ACOUSTICS OF SPANISH AND ENGLISH CORONAL STOPS

Joseph V. Casillas, Yamile Díaz and Miquel Simonet

University of Arizona, Tucson
{jvcasill, ydiaz44, simonet}@email.arizona.edu

ABSTRACT

This study explores the acoustic correlates that distinguish coronal stops (/t/, /d/) between English and Spanish. English and Spanish coronal stops are hypothesized to differ in terms of voice-onset time and place of articulation. We are particularly concerned with capturing the place of articulation difference with acoustic data, as the voice-onset time difference is well known. Specifically, we focus on English /d/ and Spanish /t/, which are phonetically-voiceless stops with a short-lag voice-onset time. Spanish /t/ has been described as being articulated at dental place, whereas English /d/ is articulated at alveolar place. Mixed-effects models explored various spectral measurements of the consonant burst and found that standard deviation, relative burst intensity, and center of gravity differed as a function of place of articulation (or language).

Keywords: Coronal stops, Spectral moments, VOT, Spanish, English

1. INTRODUCTION

Spanish and English both have coronal stops (/d, t/); however, their phonetic implementation differs. English /d/ and /t/ are produced with an alveolar place of articulation (POA) [15]. Spanish /d/ and /t/, on the other hand, are both produced with a dental POA [9]. These descriptions rest mostly on impressionistic observations. This investigation sets out to explore the acoustic correlates related to place differences amongst these segments.

An important difference between Spanish and English has to do with their treatment of the stop voicing distinction common to these languages. Both distinguish between /t/ and /d/ by exploiting the acoustic correlate voice-onset time (VOT)—the acoustic output of the coordination of glottal and supra-glottal gestures that results in a time difference between the onset of modal voicing and articulatory release—; however, the manner in which they exploit VOT differs between the two languages. In English, /d/ has a short-lag VOT and /t/ has a long-lag VOT, whereas in Spanish /d/ has a lead VOT (prevoicing) and /t/ has a short-lag VOT [13, 20].

The question of how bilinguals who speak Spanish and English use VOT to distinguish voiced and voiceless stops in their two languages has been investigated at length [19, 20]; however, studies to date have overlooked the fact that bilinguals would need to produce a difference in POA, in addition to differently exploiting VOT, in order to avoid coalescing English and Spanish coronal stops. To date, few analyses have used acoustic measures to investigate differences in POA of coronal stops; two exceptions are [17] and [18].

The fact that both Spanish and English have short-lag, phonetically voiceless stops in their sound inventory (/t/ and /d/, respectively) begs the question as to whether these two segments can be distinguished by any acoustic measures. Accounting for POA differences via acoustic metrics opens the door for new areas of research regarding bilingualism and second-language learning. Similarly to [17], the goal of the present study is to try to capture the hypothesized place difference between English and Spanish with acoustic data. The present study focuses on monolingual speakers of both languages. Our future goals include studying the behavior of Spanish-English bilinguals to determine whether they exploit the place differences between Spanish and English coronal stops, similarly to [18].

The first four spectral moments—center of gravity (COG), standard deviation (SD), skewness (SK), and kurtosis (KT)—provide acoustic measurements related to the shape of a spectrum (i.e. how the energy is distributed across frequency bands) [10]. Various investigators have used spectral moments to distinguish between place differences in fricatives [7, 11]; however, [17] is one of a reduced number of studies to use spectral moments in order to analyze place differences in stops. Specifically, Sundara examined coronal stops in French and English, which (similarly to Spanish and English) are realized with dental and alveolar place, respectively. Her investigation found differences between the two languages in relative burst intensity, COG, SD, and KT that were triggered by differences in POA. It remains an open question whether place differences between Spanish and English can be accounted for in the same manner.

2. METHOD

The goal of this investigation was to explore the acoustic correlates that differentiate Spanish from English coronal stops. We measured VOT, the first four spectral moments, and relative burst intensity (see below). After establishing the expected differences in terms of VOT, we focused on an analysis of only the two short-lag stops (English /d/ and Spanish /t/). Of particular interest was the relative importance of each of these spectral measures with regard to POA differences across the two languages.

