The acoustic realization of L2 Spanish phonetic categories and allophonic alternations by English-speaking immigrants

Mark Amengual

University of California, Santa Cruz
amengual@ucsc.edu

ABSTRACT

This experiment analyzes the acoustic realization of English and Spanish voiceless stops /p,t,k/ (*phonetic categories*) and Spanish voiced stops/approximants /b,d,g/-[β,ð,ɣ] (*allophonic alternations*) produced by native English Second Language (L2) learners of Spanish and their Spanish/Catalan/English trilingual children. Fifteen British-born, native English-speaking immigrants, who have resided in Spain for +30 years and fifteen English heritage speakers, who were raised and educated in the same Spanish/Catalan bilingual community, completed a reading-aloud task in which they produced Spanish and English /p,t,k/ and Spanish /b,d,g/-[β,ð,ɣ]. The acoustic analyses suggest that the voiced stop/approximant allophonic alternation in Spanish may be relatively easier to acquire than voiceless stops for highly proficient L2 learners that are long-term residents in the L2-speaking community. These findings add to our understanding of the sources of variation in L2 pronunciation providing insight into the differences in the acquisition of phonetic categories and allophonic alternations in the L2.

Keywords: L2 phonetic categories; Allophonic alternations; Ultimate attainment; Speech production

1. INTRODUCTION

It is well-known that the probability of acquiring a near-native or native-like accent in a foreign language diminishes as we age [1, 23, 24]. Thus, hearing and speaking with a “foreign” or “First Language (L1)-influenced” accent is minimized the earlier the language is acquired [9, 15]. However, we still do not have a clear understanding of the speech production and perception patterns of adults who learn a Second Language (L2) at a later stage in life but who have many years of experience with the language as a result of long-term and continued residence in the L2-speaking community. This brings us to the following question: What does the ultimate L2 attainment (i.e., the acquisitional “end state”) look like for highly-experienced long-term immigrants? It may be reasonable to expect that these individuals, who in many cases have spent most of their lives with intensive exposure to their L2 and are immersed in a community where their L2 is exclusively used, are those who are better equipped to overcome age of acquisition effects in their L2 speech. Several recent studies that investigate the speech patterns of long-term L1 English speakers that have immigrated from the USA to Spain [12, 13, 14] suggest that this may not be the case, showing uneven rates of success in the acquisition of a variety of sound classes in Spanish.

Additionally, most research on L2 speech has exclusively examined the acquisition of phonemic contrasts but the phonetic knowledge of L2 learners goes beyond the acquisition of two phonological systems with contrasting sounds. In comparison to how much we know about the acquisition of language-specific phonetic categories in L2 speech, we still know relatively little about the implementation of language-specific phonological processes, such as allophonic alternations and neutralization processes [4, 7, 10, 16, 21, 26]. The present study bridges this gap by comparing the acoustic realization of voiceless stops /p, t, k/ in L1 English and L2 Spanish (*phonetic categories*) and the L2 Spanish voiced stop /b,d,g/-approximant [β,ð,ɣ] (*allophonic alternation*) by advanced L1 English late L2 learners of Spanish and their early simultaneous Spanish-Catalan-English trilingual children.

In order to examine the acquisition of *new phonetic categories* we analyze the Voice Onset Time (VOT) of voiceless stops. This phonetic variable was selected because word-initial Spanish /p,t,k/ have a short VOT and are always unaspirated [p,t,k], with Spanish VOT ranging between 0 and 30 ms [2, 17, 22] whereas word-initial English voiceless stops are produced with a long lag and are aspirated [p^h,t^h,k^h], resulting in a VOT from 30 to 120ms [8, 22]. To explore the acquisition of a *new allophonic alternation*, this experiment investigates a Spanish-specific lenition process (i.e., spirantization), wherein the voiced stops /b,d,g/ in Spanish are articulated with full occlusion only phrase initially, after a homorganic nasal, and for /d/ only when there is a preceding /l/. In all other positions, including in intervocalic position, these voiced stops are commonly realized as voiced approximants [β,ð,ɣ] [18].

2. METHOD

2.1. Participants

The participants in the present study comprise 15 female British-born, native English-speaking long-term immigrants in Spain who are L2 Spanish speakers (Age: M=56.6, SD=7.7) and 15 female early simultaneous Catalan-Spanish-English trilinguals (Age: M=26.4, SD=7.3) who reside in the same Spanish-Catalan bilingual community. Participants reported normal speech and hearing and normal or corrected-to-normal vision, and received a stipend for their participation in the study.

