



Perceptual lateralization of coda rhotic production in Puerto Rican Spanish

Mairym Lloréns Monteserín¹, Shrikanth Narayanan^{2,1}, Louis Goldstein¹

¹ University of Southern California, Linguistics

² University of Southern California, Electrical Engineering and Computer Science

llorensm@usc.edu, shri@sipi.usc.edu, louisgol@usc.edu

Abstract

When speakers of Puerto Rican Spanish (PRS) produce phonemic coda taps, Spanish-speakers of other dialects often perceive these as laterals. We observed production of phonemic coda laterals and taps by a male PRS speaker in real-time MRI. Temporal and spatial characteristics of tongue tip movements during coda liquid production are inconsistent with accounts positing a categorical change from rhotic to lateral in coda for this speaker. Perceptual coding of coda tap production by naïve listeners suggests that both preceding vowel type and the relative strength of a proximal prosodic boundary may impact the proportion of the subject's phonemic taps that received a lateral percept. Results are discussed in the context of persistent difficulties in modeling the gestural representation of liquid consonants.

Index Terms: phonetics and phonology, phonological processes and models, laboratory phonology, sociophonetics

1. Introduction

Puerto Rican Spanish (PRS) is a Spanish variety with a reduced consonant inventory in the syllable coda. Some scholars have argued that a liquid lateral/rhotic tap contrast present in coda in other Spanish varieties is categorically neutralized in PRS, with rhotic taps converting to laterals and laterals surfacing unchanged. These accounts predict neutralization of contrast between minimal pairs like *alma* “soul” and *arma* “weapon” [1–5].

Contemporary studies argue against categorical neutralization of coda laterals (L) and rhotics (R) on perceptual and acoustic grounds. A perception study [6] found that native PRS speakers are able to successfully discriminate between *alma/arma* pairs uttered by a PRS speaker while Argentinian speakers' performance on identical stimuli is at chance. Acoustic analysis of the third formant in *alma/arma* pairs uttered by six native PRS speakers [7] tended to show a falling trajectory for *arma* but not *alma*, a typical sign of rhoticity. These results have been taken together to suggest that the PRS liquid inventory boasts a rhotic allophone—absent in other Spanish varieties—that can receive a lateral percept due to non-native perception. A different hypothesis is that neutralization is categorical but only in casual and not careful speech. López-Morales found that PRS lateralized coda taps occur more frequently in casual than formal register and among working-class relative to upper-class speakers [8], but his results were based on transcriptions.

Putting aside the question of what PRS speakers actually produce, coda tap lateralization in PRS is perceptually apparent [9] and PRS speakers themselves characterize the phenomena qualitatively as “turning the *r* into *l*”. How does

production of a tap that is perceived as a lateral differ from production of laterals and taps whose percepts remain distinct? Production of coda liquids in PRS has not been investigated quantitatively, leaving open the question of what articulatory processes underlie the variably occurring lateral percept of coda R.

Our study is a preliminary approach to this question. Specifically, we hypothesized that at least for some speakers or in some speaking styles, the gestural representation [10] of L and R in PRS differs in coda, as it does in onset, but that coda prosody causes L-like tongue tip movements during R production (H1). H1 predicts distinct patterns of tongue tip movement durations and magnitudes as a function of coda type (L vs. R) and proximity to phrasal boundary (phrase-medial vs. phrase-final). We also hypothesized that observed effects of prosody on R tokens would correlate with their perception as laterals by naïve listeners (H2). H2 predicts a correlation between any observed L-like movement dimensions during R production and perceptual lateralization of those L-like R tokens. Given that production goals for lateral segments is not well understood, we further reasoned that observing production of phonemic taps that were subsequently perceived as laterals would allow us to identify articulatory mechanisms for “laterality”.

2. Experiments

Production of phonemic coda rhotic taps and laterals by a male PRS speaker was elicited and imaged in real-time MRI [11]. Naïve listeners subsequently coded utterances for perceptual lateralization. Lingual articulations were measured and then compared to patterns in perceptual lateralization.

