Can locus equations model dialect-specific variation in coarticulation?

Giovina Angela del Rosso
Department of Fil.Co.Spe., Roma Tre-Sapienza University, Italy
https://doi.org/10.36505/ExLing-2021/12/0017/000490

Abstract

Cross-linguistic coarticulatory differences are experimentally challenging to disentangle. This paper aims to assess whether traditional locus equations constitute a suitable heuristic tool to highlight regional differences in coarticulation patterns. For this purpose, three highly similar dialects were chosen, i.e., Bari, Cagliari, and Palermo Italian. C-V effects were examined in CV and VC sequences (V=/a/; C=/p, t, k, f, s, m, n/). A two-step acoustic analysis was performed on 648 phonetically controlled utterances, read by 24 speakers, and sampled from the CLIPS corpus. Subsequently, traditional locus equations and alternative linear regressions were compared. The results show that locus equations do not adequately model variation of subtle dynamic patterns. Yet significant dialect-specific differences in coarticulation emerged fitting alternative linear regressions.

Keywords: acoustic phonetics; locus equations; coarticulation; dialect; Italian.

Introduction

This paper aims to investigate whether locus equations represent a proper heuristic method to highlight significant cross-dialectal differences in the acoustics of coarticulation patterns among highly similar regional dialects of Italian. Coarticulation permeates speech models since its dynamics are bonded to the long-standing issues of variability and segmentation. Its patterns were held to be constrained by articulatory and aerodynamic demands, though they were later shown to be partly language-specific (Hardcastle & Hewlett 1999). Nonetheless, identifying language-specific differences is problematic, due to the flawed comparability of phonetic material across different languages. For this reason, three highly similar language varieties are examined in this study.

Experimentally, a classic method of indirectly assessing the degree of consonant-vowel (C-V) coarticulation is represented by first order locus equations (LEs, henceforth). They also capture articulatory-acoustic relationships, carrying information about places of articulation – see Perillo et al. (2015) and Bang (2017), for reviews. Indeed, LEs are regression lines fitting scatterplots of vowel formant frequencies. F2 values are measured at onset ($F2_{ON}$, y-axis) and midpoint or steady state ($F2_{MID}$, x-axis) across various vowel contexts. Their relation is formalized as $F2_{ON} = F2_{MID} * k + c$, where *k* is the slope of the regression line (i.e., the change in F2 during the transition) and *c* is the intercept between the regression line and y-axis (i.e., F2 at the beginning of the transition). Steep regression lines indicate a high degree of coarticulation, whereas flatter slopes hint at lower degrees of coarticulation. C-V effects and the steepness of LE slopes decrease from labials to alveolars, while velars exhibit vowel-dependent split patterns (Hardcastle & Hewlett 1999).

Material and method

Coarticulation patterns were compared across three highly similar regional dialects: Bari, Cagliari, and Palermo Italian. The extensive overlap of their vocabularies and phonological systems makes employing the same set of stimuli possible. The comparability is also enhanced in lexical neighbourhood density and according to the Output Constraints model (Hardcastle & Hewlett 1999).

Read speech material was extracted from the LF subcorpus of the CLIPS (corpus of Italian). Four female and four male speakers per variety were sampled. The words in the dataset meet the following criteria: they are content polysyllabic paroxytones; the Vs are stressed; the Cs are singleton; the phonetic environment was controlled; two or three repetitions per sequence were collected, depending on resources. Overall, 648 CV (/pa, ta, ka, fa, sa, ma, na/) and VC (/at, as, an/) utterances entered the dataset (27 words * 24 speakers).

The dataset was manually segmented, annotated, and the trajectories of /a/ formant frequencies were semi-automatically analysed in Praat. The data were statistically processed in R. Specifically, the analysis was divided into two stages:

LEs were fitted using F2 frequencies extracted at two time points; $F2_{ON}/F2_{OFF}$ (of CV and VC boundaries, respectively) and $F2_{MID}$; language variety, manner, and place of articulation were implemented as additive effects;

an additional time point was analysed, i.e., 25% or 75% of the vowel duration of CV and VC sequences, respectively. F1 and F3 were measured at all three time points. These supplementary data were used to fit a series of alternative linear regressions, encompassing the same additive effects as in (I).

Finally, ANOVAs were performed to compare the models and identify the best solutions, which were then checked for overfitting; to measure accuracy and validate the results, the bootstrap resampling method was applied, through 200 random resamples with replacement.

Results

The LEs plots reveal that the datapoints and regression lines are grouped by place of articulation (Figure 1). The intercepts of labials are lower than those of alveodentals, while velars have the highest intercepts. These results are in line with previous research findings relating LEs, degree of coarticulation, and place of articulation. However, if the variety variable is included, the variability sharply increases and prevents a straightforward detection of general patterns.

This is evidenced in the statistical analysis, summarised in Table 1. The ANOVAs identified M1 as the best LE-based model of stage I, accounting for 59% of the variation. However, M1 does not include language-specific differences since it relies solely on $F2_{MID}$, C place, and mode. Including the variety as an additive effect does not significantly improve the predictive power.