The British-born L1-English-L2 Spanish group consisted of British Expatriates (BE) who were born, raised, and educated in Scotland or England. Most of these native English speakers immigrated to Spain in their early 20s, settled on the island of Majorca, and married a native Spanish-Catalan bilingual partner. At the time of testing they reported (i) that they had been residing in Spain for an average of 32.6 years (SD=10.1), (ii) that they spoke in Spanish with their partners and in the community and English with their children, (iii) that even though they understood Catalan they did not regularly speak Catalan, and (iv) that they were not native speakers of any other language. The early simultaneous trilingual group consisted of English Heritage Speakers (EHS) who were raised, and educated in Majorca (Spain). They had grown up in a trilingual environment speaking English with their mothers, Catalan with their fathers, and Spanish and Catalan in the community.

2.2. Materials and Recording Procedure

The production of the target Spanish and English voiceless stops /p,t,k/ and Spanish /b,d,g/-[β,ð,ɣ] was elicited in a reading-aloud task. The materials consisted of 30 word-initial voiceless stops (10 /p/, 10 /t/, 10 /k/) for each language, in stressed position and preceding a low vowel (e.g., *cama* ‘bed’, *casa* ‘house’, cat, catch) and 30 word-initial and 30 intervocalic (10 bilabial, 10 dental, 10 velar) Spanish voiced stops in stressed position and preceding a mid vowel (e.g., *boca* ‘mouth’, *jabón* ‘soap’). Each target item was embedded in a carrier phrase: *I say* TARGETWORD *today* (English), *Puedo decir* TARGETWORD *también* (Spanish). Participants produced two repetitions of each experimental item.

The production task was conducted in a quiet room in the home of each participant, who was comfortably seated in front of a computer display. Participants were told that the study involved reading sentences on a computer screen and that their speech would be recorded for subsequent acoustic analysis. Each sentence was presented on a computer screen and participants were asked to read the sentences clearly and with a natural pace, producing “clearly enunciated speech” [20]. The speech samples were recorded using a head-mounted microphone (Shure SM10A) and a solid-state digital recorder (Marantz PMD660), digitized (44kHz, 16-bit quantization), and computer-edited for subsequent acoustic analysis.

2.3. Acoustic Analysis

The VOT values of the voiceless stops in Spanish and English were obtained from the waveform in Praat [6] by measuring the time interval between the stop release and the onset of voicing as discerned on the waveform as periodic (repeating) cycles. The measurement (rounded to the nearest decimal) was determined from the beginning of the burst (identified by a sharp spike where the waveform changes from quiescent to transient) to the beginning of the first regularly repeating voicing cycle. The point in the first glottal cycle that was counted as the onset of voicing was the initial zero crossing in the waveform.

Two gradient measures of intensity from the target word-initial and intervocalic Spanish /b,d,g/-[β,ð,ɣ] were taken using Praat. The first was the lowest intensity point of the voiced stop/approximant, and the second was the intensity peak of the following vowel. For each item, the valley of the consonant and the peak of the following vowel on the intensity curve were annotated. All points were manually marked, and a Praat script was run to extract all points and subtract all the valleys from their corresponding peaks. The degree of spirantization was determined by the difference between the intensity minimum of the consonant and the intensity maximum of the following tautosyllabic vowel (*IntDiff*; [19]). The more lenited the consonant, the smaller the *IntDiff* (dB) is expected to be with respect to the following vowel. Previous studies have reliably used this measure to determine the degree of spirantization in the production of approximants in Spanish [3, 11].

Each participant produced 120 target voiceless stops and 120 voiced stops/approximants, for a total of 7,200 tokens. Twelve voiceless stops and 54 voiced stop/approximant tokens were excluded due to mispronunciations or recording errors. As a result the dataset comprised 3,588 VOT (ms) and 3,546 intensity (dB) measurements.

3. RESULTS

3.1. Voice Onset Time

In order to compare the VOT values of BE and EHS as a function of their language and point of articulation, a dataset was created including the average over subjects as a condition of language (Spanish, English) and phoneme (/p/, /t/, /k/). As shown in Figure 1, both groups maintain language-specific VOT values in their production of voiceless stops in each language. In other words, they produce distinct VOT values in English and Spanish.

