Articulatory Coordination in L2-Speakers of Spanish

Stephen J. Tobin[1,2], Mark Gibson[3], Stavroula Sotiropoulou[4], Adamantios I. Gafos[2,4]

[1]University of Michigan, [2]Haskins Laboratories, [3]Universidad de Navarra, [4]Universität Potsdam
sjtobin@umich.edu, mgibson@unav.es, stsotiro@uni-potsdam.de, gafos@uni-potsdam.de

ABSTRACT

Although recent research has shed some light on relative timing in laryngeal and oral articulatory gestures, little is known about articulatory patterns of second language (L2) users along the trajectory from novice to expert. We report results of an electromagnetic articulographic (EMA) investigation of two native English speakers' L2 Spanish—languages in which voicing and word-initial overlap in C1C2 clusters are markedly different. We examine how voicing implementation and onset cluster overlap change as a function of L1 by comparing these speakers to previously recorded native Spanish speakers (under review). Analyses of L1 and L2 groups' articulatory overlap in voiced- and voiceless-stop clusters revealed subtle group differences, the L2 speakers not producing some fine-grained native differences in overlap. However, analyses of voiceless-stop VOTs in singleton/complex clusters revealed substantial differences partially attributable to syllable complexity. Results will be discussed in light of recent research addressing the acquisition of novel coordination patterns in an L2.

Keywords: Speech production; articulatory timing; oral-laryngeal coordination; English; Spanish; L2 speech.

1. INTRODUCTION

A considerable body of research on articulatory timing has emerged over the past 40 years [5], whereby particular types of coordination [18] have been found to intricately reflect the syllabic [8] and prosodic [7] structure of an utterance. For example, English syllable-onset consonant clusters appear to be coordinated with their tautosyllabic vowel by aligning some temporal landmark at the centre of the cluster (C-centre) with the vowel [5,6]. Investigations of other languages have revealed substantial language-specific as well as speaker-specific variation in onset cluster timing [16]. Italian, for example, shows C-centre organization in stop+liquid onsets but a more sequential organization in other word-initial clusters (e.g., /s/+stop as studied in [19]) where only the immediately prevocalic consonant enters into a stable relation with the vowel, with any additional preceding consonants appended to the left (hence, ‘sequential’) of the inner CV. Varieties of Arabic, on the other hand, seem to exhibit exclusively sequential organization in onsets [15,28]. Further, recent investigations [29,17] have revealed relatively little overlap in the timing of oral gestures in onset clusters of peninsular Spanish, similar to Arabic. This is so even though Spanish is claimed to allow clusters as syllable onsets whereas Arabic is claimed to preclude these ([14] and references therein), suggesting that the phonological property of admitting/precluding clusters as syllable onsets does not uniquely determine oral gesture timing schemas across languages. These recent articulatory findings on Spanish are consistent with a long literature on previous acoustic investigations, in which open transitions or intrusive schwa-like vocalic elements between the onset consonants of Spanish clusters are reported [4,21,24,25].

In addition to this variation in coordination among oral gestures, their interplay with any laryngeal devoicing gestures has long be known to be language-specific and, more recently, has been shown to contribute to the timing of the oral gestures. At a fundamental level, the coordination of this laryngeal gesture with a single oral gesture as instantiated in voice onset time (VOT) has been long established to be language-specific [22], Spanish exhibiting prevoicing in voiced stops and short positive VOTs in voiceless stops, and English exhibiting contextually determined prevoicing or short positive VOTs in voiced stops and longer positive VOTs in voiceless stops. More recently, patterns of convergence towards native-like VOTs have been reported to be L1-dependent [31]. Further recent work indicates that the addition of a laryngeal gesture may also affect the coordination of oral gestures. Thus, French [3] exhibits no substantial impact of voicing but German and Spanish [16] show decreased overlap when the initial stop is voiceless.

Here, we asked how patterns of coordination in second-language (L2) speakers compare with those of native speakers. The question of how native speakers of other languages approximate these patterns of coordination as they learn an L2 has received little attention. Investigating the effect of L1 on the level of precision with which gestural coordination of an L2 is approximated will further our understanding of the relative stability of these coordination patterns beyond the well-established stability of in-phase and anti-phase timing [17], and will further aid in our understanding of how new phonetic categories are learned [13]. Languages vary in the number of onset gestures that can be coordinated with one another (cf. cluster complexity), in the language-specific gestural demands of the cluster constituents (e.g., rhotics which are approximants in English, uvular fricatives/trills in French and German, and alveolar taps in Spanish), as well as in the phasing with which they may be coordinated. Thus, the number, identity, and phasing of onset gestures of an L2 learner's L1 may impact the precision with which (s)he approximates coordination patterns in the L2.