2.1. Articulatory Study

This part of the study consisted of fourteen MRI scans.

2.1.1. Speaker

One male native speaker of PRS participated in this study. He reported no known speech or hearing impediments.

2.1.2. Stimuli

A list of minimal pairs differing with respect to coda liquid type—L vs. R—was constructed. One L/R pair contained word-final phonemic coda liquids preceded by the mid vowel /e/. Three pairs of words with a final “-ma” syllable elicited word-medial coda liquids, one L/R preceded by /e/ and two preceded by the low vowel /a/. All words were produced as part of sentence frames and occurred either intonational-phrase-medially or intonational-phrase-finally.

2.1.3. Data Acquisition

Imaging was performed on a 1.5 T scanner (Signa Excite HD, GE Healthcare; Waukesha, WI). A body coil was used for radio frequency (RF) transmission, and a custom eight-channel upper-airway receive coil array was used for RF signal reception. Real-time MR imaging data was acquired using a time interleaved spiral acquisition, where the angle between successive spirals is dictated by the golden angle ~ 222.49 degrees. Other important parameters include: slice thickness = 5 mm, TR = 6.004 ms., spatial resolution: 2.4mm2. Images were reconstructed by combining every four spiral interleaves, i.e., with a resultant time resolution of around 24ms. An iterative constrained image reconstruction algorithm exploiting sparse properties of the data in the temporal finite difference transform domain was employed [12]. Review [11,12] for further technical details regarding the rtMRI protocol employed in this study.

Two image planes were acquired simultaneously—mid-sagittal and coronal—but coronal slice data was not analyzed for this study.

Audio data of the speaker's utterances was simultaneously recorded from within the scanner at a sampling frequency of 100 kHz and was subsequently down-sampled to 20 kHz and noise-corrected. See [13] for further technical details regarding this aspect of the procedure.

MR scanning and audio acquisition occurred while the speaker lay supine inside the scanner and produced sentence frames that were blocked according to boundary condition.

Casual register speech was obtained as follows. The speaker was trained on a set of images representing the stimuli list. He viewed the list of images on a computer screen in random order with minimal instruction from the researcher, who only repeated the image-word correspondence twice. The speaker never saw the word list orthographically during training or the experiment. He was allowed to practice with the picture set for as long as he wished. The speaker was then instructed as follows: each 15-second scan would begin with his reading out loud a sentence frame presented visually, reflected on a small mirror on the ceiling of the scanner tube. This sentence would contain a blank space. Next, the speaker would be presented with pictures from the list he trained on. For each presented picture, he had to complete that block's sentence frame by inserting the matching word into the blank space. Because scans were blocked according to boundary condition, each scan required the speaker to use a single sentence frame.

2.1.4. Data Analysis

Tongue movements were estimated from MR videos by calculating the changes in mean pixel intensity (MPI) within regions-of-interest (ROIs) of the vocal tract. As tissue moves into the ROI, MPI increases. To minimize noise or random fluctuations in pixel intensity, signals were smoothed by a locally weighted linear regression [14]. An automatically-derived midline of the airspace in a vocal tract mean image is calculated using a dynamic programming algorithm [15]. Pixels along the midline are manually selected to serve as the centers of circular ROIs. ROIs selected were, from anterior (lips) to posterior (pharynx): labial, anterior tongue tip, posterior tongue tip, tongue body and pharynx. Only lip, tongue tip and tongue body regions were considered here.

Temporal landmarks were identified algorithmically using a velocity threshold in a manually located measurement

window on the MPI time-functions (algorithm by Mark Tiede, Haskins Laboratories). These landmarks included movement onset, time of first peak velocity, constriction onset, maximum constriction and constriction offset. Movement onset is defined as the point in time where the velocity signal first crossed the $\pm 10\%$ threshold of the first peak velocity; constriction onset is defined as the point in time when the velocity signal falls below that same threshold having surpassed it. Maximum constriction is defined as the zero-crossing point in the velocity signal; constriction offset is defined as the time of the first threshold-crossing before the second peak velocity within the measurement window.