The dataset was submitted to a mixed-model ANOVA, which was performed using R (R Core Team, 2022), with *Speaker Group* as between-subjects factor, *Language* and *Phoneme* as within-subjects factors, and *Subject* as the random term. The mixed-design ANOVA yielded significant main effects of *Speaker Group* (F(1,28) = 22.3, p<0.001), *Language* (F(1,140) = 573.7, p<0.001) and *Phoneme* (F(2,140) = 53.1, p<0.001). In addition, there was a significant interaction between *Language* and *Phoneme* (F(2,140) = 4.1, p<0.05).

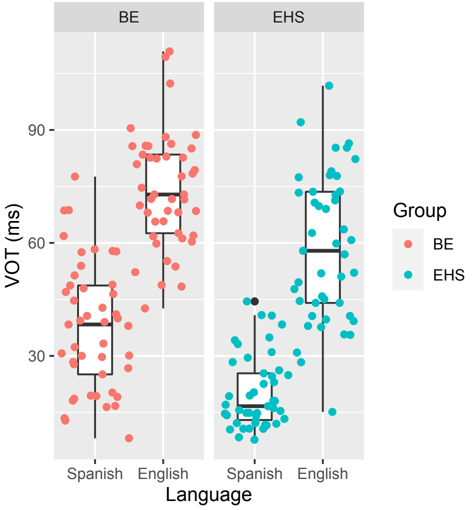

Figure 1: VOT values as a function of group and language.

Figure 2 presents the Spanish VOT values (ms) for each phoneme produced by the BE and EHS. As the main focus of this study is the acoustic realization of L2 Spanish voiceless stops, pairwise comparisons between both participant groups were performed using Tukey's HSD test. In all cases, paired t-tests and effect sizes calculated by means of paired Cohen's *d* are reported.

An analysis of the Spanish VOT data demonstrates that the BE group is far from producing VOT values in the Spanish target range, as shown in the production of the EHS group. The pairwise comparisons revealed a significant difference between the VOT values produced by BE and EHS for /p/ (diff. = 14.69, t(14) = 4.58, p<0.001, Cohen's d=1.18), /t/ (diff. = 19.3, t(14) = 5.2, p<0.001, Cohen's d=1.34), and /k/ (diff. = 18.19, t(14) = 3.99, p<0.001, Cohen's d=1.03).

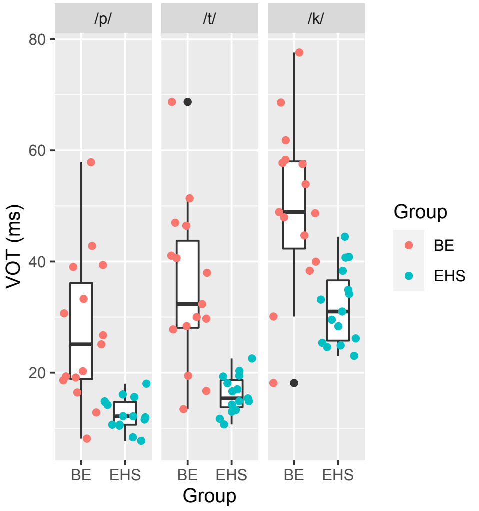

Figure 2: VOT values of Spanish /p,t,k/ by BE and EHS.

3.2. Spirantization

The Intensity Difference (*IntDiff*) data are plotted in Figure 3. The figure includes two panels: one for each speaker group with a boxplot that represents the post-pausal phrase-initial productions, and the other that corresponds to the intervocalic realizations. These data show that *IntDiff* is consistently larger for phrase-initial voiced stops than for intervocalic voiced stops, demonstrating that both groups are applying the spirantization rule in Spanish.

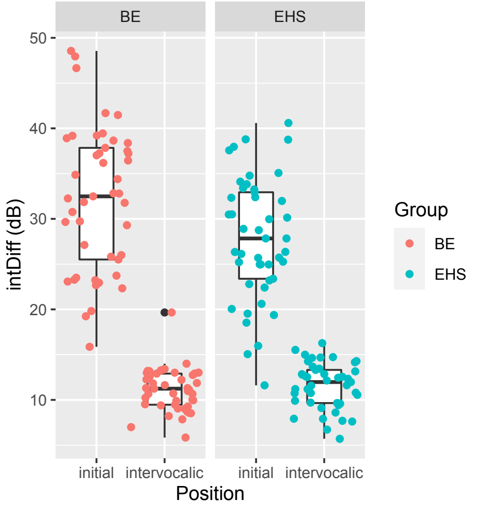

Figure 3: *IntDiff* of word-initial and intervocalic /b,d,g/-[β,ð,ɣ] as a function of place of articulation.