2.1. Speakers

In order to address the aforementioned issues, we recorded the speech of 16 female participants. Eight were native Spanish speakers between the ages of 18 and 23, all of which were recruited from the *Universitat de les Illes Balears* campus community and were born and raised on the island of Majorca, Spain. Eight were native English speakers and were undergraduate students at the University of Arizona, born and raised in the US Southwest. The Spanish speakers had studied some English in Spain, and the English speakers had studied some Spanish in the U.S., but none of the speakers were able to maintain a basic conversation in their "second" language.

2.2. Materials and Procedure

We devised a list of 48 target words, 24 in English and 24 in Spanish. The target words contained the voiced and voiceless coronal stops of both languages in word initial position. For each language there was a total of 24 words, 12 beginning with /d/ and 12 beginning with /t/, equally divided between stressed and unstressed syllables. All stops were followed by a low vowel (/a/ for Spanish and /æ, ɑ/ for English). (See [17].)

In order to collect the acoustic data we used the "delayed repetition technique" widely used in bilingual-speech research [5]. The materials were read by 6 male native speakers of these languages: 3 native English speakers (recorded in Austin, Texas) and 3 native Spanish speakers (recorded on Majorca, Spain). These acoustic materials were used as auditory stimuli to be repeated outloud by the 16 female speakers whose speech is analyzed here.

The speakers produced the target words in the carrier phrase "_ is the word" or the Spanish equivalent ("_ es la palabra"). All words not containing coronal stops were considered distractors. The computer program Praat [3] presented the sentences randomly in auditory form and the speakers were asked to listen to the entire sentence and then repeat it outloud after a beep at their own pace. They were not asked to imitate the voices of the male talkers, but to produce the sentences in their "own way."

The English data were recorded in a sound attenuated booth on the campus of the University of Arizona. The Spanish data were obtained in a quiet classroom on the campus of the *Universitat de les Illes Balears*. In order to carry out the recordings we used a Shure SM10A dynamic head-mounted microphone, a Sound Devices MM-1 microphone pre-amplifier and a Marantz PMD660 digital speech recorder. The signal was digitized at 44.1 kHz and 16-bit quantization.

Each participant provided the dataset with 72 coronal stops (24 target words $\times$ 3 repetitions). Thus, a total of 1,152 tokens were recorded (24 words $\times$ 3 repetitions $\times$ 16 participants = 1,152 stops). Our initial analysis of VOT utilizes the entire dataset; however, for subsequent analyses of burst measurements we took a subset of this data (exactly half) containing only Spanish /t/ and English /d/, as these are the stops that are not easily be distinguished by VOT. Five tokens were removed due to mispronunciations or extraneous noise leaving a total of 571 tokens for burst analyses.

2.3. Measurements

The digitized sound files were low-pass filtered at 11.025 kHz. For each of the coronal stops, synchronized waveform and spectrographic displays were used to mark the onset of modal voicing and of the burst. The onset of voicing was taken to be the upwards zero-crossing of the first periodic pattern found in the oscillogram [12]. Voice-onset time was calculated as the difference (in ms) between the onset of modal voicing and the onset of the burst.

Unlike in [17], the duration of the burst was determined semi-automatically. For short-lag stops, the burst was equal to the duration of VOT (see above). Thus, in cases in which VOT was positive but smaller than 25 ms, the burst was variable. For stops with long-lag and lead VOT, the burst was exactly 25 ms. That is, the onset of the burst was determined by hand, but the offset of the burst was set to occur 25 ms after the onset—i.e., the longest bursts were 25 ms.

In order to calculate relative burst intensity (RI), we extracted the intensity of the burst (in dB), as well as the mean intensity of the following vowel, which was also manually segmented. RI was the difference between the intensity of the vowel and the intensity of the burst. All spectral measures (COG, SD, SK, KT) were derived from the spectral enve-

lope (which ranged from 60 Hz to 11.025 kHz.). Tokens for which a clear burst could not be established were removed.