The following tongue tip movement intervals were measured: time to peak velocity (Pvel), time to constriction onset (CloDur), time to maximum constriction (MaxDur), plateau duration (PlatDur). Durational quantities were measured in milliseconds. Spatial characteristics of movements were derived by measuring mean pixel intensity (MPI) at the time of maximum constriction in each timeseries. Higher values represent larger or more open constrictions.

Although there are inherently four prosodic positions for coda liquids (word-internal vs. word-final, phrase-medial vs. phrase-final), limitations of the current study resulted in an unbalanced list of stimuli; that is, no word-final coda liquids with a preceding /a/ were elicited. For this reason both word-final coda liquids, preceded only by /e/, and word-medial coda liquids, preceded by both /e/ and /a/, are collapsed here in the "syllable-only" (SO) boundary condition when they occur phrase-medially. The same goes for the word-medial and word-final coda liquids occurring phrase-finally in the "phrase-final" (PP) boundary condition. In order to determine if average tongue tip movement durations and magnitude differed significantly as a function of experimental conditions, three-way ANOVAs were run with Coda Type (L vs. R), Vowel Type (A vs. E) and Boundary (Phrase-final vs. Phrase-medial) as factors. P-values less than .05 were considered significant.

2.1.5. Results

There was a significant effect of coda type on the average time to peak velocity (Pvel) at the $p < .05$ level [$F(1)=7.46$, $p=0.009$]. There was also a significant effect of coda type on the duration of the movement toward closure (CloDur) at the $p < .05$ level [$F(1)=6.04$, $p=0.02$]. For both Pvel and CloDur, movements for L were longer than movements for R.

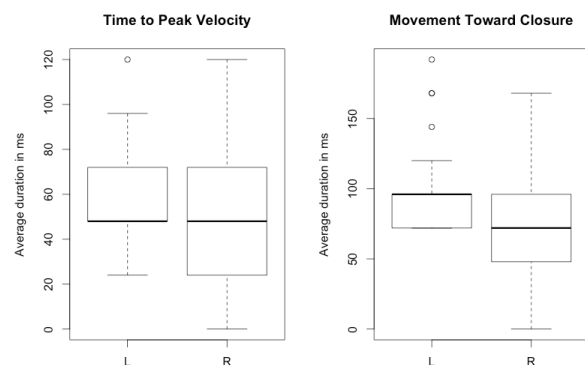


Figure 1: Effect of coda type on tongue tip movement interval durations. L=phonemic lateral, R=phonemic tap.

There was also a significant effect of preceding vowel on average tongue tip Pvel [$F(1)=17.48$, $p=0.0001$] and CloDur [$F(1)=15.47$, $p=0.0002$] at the $p<.05$ level. Movements following E were always longer than movements following A.

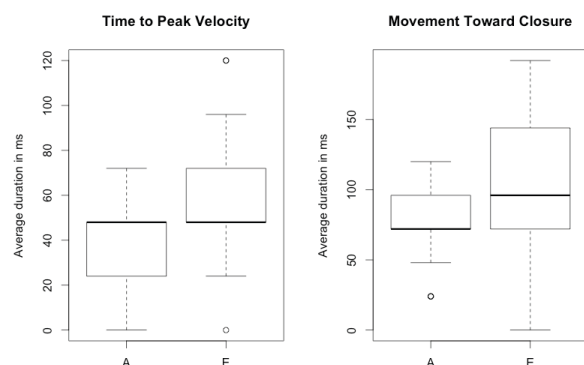


Figure 2: Effect of preceding vowel on tongue tip movement interval durations. A=low vowel /a/, R=mid vowel /e/.

There was a significant effect of boundary type on tongue tip Pvel at the $p<.05$ level [$F(1)=4.45$, $p=0.04$] but no significant effect of boundary type on CloDur. Proximity to the end of the intonational phrase was associated with longer time to peak velocity. There was also a significant interaction between boundary and vowel type on both Pvel [$F(1)=15.33$, $p=0.0003$] and CloDur [$F(1)=21.12$, $p=2.63e-05$] for the tongue tip at the $p<.05$ level. **Phrase-finally**, tongue tip movements during liquid production following /e/ were longer than those following /a/ but **phrase-medially**, movements following /a/ were longer than those following /e/.