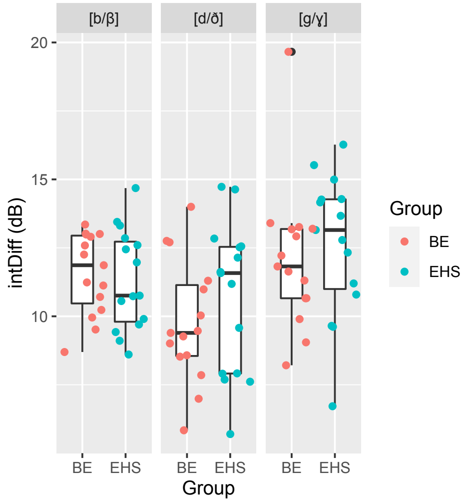

Figure 4: *IntDiff* of intervocalic /b,d,g/-[β,ð,ɣ] as a function of place of articulation.

The *IntDiff* data were also analyzed through mixed-model ANOVAs with *Speaker Group* as between-subjects factor, *Phoneme* ([b/β, d/ð, g/ɣ] and *Word Position* (word-initial, intervocalic) as within-subjects factors, and *Subject* as the random term. The mixed-design ANOVA yielded significant main effects of *Phoneme* (F(2,140) = 9.7, p<0.001), and *Word Position* (F(2,140) = 882.2, p<0.001). In addition, there were significant interactions between *Speaker Group* and *Word Position* (F(1,140) = 14.4, p<0.001) and between *Phoneme* and *Word Position* (F(2,140) = 6.3, p<0.001). The model did not yield a significant effect of *Speaker Group*.

4. DISCUSSION AND CONCLUSIONS

This study examined the acoustic realization of L2 Spanish voiceless stops /p,t,k/ (*new phonetic categories*) and the L2 Spanish voiced stop /b,d,g/-approximant [β,ð,ɣ] allophonic distribution (*new allophonic alternation*) produced by advanced late learners of Spanish, who are British Expatriates (BEs) and their Spanish-Catalan-English trilingual children, who are English Heritage Speakers (EHSs).

In terms of their acoustic realization of voiceless stops /p,t,k/, the results indicate that even though both groups produced acoustically distinct VOT values in each language, the Spanish VOT values of the voiceless stops produced by the BEs converge towards the English-like range, with intermediate ("compromise") values between the VOTs in Spanish and English as produced by the EHSs. With respect to the acquisition of the voiced stop/approximant allophonic alternation in Spanish, both groups produced a more lenited consonant in intervocalic position than in phrase-initial position, showing that they had acquired the Spanish spirantization rule, producing comparable degrees of spirantization in intervocalic position. These results indicate that BEs seem to fall short of target-like performance with regards to their VOT values but they more closely approximate the *IntDiff* values of their native Spanish-speaking children.

With respect to their ultimate attainment in the acquisition of L2 Spanish voiceless stops (*phonetic categories)* and spirantization in intervocalic voiced stops (*allophonic alternation*), on the one hand, the VOT data appear to show signs of fossilization (arrested development) in the acquisition of L2 learners who are truly exceptional in the amount of extended Spanish immersion and usage they have had during their lives. On the other hand, their acoustic realization of the allophonic alternation between /b,d,g/-[β,ð,ɣ] in Spanish indicates a different, more target-like, acquisitional path. As the data might suggest, are allophonic alternations "easier" to acquire than language-specific phonetic categories? If so, why would this be the case?

The Second Language Perceptual Assimilation model (PAM-L2; [5]) postulates that there are at least two levels involved in cross-linguistic sound interactions: the phonetic and the phonological level. In the acquisition of L2 sounds, the notion of phonetic category coexists with that of the complementary distribution found in allophonic alternations, such as spirantization. The results of this experiment suggest that, at least for experienced late L2 Spanish learners, [β,ð,ɣ] may be easier to acquire than /p,t,k/ because intervocalic articulatory reduction (lenition) can be predicted by a phonological process that is highly frequent in discourse.

Because L2 phonetic categories, such as voiceless stops for English-Spanish bilinguals, consist of cross-linguistically 'similar' sounds but with subtle, fine-grained differences (a tight connection with an L1 category), this may impede 'target-like' acquisition. In contrast, sounds that are only loosely connected with L1 categories (such as [β,ð,ɣ]) may be particularly easier to learn because they are in complementary distribution with /b,d,g/ and the bilingual individual may be more likely to notice, over time, that this phonological process is specific to the L2, and sufficiently different from comparable sounds in the same phonological context in their L1. This predictability and high frequency may attract attention, which would be conducive to phonetic learning.

