# Evaluation of segmentation approaches and constriction degree correlates for spirant approximant consonants

Mauricio A. Figueroa Candia, Bronwen G. Evans

Speech Hearing and Phonetic Sciences, UCL (University College London)
m.figueroa.12@ucl.ac.uk, bronwen.evans@ucl.ac.uk

## ABSTRACT

The segmentation of Spanish spirant approximant consonants poses several methodological challenges given the gradual transition from these consonants to neighbouring segments. Most studies to date employ manual segmentation approaches, although these methods are highly subjective and time-consuming. Alternative automated approaches have used the minimum intensity in the consonant and maxima from the surrounding segments as landmarks for relative intensity measurements.

Multinomial Logistic Regression (MLR) analyses were used to assess whether acoustic measurements obtained from manual and automated segmentation approaches are able to predict the level of a known categorical outcome variable (*phoneme category*). A third analysis was included to explore the predictive capabilities of data from several constriction degree correlates, based on relative intensity or spectral energy measurements.

Results showed a slight advantage for the automated segmentation method. As for constriction degree correlates, some measurements – Intensity Ratio and Intensity Difference A – were unable to fully predict the outcome categorical variable.

**Keywords**: segmentation, approximants, constriction correlates, MLR

## 1. INTRODUCTION

Spirant approximants [β̞], [ð̞] or [ɣ̞] are to be found in complementary distribution with other allophones from /b, d̪, g/ in all dialects of Spanish [1,11,15]. Although an important degree of variation is to be found between dialects (see for example [3,6]), a general property of these consonants is that they display a gradual transition to neighbouring segments, which makes segmentation difficult [18], thus hampering the extraction of acoustic data.

Despite this methodological problem, several studies to date provide acoustic measurements for these consonants, normally resorting to manual segmentation (e.g., [9,10]). While it is not always clear how manual segmentation has been executed (e.g., [4,7]), most previous studies declare using auditory and visual cues from waveforms, spectrograms, intensity and formant contours to identify the most likely location of the approximant consonant [4,6,9].

A somewhat different approach consists in the automated identification of the minimum intensity in the consonant and of maxima intensity values in the surrounding segments [4,8,13]. Several intensity measurements based on these intensity landmarks have been calculated, usually as correlates for the consonant's degree of constriction (e.g., [8,9,14]).

So far, with the exception of intensity-related measurements, there have been no attempts to extract other types of acoustic measurements from points of maximum and minimum intensity in consonants such as [β̞], [ð̞] or [ɣ̞] and their surrounding segments. Yet, if the intensity minimum in a spirant approximant is assumed as being the place where the consonant's properties are best exemplified, additional acoustic measurements can be extracted at this point, thus circumventing the subjectivity associated with manual segmentation.

Along the same lines, no comparisons of datasets originating from different segmentation methods have been carried out for spirant approximants yet. In the case of acoustic constriction degree correlates, data for intensity ratios, intensity differences and maximum intensity change velocity has been correlated to electromagnetic articulometry data for the constriction of /b/ using linear regression. Results showed that intensity ratios without low pass filtering were best correlated to the articulatory domain [14]. The effect of the preceding segment in the constriction degree for /d̪/ using intensity differences, spectral tilt and maximum intensity change velocity has also been evaluated via mixed-effects models in regression analyses. The results showed that maximum velocity was most likely the best predictor [8].

This study uses Multinomial Logistic Regression (MLR) to assess how acoustic measurements obtained through two different segmentation approaches (manual and automated) are able to predict the levels of the categorical outcome variable *phoneme category* (levels: /b/, /d̪/ and /g/) [5]. The

|  | /b/ |  |  |  | /d/ |  |  |  |
|---|---|---|---|---|---|---|---|---|
|  |  | 95% CI for odds ratio |  |  |  | 95% CI for odds ratio |  |  |
|  | Estimate (SE) | Lower (2.5%) | Odds ratio | Upper (97.5%) | Estimate (SE) | Lower (2.5%) | Odds ratio | Upper (97.5%) |
| Intercept | -2.24(0.33)*** | 0.06 | 0.11 | 0.20 | -2.22(0.33)*** | 1.00 | 1.00 | 1.00 |
| Intensity | 0.03(0.01)*** | 0.06 | 0.11 | 0.21 | 0.01(0.01) | 1.00 | 1.00 | 1.00 |
| F1 | 0.00(0.00)*** | 1.02 | 1.03 | 1.04 | 0.00(0.00)*** | 1.00 | 1.00 | 1.00 |
| F2 | 0.00(0.00)* | 1.00 | 1.01 | 1.02 | 0.00(0.00)*** | 1.00 | 1.00 | 1.00 |