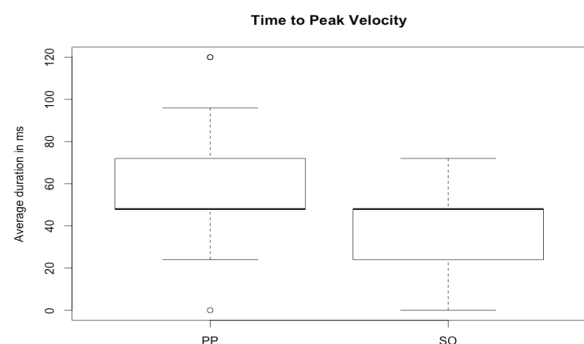


Figure 3: Effect of boundary on tongue tip time to peak velocity. PP=phrase-final, SO=phrase-medial.

There was a significant effect of coda type on movement magnitudes in the $p<.05$ level [$F(1)=4.05$, $p=0.05$]. L closures achieved tighter constrictions than R closures. Interestingly, when separated according to vowel type, average movement magnitude was distinguishable as a function of coda type following /a/ vowels but not following /e/ vowels.

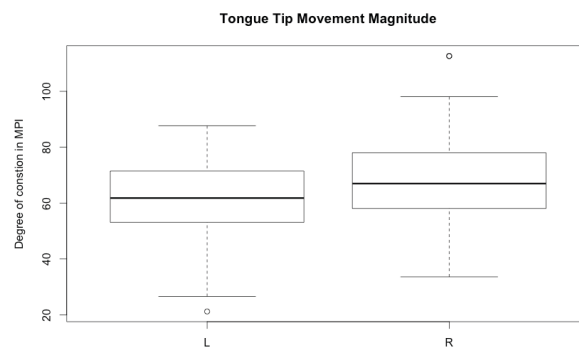


Figure 4: Effect of coda type on tongue tip movement magnitude. Smaller numbers represent tighter constrictions; scale is reversed to represent proximity to roof of mouth.

After perceptual coding (see following section), the set of coder-agreed lateralized R tokens were compared to R tokens that did not receive a lateral percept by both coders. This comparison showed a non-significant trend toward longer (i.e., more L-like) tongue tip time to constriction onset. Magnitude of these movements also showed a non-significant trend toward tighter maximum (i.e., more L-like) constrictions.

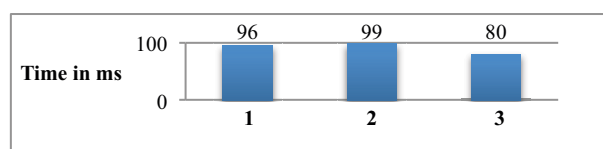


Figure 5: Trend of lateralized R tokens having longer tongue tip movement durations relative to non-lateralized R tokens. 1=average for -aL sequence. 2=average for lateralized -aR sequence. 3=average for non-lateralized -aR sequence.

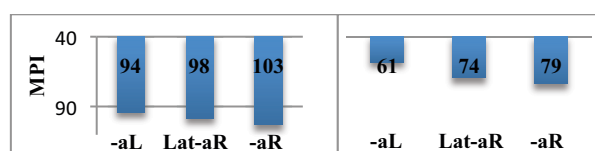


Figure 6: Trend of lateralized R tokens having L-like movement magnitudes. Only low vowel items considered. -aL=L tokens. Lat-aR=coder-agreed lateralized R tokens. -aR= non-lateralized R tokens. Left: Average tongue body maximum constriction. Right: Average tongue tip maximum constriction. Scale is identical in both graphs.