2.4. Statistical analyses

A linear mixed-effects model was fit to each acoustic measure detailed above (i.e. VOT, RI, COG, SD, SK, and KT). The analysis of VOT included the entire dataset. Language (Spanish, English) and consonant (/d/, /t/) were fixed effects. Individual speaker and word items were random effects [1], with random slopes for subjects and items for the effect consonant [2]. Statistical significance of group, consonant, and the group by consonant interaction were assessed using hierarchical partitioning of variance via nested model comparisons. Simultaneous Tests for General Linear Hypotheses analyzed all pairwise comparisons using Tukey Contrasts with adjusted p-levels.

Subsequent analyses of residualized burst measurements only included Spanish /t/ and English /d/ data. The present study was concerned with analyzing the acoustic correlates (aside from VOT) that could account for POA differences. Because all burst metrics are directly related to VOT, the effect of this variable was partialed out of the burst measurements as a method of reducing multicollinearity between predictors. In order to accomplish this, separate models were fit with each burst measurement regressed on VOT. The residuals of these models were then used as the predictors for all analyses.[1] Thus, each omnibus model directly compared Spanish /t/ to English /d/.[2] In each case, individual speaker and word items were given random intercepts. We report marginal R^2 and conditional R^2 as an indication of goodness of fit for all models [14]. Marginal R^2 provides a measure of variance explained without mixed-effects and conditional R^2 includes them.

The second analysis explored the extent to which each of the acoustic measures could provide useful information about the POA of the phonetically voiceless segments. To this end, we divided the dataset into two subsets of Spanish /t/ and English /d/ stops: a training set, comprised of 75% of the data, and a testing set, comprised of 25% of the data. We then used RI, COG, SD, SK and KT as predictors in a forward selection logistic regression model in which phoneme identity (Spanish /t/, English /d/) was the criterion variable. Causal priority was given to the correlates found to best predict POA in [17]. After building the model on the training subset, we used cross-validation to predict the identity of the stops in the testing subset.

3. RESULTS

3.1. VOT

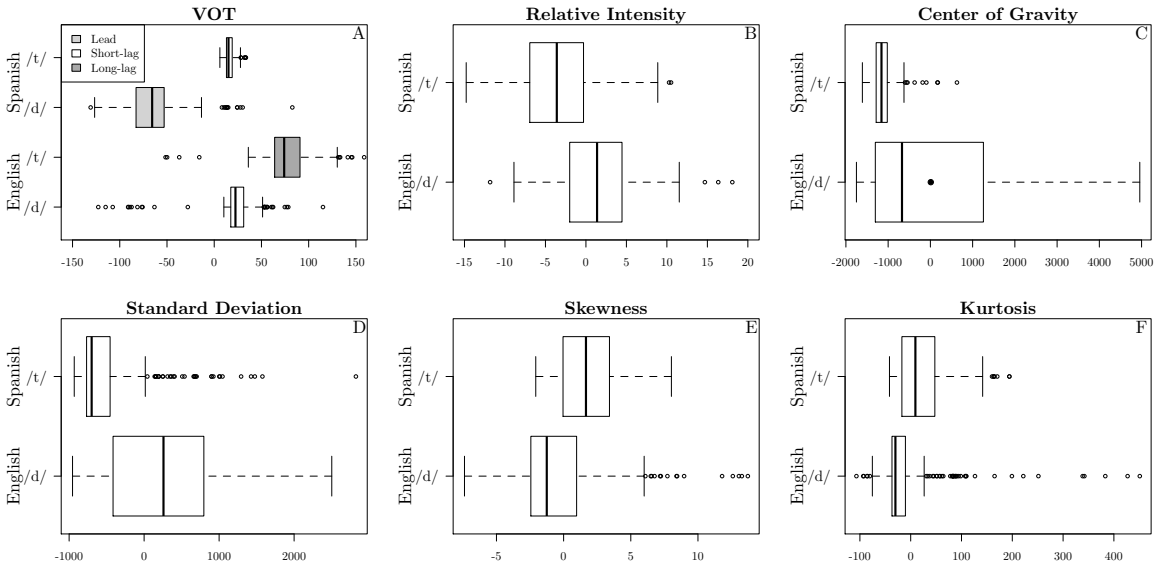
The analysis of the VOT data revealed a main effect of language ($\chi(2) = 44.07$; $p < 0.001$), consonant ($\chi(2) = 56.39$; $p < 0.001$), as well as a language by consonant interaction ($\chi(1) = 10.54$; $p < 0.002$). Pairwise comparisons showed that all of the coronal stops differed from each other ($p < 0.001$), with the exception of the Spanish /t/ *vs.* English /d/ short-lag stops. The mixed model provided the best fit for the data (conditional $R^2 = 0.87$; marginal $R^2 = 0.82$). Figure 1A plots VOT as a function of language and consonant. The light gray box shows that Spanish /d/ was produced with lead VOT ($\bar{x} = -64.48 \pm 28.56$ SD), and the dark gray box shows that English /t/ was produced with long-lag VOT ($\bar{x} = 77.63 \pm 25.63$ SD). The white boxes represent the short-lag stops of English and Spanish. VOT for English /d/ was slightly longer ($\bar{x} = 22.13 \pm 26.69$ SD) than Spanish /t/ ($\bar{x} = 16.18 \pm 5.08$ SD); however, this difference was negligible, likely due to the high rate of variability for English /t/. Thus, in these data VOT can account for differences between all coronal stops except for those that are manifested through short-lag VOT: English /d/ and Spanish /t/.