In closing, two questions can be posed for future research: how close to a 'native-like' range must a learner realize these L2 segments for the acoustic differences to be inconsequential in contributing to a foreign accent? How are these sounds perceived in tandem with other segments that are similarly prone to cross-linguistic influence in L2 speech acquisition?

5. ACKNOWLEDGMENTS

I would like to express my gratitude to Magaly Gastélum, Rebecca Hernández, Serene Tseng, Lindsey Archer, Daisy Soto, Caitlyn Welsh, Katrina Escher, Richard Bibbs, and Chase Rodriques from the UCSC Bilingualism Research Lab for their assistance with the acoustic analyses. This research was supported by the Hellman Fellows fund.

6. REFERENCES

[1] Abrahamsson, N., Hyltenstam, K. 2009. Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language learning*, *59*(2), 249-306.

[2] Abramson, A. S., Lisker, L. 1973. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1, 1–8.

[3] Amengual, M. 2019. Type of early bilingualism and its effect on the acoustic realization of allophonic variants: early sequential and simultaneous bilinguals. *Int. Journal of Bilingualism,* 23(5), 954-970.

[4] Amengual, M., Simonet, M. 2020. Language dominance does not always predict cross-linguistic interactions in bilingual speech production. *Linguistic Approaches to Bilingualism,* 10(6), 141-166.

[5] Best, C.T., Tyler, M.D. 2007. Nonnative and second-language speech perception: commonalities and complementarities. In: Bohn, O.-S., Munro, M. (eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*, 13-34. Amsterdam: John Benjamins Publishing.

[6] Boersma, P., Weenink, D. 2021. Praat: doing phonetics by computer [Computer program], Version 6.1.39.

[7] Braun, B., Lemhöfer, K., Mani, N. 2011. Perceiving unstressed vowels in foreign-accented English. *The Journal of the Acoustical Society of America*, *129*(1), 376-387.

[8] Cho, T., Ladefoged, P. 1999. Variation and universals in VOT: Evidence from 18 languages, *Journal of Phonetics 27*(2), 207–229.

[9] DeKeyser, R. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(4), 499–534.

[10] Dmitrieva, O., Jongman, A.,Sereno, J. 2010. Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics*, *38*(3), 483-492.

[11] Eddington, D. 2011. What are the contextual phonetic variants of in colloquial Spanish? *Probus, 23*(1), 1–19.

[12] Face, T. 2021. Ultimate attainment of Spanish laterals by native English-speaking immigrants in Spain. *Journal of Second and Multiple Language Acquisition*, 9(1), 167-180.

[13] Face, T. L. 2018. Ultimate attainment in Spanish spirantization: The case of US-born immigrants in Spain. *Spanish in Context*, *15*(1), 27-53

[14] Face, T. L., Menke, M. R. 2020. L2 acquisition of Spanish VOT by English-speaking immigrants in Spain. *Studies in Hispanic and Lusophone Linguistics*, *13*(2), 361-389.

[15] Flege, J. E. 2003. Assessing constraints on second-language phonetics and phonology. In: Meyer, A. Schiller, N. (eds.), Phonetics and phonology in language comprehension and production, differences and similarities, 319–355. Berlin, Germany: Mouton de Gruyter.

[16] Flege, J. E., Bohn, O. S. 1989. An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition*, 11, 35-62.

[17] Flege, J., Eefting, W. 1986. Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, 43, 155–171.

[18] Harris, J. W. 1969. *Spanish phonology*. Cambridge, MA: MIT Press.

[19] Hualde, J. I., Simonet, M., Nadeu, M. 2011. Consonant lenition and phonological recategorization. *Laboratory Phonology*, 2(2), 301–329.

[20] Kessinger, R.H., Blumstein, S.E. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25, 143-168.

[21] Lee, B., Guion, S. G., Harada, T. 2006. Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition*, 487-513.

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

[23] Moyer, A. 1999. Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, 21(1), 81-108.

[24] Patkowski, M. 1990. Age and accent in a second language: A reply to James Emil Flege. *Applied Psycholinguistics*, 11, 73–89.

[25] R Core Team (2022). R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria), http://www.R-project.org/ (Last viewed November 14, 2022).

[26] Smith, B. L., Hayes-Harb, R., Bruss, M., Harker, A. 2009. Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics,* 37(3), 257-275.