2.2. Perceptual Coding

2.2.1. Raters

Two advanced undergraduate students of phonetics recommended by faculty at a neighboring university served as naïve coders for perceptual lateralization. They were only required to be familiar with the linguistics term “lateral”. They

had no significant, sustained contact with Spanish apart from residing in the city of Los Angeles.

2.2.2. Stimuli

Recordings of utterances produced by the speaker in the articulatory portion of this study were spliced in the Praat software [16] to create a set of word-sized audio stimuli corresponding to all target words uttered with the R coda type, a total of 40 items. Target words with the L coda type were not coded because our goal was to characterize lateral percepts for taps, not percepts for phonemic laterals.

2.2.3. Procedure

Coders completed a forced choice task implemented as a Praat experiment script on a personal laptop. Target R words clipped from the speaker's utterances were presented in random order in a single block. They had one task: to click a "YES" button on the screen if they heard a lateral in the word played back to them; "NO" if otherwise. A separate button on the screen allowed them to replay each item no more than five times. They sat in a quiet room wearing a pair of Pioneer HDJ-1500 professional headphones and took around fifteen minutes to complete the task.

2.2.4. Analysis

An item-by-item analysis was first carried out to determine the rate at which coders agreed in the coding task. Only the R items that both coders perceived as lateralized were subjected to further analysis.

2.2.5. Results

Coders tended to agree more on perceptual lateralization for target words with a word-internal coda R (92%) than for words with a word-final coda R (50%).

Among the set of coder-agreed lateralized R tokens, there was an effect of preceding vowel. No target words with preceding mid-vowels were perceived as lateralized by both coders. Tokens with a preceding /a/ showed an effect of boundary. In this condition, 20% of instances of word-final coda R were perceived as lateral by both coders, whereas word-medial coda R was perceived as lateral by both coders in 100% of instances.

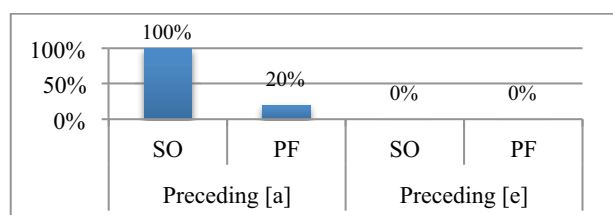


Figure 7: Rate of coder-agreed lateralization in each boundary condition by vowel. SO=syllable boundary, PF=phrase boundary.

3. Discussion

Our hypothesis that gestural representation for L and R differs in coda was supported by the finding that tongue tip interval durations and movement magnitude were distinguishable as a function of coda type. These results add further support to the notion that articulatory events resulting in perceptual lateralization do not necessarily constitute categorical

neutralization for all speakers. Since our study looked at only one speaker, our claim is not that all PRS speakers maintain distinct production patterns for L and R in coda but rather that distinct L and R production in coda can and does result in lateral percepts for R tokens by non-native listeners.

Proximity to different prosodic boundaries influenced lingual articulations, mirroring previous findings [17]: movement durations in proximity to a phrase boundary were lengthened relative to those same movements phrase-medially. Boundary strength influenced coder agreement on perceptual lateralization as well, with more agreement phrase-medially than phrase-finally. Boundary also influenced the rate of agreed perceived lateralization, with more lateralization phrase-medially than phrase-finally. However, the extent to which these results support our hypotheses is unclear due to observed effects of vowel type.

Vowel type influenced all measurements, contrary to our expectations. Though individual coders heard lateralization in some E tokens, no coder-agreed lateralization was found among these in either boundary condition. In contrast, coders agreed that **all** of phrase-medial R tokens preceded by A were lateralized, as well as some phrase-final ones. To put another way, the hypothesized effect of boundary on perceived lateralization was found but only in the context of preceding low vowels. These results would suggest that variable perceptual lateralization of coda rhotics in PRS is largely related to vocalic context, not prosody; however, the first author of this study, a native PRS speaker, reports hearing lateralization in many E tokens. Coders' perceptual expectations for vowel type and coda L may have lead them to reject lateralized R tokens.