**Table 1**: MLR results for the manual segmentation approach. Non significant main effects have been highlighted. Significance codes: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$. Log-Likelihood (unexplained variability): -5334.9; McFadden $R^2$ (effect size): 0.017299; Likelihood ratio test (significant variability explained by model): 187.83, $p < 0.001$.

choice for this particular outcome variable was made under the assumption that the acoustic information that distinguishes spirant approximants originating from these three phonological units is present in the measurements undertaken, provided that the segmentation method is adequate.

The same approach using MLR will be employed to assess the predictive capabilities of several acoustic constriction degree correlates, most of them relative intensity measurements.

## 2. METHODS AND PROCEDURES

### 2.1. Measurements

Acoustic measurements were extracted from a Chilean Spanish corpus containing 3 speech styles from 10 native male and female adult speakers (n = 4948: /g/ 33%, /b/ 36%, /d̪/ 31%). All data was extracted and some processes automated using Praat [2].

The extraction of acoustic parameters via manual segmentation was aided by visual cues from the waveform and spectrogram, and formant and intensity curves. These cues were used to identify likely boundaries inside each consonants' transitions to neighbouring segments. Intensity, F1 and F2 were measured as mean values from the inner 50% for each token. For the automated approach, intensity, F1 and F2 values were extracted from the points of minimum intensity inside the consonant.

Additionally, the following acoustic constriction degree correlates were calculated:

- Intensity Ratio (IntRatio): Consonant minimum intensity divided by the following segment's maximum intensity [3,14].
- Intensity Difference A (IntDiff_A): Consonant minimum intensity subtracted from the following segment's maximum intensity [8,14].
- Intensity Difference B (IntDiff_B): Consonant minimum intensity subtracted from the preceding segment's maximum intensity [12].
- Maximum Velocity (MaxVel): Maximum intensity from first differences (0.001 ms interval) between consonant intensity minimum and following segment's intensity maximum [8,9,14].

  For first differences at *n* intensity samples $I_i$:

  (1) $$\Delta I_i = I_i - I_{i-1}$$

  Maximum Velocity is defined as:

  (2) $$MaxVel = \max_{i=1,\ldots,n} \Delta I_i$$

- Minimum Velocity (MinVel): Minimum intensity from first differences (0.001 ms interval) between consonant intensity minimum and following segment's intensity maximum [9] (see Equation 3).

  (3) $$MinVel = \min_{i=1,\ldots,n} \Delta I_i$$

- Spectral Tilt (SpecTilt): Spectral energy difference between a low 50-500 Hz band and a high 500-5000 Hz band [8,17], measured in this case between the consonant intensity minimum and the following intensity maximum (see Equation 4).

  (4) $$lfb_{(50-500\,Hz)} - hfb_{(500-5000\,Hz)}$$

| | /b/ | | | | /d/ | | | |
|---|---|---|---|---|---|---|---|---|
| | | 95% CI for odds ratio | | | | 95% CI for odds ratio | | |
| | Estimate (SE) | Lower (2.5%) | Odds ratio | Upper (97.5%) | Estimate (SE) | Lower (2.5%) | Odds ratio | Upper (97.5%) |
| Intercept | -2.09(0.29)*** | 0.07 | 0.12 | 0.22 | -1.93(0.29)*** | 0.08 | 0.15 | 0.26 |
| Intensity | 0.03(0.01)*** | 1.02 | 1.03 | 1.04 | 0.02(0.01)** | 1.00 | 1.02 | 1.03 |
| F1 | 0.00(0.00)*** | 1.00 | 1.00 | 1.00 | 0.00(0.00)*** | 1.00 | 1.00 | 1.00 |
| F2 | 0.00(0.00)* | 1.00 | 1.00 | 1.00 | 0.00(0.00)*** | 1.00 | 1.00 | 1.00 |

**Table 2**: MLR results for the automated segmentation approach. Significance codes: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$. Log-Likelihood: -5348.2; McFadden $R^2$: 0.014856; Likelihood ratio test: 161.3, $p < 0.001$.

