The gradient nature of *s*-lenition in Caleño Spanish. *Language Variation and Change* 23(2), 223–243.

[27] Horn, M. 2013. *An acoustic investigation of postaspirated stops in Seville Spanish*. PhD thesis Ohio State University.

[28] Ladefoged, P., Maddieson, I. 1996. *The Sounds of the World's Languages*. Blackwell Publishers Ltd.

[29] Klatt, D. H. 1975. Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters. *Journal of Speech and Hearing Research* 18(4), 686–706.

[30] Labov, W. 1990. The intersection of sex and social class in the course of linguistic change. *Language variation and change* 2(2), 205–254.

[31] Monroy, R., Hernández-Campoy, J. M. 2015. Murcian spanish. *Journal of the International Phonetic Association* 45(2), 229–240.

[32] Möbius, B. 2003. Rare events and closed domains: Two delicate concepts in speech synthesis. *International Journal of Speech Technology* 6(1), 57–71.