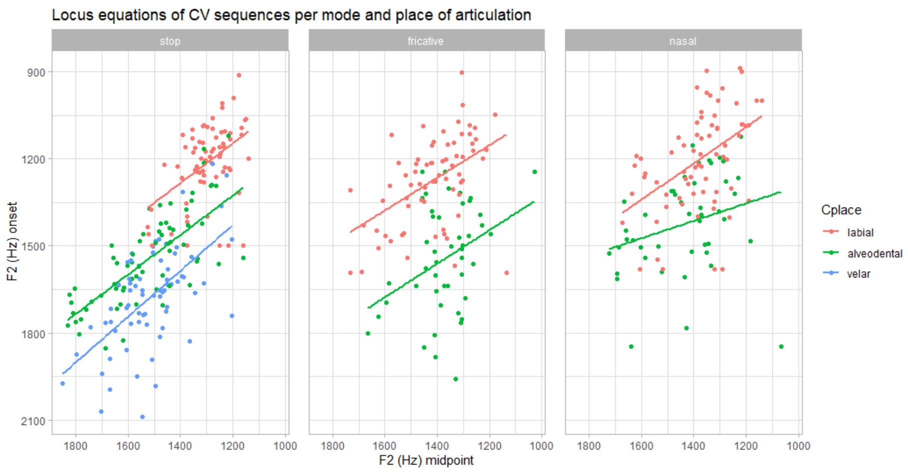

Figure 1. LEs fitting $F2_{MID}$-$F2_{ON}$ scatterplots, not divided per variety.

Table 1. Summary of the best linear models.

response		$F2_{ON}$ (CV)	$F2_{25\%}$ (CV)	$F1_{ON}$ (CV)	$F1_{OFF}$ (VC)
model		M1	M2	M3	M4
explanatory variable		$F2_{MID}$	$F2_{MID}$	$F1_{MID}$	$F1_{MID}$
		.54*** (.45, .62)	.82*** (.79, .86)	.44*** (.38, .49)	.69*** (.58, .81)
C place	labial	-222*** (-250, -194)	-80*** (-92, -69)	30*** (16, 44)	N/A
	velar	124*** (83, 165)	71*** (56, 87)	5 (-14, 24)	N/A
C mode	nasal	-82*** (-115, -48)			-74*** (-104, -45)
	stop	-23 (-55, 10)			-6 (-39, 27)
variety	CA		-19** (-31, -6)	22** (7, 38)	66*** (35, 97)
	PA		-15** (-28, -3)	24** (8, 39)	64*** (32, 95)
constant		761*** (633, 889)	282*** (228, 336)	301*** (258, 343)	77 (-15, 169)
R^2 (adj. R^2)		.59 (.58)	.85 (.85)	.34 (.34)	.42 (.41)
residual SE		159	66	82	107
F-statistic		$128_{(5, 466)}$ ***	$502_{(5, 466)}$ ***	$46_{(5, 466)}$ ***	$27_{(5, 186)}$ ***

Nevertheless, adding the variety as an effect reveals statistically significant differences in all best models fitted in stage II. In particular, M2 employs $F2_{MID}$ to predict $F2_{25\%}$ and models 85% of the variation in the data, yielding the highest accuracy score among the regressions fitted in this study. M3 and M4 are especially noteworthy because they harness the first formant of CV and VC utterances, respectively – M3 uses $F1_{MID}$ to predict $F1_{ON}$ and M4 predicts $F1_{OFF}$ through $F1_{MID}$. Finally, all models shown in Table 1 were validated after bootstrap resampling. No signs of overfitting were found; thus, no subsequent adjustments were made.

Discussion and conclusion

This study confirms that LEs successfully model C-V effects across a variety of contexts, strongly depending on the place of articulation (Hardcastle & Hewlett 1999; Perillo et al. 2015; Bang 2017). Less consistent results were obtained as language-specific effects were evaluated in highly similar dialects.

In stage I, LEs failed to detect cross-dialectal differences, but they were later shown to be significant (stage II). Using more detailed data was crucial to avoid neglecting relevant sociolinguistic variation. Unlike Öhman (1966), this study shows that not only F2 but also F1 trajectories encode information that can be exploited to model regional differences in coarticulatory effects; yet no clear pattern emerged by relying on F3. The alternative regressions mimic the capability of LEs to capture variations between two time points. However, it is not clear whether this involves any articulatory meaning.

These findings support previous research indicating that relevant phonetic information is not exclusively conveyed by few «magic moments» (Carignan et al. 2020: 2). Conversely, relying only on well-known simple metrics runs the risk of oversimplification, especially affecting time-varying phenomena.

References

Bang, H.-Y. 2017. The acoustic counterpart to coarticulation resistance and aggressiveness in locus equation metrics and vowel dispersion. Journal of the Acoustical Society of America 141(4), EL345–EL350.

Carignan, C., Hoole, P., Kunay, E., Pouplier, M., Joseph, A., Voit, D., Frahm, J., Harrington, J. 2020. Analyzing speech in both time and space. Laboratory Phonology 11(1: 2). 1–26.

Hardcastle, W. J., Hewlett, N. (eds.) 1999. Coarticulation: Theory, data and techniques. New York, Cambridge University Press.

Öhman, S. 1966. Coarticulation in VCV Utterances: Spectrographic Measurements. The Journal of the Acoustical Society of America 39(1). 151–168.

Perillo, S., Bang, H.-Y., Clayards, M. 2015. Locus equation metrics as an index of coarticulation resistance: The effect of prosodic prominence. Proceedings of Meetings on Acoustics 25.