### 2.2. Analyses

Three MLR analyses were conducted in R [16] (package: *mlogit*). The reference level was defined as /g/ in order to obtain more conservative results, given that /b/ and /d̪/ have been shown to be acoustically more similar to each other [6].

The analysis for the manual approach data included intensity, F1 and F2 as main factors. The same main factors were included for the dataset generated via the automated approach. Finally, the analysis for the constriction degree correlates included IntRatio, IntDiff_A, IntDiff_B, MaxVel, MinVel and SpecTilt as main factors.

No multicollinearity was found for the analyses pertaining to the segmentation approaches, however high correlations were found for some constriction degree correlates (IntRatio against IntDiff_A: $r(4946) = -.993$; IntRatio against MaxVel: $r(4946) = -.859$; MaxVel against IntDiff_A: $r(4946) = .851$).

## 3. RESULTS

### 3.1. Manual approach

The results for the MLR analysis for the manual segmentation approach dataset show that intensity, F1 and F2 are able to predict /b/ instead of the reference level /g/ with statistical significance (see Table 1). Only F1 and F2 are able to predict /d̪/ instead of the reference level with statistical significance.

### 3.2. Automated approach

The MLR results for the dataset generated through the automated approach (see Table 2) showed that all the main factors were able to predict both /b/ and /d̪/ levels instead of the reference level, with statistical significance.

| | /b/ | | | | /d/ | | | |
|---|---|---|---|---|---|---|---|---|
| | | 95% CI for odds ratio | | | | 95% CI for odds ratio | | |
| | Estimate (SE) | Lower (2.5%) | Odds ratio | Upper (97.5%) | Estimate (SE) | Lower (2.5%) | Odds ratio | Upper (97.5%) |
| Intercept | 5.35(3.19). | 0.40 | 210.86 | 110520.9 | 4.31(3.16) | 0.15 | 74.42 | 36720.6 |
| IntRatio | -4.43(3.18) | 0.00 | 0.01 | 6.00 | -3.80(3.15) | 0.00 | 0.02 | 10.61 |
| IntDiff_A | -0.18(0.06)** | 0.75 | 0.84 | 0.93 | -0.20(0.06) | 0.73 | 0.81 | 0.91 |
| IntDiff_B | -0.10(0.01)*** | 0.89 | 0.91 | 0.93 | -0.08(0.01)*** | 0.90 | 0.92 | 0.94 |
| MaxVel | 3.06(0.43)*** | 9.24 | 21.31 | 49.17 | 4.49(0.42)*** | 38.74 | 88.83 | 203.65 |
| MinVel | 1.49(0.70)* | 1.12 | 4.43 | 17.54 | 4.24(0.79)*** | 14.89 | 69.57 | 325.06 |
| SpecTilt | -0.01(0.00)*** | 0.98 | 0.99 | 0.99 | -0.01(0.00)*** | 0.99 | 0.99 | 1.00 |

**Table 3**: MLR results for constriction degree correlates. Non significant main effects have been highlighted. Significance codes: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$. Log-Likelihood: -5268; McFadden $R^2$: 0.029626; Likelihood ratio test: 321.67, $p < 0.001$.

### 3.3. Constriction correlates

The results for the MLR analysis comparing different constriction degree correlates (see Table 3) show that IntDiff_B, MaxVel, MinVel and SpecTilt are good predictors for /b/ and /d̪/. IntDiff_A was only able to predict /b/ with statistical significance. IntRatio was unable to predict any of the levels as different from the reference level.

## 4. DISCUSSION

The datasets obtained through the two segmentation approaches show similar capabilities at predicting *phoneme category* levels. However, the main factor intensity was unable to predict /d̪/ as different from the baseline level (/g/) for the manual segmentation approach. This suggests a slight advantage for the automated approach over the alternative manual one in this regard.

Although manual approaches display some unique advantages – i.e., an actual linguistic unit is being segmented, thus being a more ecologically valid approach – they are based on a subjective analysis and, most importantly, are considerably time consuming.