3.2. Burst measurements

Figure 1B plots RI of Spanish /t/ and English /d/. The data were best fit using the mixed-effects model (conditional $R^2 = 0.66$; marginal $R^2 = 0.20$). The analysis revealed that English /d/ was 4.81 dB $\pm$1.77 standard errors (SE) higher that Spanish /t/ ($t = -2.72$; $p < 0.02$).

The COG data were also best fit using the mixed-effects model (conditional $R^2 = 0.68$; marginal $R^2 = 0.19$). Spanish /t/ was 1131 Hz $\pm$ 455 SE lower than the average English /d/ ($t = -2.48$, $p < 0.03$; see Figure 1C). Regarding SD, the mixed-effect model accounted for 61% of the variance (vs. marginal R^2 of 26%). The SD values for Spanish /t/ were 763 Hz $\pm$ 222 SE lower than English /d/ ($t = -3.44$, $p < 0.004$; see Figure 1D). The analysis of SK (Figure 1E) revealed that Spanish /t/ was 2.10 ± 0.87 SE units higher than English /d/ ($t = 2.40$; $p < 0.03$). Again, the data were best fit using the mixed effects model (conditional $R^2 = 0.45$; marginal $R^2 = 0.12$). Lastly, the KT data had the least amount of variance explained by the model (conditional $R^2 = 0.26$; marginal $R^2 = 0.06$). Spanish /t/ was 29.02 ± 14.30 se units higher than English /d/; however, this difference was not significant at our specified alpha level ($t = 2.03$, $p = 0.06$; see Figure 1F). In sum,

Figure 1: VOT and burst measures of Spanish and English coronals.



all of the burst measurements, with the exception of kurtosis, differed as a function of language. This is taken as an indication that these metrics successfully accounted for place differences between the short-lag stops of English and Spanish.

The next step was to analyze the relative contribution of the burst measurements. The training subset of the data was analyzed via logistic regression, with the burst metrics as fixed effects for predicting the short-lag stop phoneme identity (Spanish /t/, English /d/). The model eliminated SK and KT from the analysis. Table 1 summarizes the results. Nagelkerkes' pseudo R^2 is reported to give an indication of goodness-of-fit of each predictor as it was entered into the model. SD accounted for the largest amount of the variance (35%), followed by RI (5%) and COG (2%).