If we put aside the effects of vowel and consider only A tokens, the trend toward L-like (i.e., longer and more constricted) tongue tip movements during R tokens that were subsequently perceived as lateralized by both coders lends partial support to our hypothesis that some prosodic factor which influences R production results in a higher rate of perceptual lateralization. Regardless, shortcomings of the current study render it unable to tease apart effects from vowel and boundary strength on both lingual articulations and perceptual coding.

4. Conclusions

This study provided articulatory support for the hypothesis that L and R have distinct representation in the PRS coda despite variable perceptual lateralization of rhotics in this context. Future work should acquire a fully balanced data set including all vowels and liquids in both onset and coda. Furthermore, coder disagreement and reduced rate of lateral percepts at phrase boundaries may have arisen due to observed velum lowering preceding the phrase-final pause. In other words, velum lowering towards rest position may have added nasality to coda R utterances, leading coders to categorize these as nasalized rather than lateralized. This and other problems already mentioned should be taken into account when designing future experiments investigating coda liquid production in PRS.

5. Acknowledgements

Funding was provided by National Institutes of Health grants T32 DC009975 and RO1 DC007124. The authors wish to thank Drs. Dani Byrd, Asterios Toutious and Sajan Goud Lingala, as well as Colin Vaz and Reed Blaylock.

6. References

- [1] J. Matluck, "Fonemas finales en el consonantismo Puertorriqueño," *NRFH*, vol. 15, pp.332-342, 1961.
- [2] J.M. Guitart, "Conservative versus radical dialects in Spanish: Implications for language instruction," *Bilingual Review*, vol. 5, no. ½, 1978.
- [3] D.L. Canfield, "Spanish Pronunciation in the Americas," Chicago: University of Chicago Press, 1981.
- [4] D. Sankoff and P. Rousseau, "Statistical evidence for rule ordering," *Lang. Var. and Change*, vol.1, no.1, pp.1-18, 1989.
- [5] J. Lipski, "El español de America," Madrid, Spain: Editorial Cátedra, 1994.
- [6] M. Paz, "Retroflexion of post-nuclear /r/ in Puerto Rican Spanish," *Proceedings of the 9th Hispanic Linguistics Symposium*, November, 2005.
- [7] M. Simonet, M. Rohena-Madrazo and M. Paz, "Preliminary evidence for incomplete neutralization of coda liquids in Puerto Rican Spanish," *Proceedings of the 3rd Annual Conference on Laboratory Approaches to Spanish Phonology*, 2008.
- [8] H. López-Morales, "Estratificación social del español en San Juan de Puerto Rico," Distrito Federal, México: Publicaciones del Centro de Ling. Hispánica, 1983.
- [9] T. Navarro Tomás, "El español en Puerto Rico," San Juan, Puerto Rico: Universidad de Puerto Rico, 1948.
- [10] C.P. Browman and L.M. Goldstein, "AP: an overview," *Phonetica*, vol.49, pp.155-180, 1992.
- [11] S. Narayanan et al., "An approach to real-time magnetic resonance imaging for speech production," *JASA*, vol.115, no.4, pp.1771-1776, 2004.
- [12] SG Lingala et al., "High spatio-temporal resolution multi-slice real time MRI of speech using golden angle spiral imaging with constrained reconstruction, parallel imaging and a novel upper airway coil," *Proceedings ISMRM 23rd Scientific Sessions*, 2015.
- [13] C. Vaz, V. Ramanarayanan, S. Narayanan, "A two.step technique for MRI audio enhancement using dictionary learning and wavelet packet analysis," *Proceedings Interspeech*, 2013.
- [14] A. Lammert et al., "Statistical methods for estimation of direct and differential kinematics of the vocal tract," *Speech Communication*, vol.55, pp.147-161, 2013.
- [15] R. Bellman, "Dynamic programming," Princeton: Princeton University Press, 1957.
- [16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," Computer Program, Version 6.0.15.
- [17] D. Byrd and E. Saltzman, "Intragestural dynamics of multiple prosodic boundaries," *J. of Phonetics*, vol.26, pp.173-199, 1998.