While the automated approach is unable to isolate segmental units and thus is incapable of providing duration measurements for the segment, it has the advantages of being highly efficient and less error prone. As long as duration is not a variable of interest, this approach seems to be able to replace manual segmentation reliably. Moreover, the results suggest that the automated measurements seem to convey the relevant acoustic differences upholding the /b, d̪, g/ distinction for spirant approximants, despite the fact that they were taken from the consonant's point of minimum intensity only.

In contrast to previous research using articulatory data [14], IntRatio, one of the constriction degree correlates, was unable to predict the selected categorical outcome variable. It is unclear why this might be the case, but one possibility is that this constriction degree correlate is only reliable at discriminating within a phonological category (/b/ in previous accounts [14]), but that it fails to convey relevant intensity differences between said categories.

Instead, the best predictors appeared to be IntDiff_B, MaxVel, MinVel and SpecTilt, although IntDiff_A was also able to predict one level of the outcome variable. The fact that most constriction degree correlates were able to perform reliably suggests that the phonemic categories evaluated differ in some aspects of their intensity constitution, which might be indicative of further differences in how constriction is represented acoustically.

## 5. REFERENCES

[1] Barlow, J. A. 2003. The stop-spirant alternation in Spanish: Converging evidence for a fortition account. *Southwest Journal of Linguistics* 22(1), 51–86.

[2] Boersma, P., Weenink, D. 2014. *Praat: doing phonetics by computer* (version 5.4) [Computational programme]. URL: http://www.praat.org/

[3] Carrasco, P., Hualde, J. I., Simonet, M. 2012. Dialectal differences in Spanish voiced obstruent allophony: Costa Rican versus Iberian Spanish. *Phonetica* 69(3), 149–79.

[4] Colantoni, L., Marinescu, I. 2010. The scope of stop weakening in Argentine Spanish. In: Ortega-Llebaria, M (ed), *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, Austin, 100–114.

[5] Field, A., Miles, J., Field, Z. 2013. *Discovering Statistics Using R*. Great Britain: Sage.

[6] Figueroa, M., Evans, B. G. 2014. Cue weighting of acoustic variables in the perception of Chilean Spanish approximant consonants (poster). *2014 Meeting of the British Association of Academic Phoneticians*. Oxford.

[7] Hualde, J. I., Shosted, R., Scarpace, D. 2011. Acoustics and articulation of Spanish /d/ spirantization. *Proceedings of the International Congress of Phonetic Sciences XVII*, Hong Kong, 906–909.

[8] Hualde, J. I., Simonet, M., Shosted, R., Nadeu, M. 2010. Quantifying Iberian spirantization: acoustics and articulation. *Proc. LSRL 40*, Seattle, 26–28.

[9] Kingston, J. 2008. Lenition. In: Colantoni, L. and Steele, J. (eds), *Proc. 3rd Conference on Laboratory Approaches to Spanish Phonology*, Toronto, 1–31.

[10] Martínez Celdrán, E. 1984. Cantidad e intensidad en los sonidos obstruyentes del castellano: hacia una caracterización acústica de los sonidos aproximantes. *Estudios de Fonética Experimental* 1, 73–129.

[11] Martínez-Celdrán, E. 2004. Problems in the classification of approximants. *Journal of the International Phonetic Association* 34(2), 201–210.

[12] Martínez-Celdrán, E., Regueira, X. L. 2008. Spirant approximants in Galician. *Journal of the International Phonetic Association* 38(01), 51-68.

[13] Ortega-Llebaria, M. 2003. Effects of Phonetic and Inventory Constraints in the Spirantization of Intervocalic Voiced Stops: Comparing two Different Measurements of Energy Change. *Proc. 15th ICPhS* Barcelona, 2817–2820.

[14] Parrell, B. 2010. Articulation from acoustics: estimating constriction degree from the acoustic signal. *2nd Pan-American/Iberian Meeting on Acoustics*, Cancún.

[15] Parrell, B. 2011. Dynamical account of how /b, d, g/ differ from /p, t, k/ in Spanish: Evidence from labials. *Laboratory Phonology* 2(2), 423–449.

[16] R Core Team 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

[17] Sluijter, A. M. C., van Heuven, V. J. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100(4), 2471-2485.

[18] Turk, A., Nakai, S., Sugahara, M. 2006. Acoustic segment durations in prosodic research: A practical guide. *Methods in Empirical Prosody Research* 3, 1–28.