Table 1: Regression analysis of /d/-/t/.

Metric	R^2	R^2_{change}	χ^2_{change}	*p-value*
SD	.353	.353	175.39	< 0.001
COG	.376	.023	14.01	< 0.001
RI	.425	.049	30.01	< 0.001
SK	.426	.001	0.08	> 0.05
KT	.432	.006	4.36	< 0.04

Finally, the best fit model was used to predict the identity of the short-lag stops of the testing subset. The model performed with 87% accuracy (out of sample error = 0.13). That is, given the data for SD, RI and COG, it was possible to predict whether the stop was Spanish /t/ or English /d/ on 87% of the testing subset (142 tokens).

4. DISCUSSION AND CONCLUSION

The results of the analyses showed that the phonetically voiceless coronal stops of English and Spanish can be distinguished by RI and by the spectral shape of the stop burst. Importantly, the present study indicates that the place of articulation differences described for Spanish /t/ and English /d/ are best accounted for using measures of SD, RI, and COG. SK and KT did not significantly contribute to predicting the POA of the short-lag stops in our data.

Our results partially corroborate the findings of [17]. As is the case in the Canadian varieties of English and French investigated in [17], the two varieties of English and Spanish investigated here differ in place of articulation of their coronal stops—alveolar (English) and dental (French, Spanish), according to impressionistic descriptions. In our data, similar to [17], values of SD and other burst measures varied across the two languages; however, different from our findings, she also encountered significant differences for kurtosis.

The present study contributes language-specific acoustic characteristics of bursts in the short-lag coronal stops of two monolingual varieties of English and Spanish. Among other things, the findings provide base acoustic descriptions for future studies on Spanish-English bilinguals.

5. REFERENCES

[1] Baayen, R. H., Davidson, D. J., Bates, D. M. Nov. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390–412.

[2] Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. Apr. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 255–278.

[3] Boersma, P., Weenink, D. 2012. Praat: doing phonetics by computer. *Glot International* 5, 341–345.

[4] Cole, J., McMurray, B., Munson, C., Linebaugh, G. 2010. Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics* 38, 167–184.

[5] Flege, J. E., Munro, M., MacKay, I. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16, 1–26.

[6] Fowler, C. A. 1984. Segmentation of coarticulated speech in perception. *Perception & Psychophysics* 36, 359–368.

[7] Gordon, M., Barthmaier, P., Sands, K. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32, 141–174.

[8] Gow, D. W. 2003. Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics* 65, 575–590.

[9] Hualde, J. I. 2005. *The sounds of Spanish*. Cambridge University Press.

[10] Jones, M. J., Knight, R.-A. 2013. *Bloomsbury Companion to Phonetics*. London, UK: A & C Black.

[11] Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108, 1252–1263.

[12] Lieberman, P., Blumstein, S. E. 1988. *Speech physiology, speech perception, and acoustic phonetics*. Cambridge, UK: Cambridge University Press.

[13] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384–422.

[14] Nakagawa, S., Schielzeth, H. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4, 133–142.

[15] Picard, M. 1987. *An introduction to the comparative phonetics of English and French in North America* volume 7. Amsterdam: The Nethrlands: John Benjamins Publishing.

[16] R Core Team, 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[17] Sundara, M. 2005. Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *The Journal of the Acoustical Society of America* 118, 1026–10037.

[18] Sundara, M., Polka, L., Baum, S. 2006. Production of coronal stops by adult simultaneous bilinguals. *Bilingualism: Language and Cognition* 9, 97–114.

[19] Williams, L. 1977. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics* 21, 289–297.

[20] Williams, L. 1979. The modification of speech perception and production in second-language learning. *Perception & Psychophysics* 26, 95–104.

---

[1] See [6], [8] and [4] for discussion and examples of this approach.

[2] Degrees of freedom for hypothesis tests were derived using the Satterthwaite approximation as implemented in the *lmerTest* package in R [16].