Among experimental investigations into L2 articulation, work of Davidson and colleagues [10,12] figures prominently. In these acoustic and ultrasound investigations, native speakers of English and of Slavic languages were recorded producing real and nonce words of Slavic. Some of these contained initial clusters that are phonologically illegal in English (e.g., /zg/). The English speakers' initial tongue body position for these illegal clusters were shown to match English /sC/ clusters more closely than /səC/, indicating that participants were not epenthesizing a targeted schwa, although vocalic material was present between the non-native cluster consonants in the acoustic record. Davidson and colleagues ascribe this to insufficient overlap among the onset gestures in these unfamiliar clusters. They also found that acoustic durations of the consonants were longer among L2 than L1 speakers, consistent with the unfamiliarity of these sequences and lesser gestural overlap among the L2 speakers [11].

In the present investigation, we report articulatory coordination patterns of two English-L1, L2 speakers of Castilian Spanish and we compare these to an existing data set of native speakers of Castilian Spanish (*n*=6) [16]. We are currently collecting data from additional L2 speakers. At the outset, there are reasons to expect that the articulatory patterns of our L2 speakers will closely approximate those of native speakers. This expectation is based on characteristics of Spanish and English phonotactics and on the participants' Spanish proficiency. English exhibits both C-centre timing (in onsets) and sequential timing (in codas) [23]. Thus, sequential coordination forms part of the existing coordinative repertoire of our two participants. Additionally, English phonotactics allows three onset consonants whereas Spanish allows maximally two. Thus, our participants will not be challenged by the number of gestures to be coordinated but may be challenged by the presence of a tap, since taps do not form part of clusters in American English, and are absent from most varieties of British English. Finally, both of our L2 speakers had learned Spanish to an advanced level (see section 2.1) and have been judged to be of native or near-native proficiency by native speakers of Spanish.

We make the additional prediction that participants' voiceless stop VOTs will also approximate those of the native speakers, given that VOT (aspiration) of singleton initial voiceless stops is an aspect of pronunciation that receives attention in Spanish language teaching practice [27]. However, the question of whether their expertise with VOT will transfer to clusters is less clear, particularly since this is a level of phonetic complexity that is not commonly considered in Spanish pronunciation instruction.

2. METHOD

2.1. Participants

L2 participants were two native speakers of English. The speakers differed somewhat in their language background. S1 was a native speaker of American English, was immersed in Spanish from early childhood and had resided in a Spanish-speaking country for the 17 years preceding the experiment. S2 was a native speaker of British English who learned Spanish in an academic context, supplemented with visits to Spain and an academic year abroad in Madrid, but had only made occasional use of Spanish over the 15 years prior to the experiment. Both speakers had learned Spanish to an advanced level (Common European Framework of Reference level C2 [9]). The native Spanish-speaking participants were six speakers from Central Spain.

2.2. Stimuli & Procedure

Stimuli were 42 disyllabic words of Spanish with singleton (C1V) and cluster (C1C2V) onsets. In the latter, C1 was a stop from the set /b, g; p, t, k/ and C2 was a lateral or a tap, /l, ɾ/. Words were spoken in the sentence 'Di ___ por favor' /di ___ poɾ faβoɾ/. Having given informed consent in accordance with Universität Potsdam's Ethikkommission, participants sat in a sound-treated booth and stimuli were presented on a computer monitor. Five repetitions of each stimulus word were collected. Data were collected using a Carstens AG501 Electromagnetic Articulograph. Audio data were recorded with a t.bone EM 9600 unidirectional microphone placed 1 metre from the participant.

2.3. Data Processing & Measurement

Data were corrected for head-movement and rotated into a standard orientation using standard procedures. Measurement of data took place in Mview using adaptations of MATLAB scripts developed by Tiede and colleagues [30].

Gestural overlap was defined as the latency between the release of C1's gestural plateau and the onset of the C2 gestural plateau. In measuring gestural overlap via the difference between these two landmarks, negative values imply gestural plateau overlap and positive values indicate the existence of an inter-plateau interval (henceforth IPI).

VOT of voiceless stop-initial words was defined as the lag from the release burst following the stop closure silence in the acoustic waveform to the onset of voicing, determined with reference to the presence of periodicity in the waveform and the presence of a voice bar in the spectrogram.

2.6. Analysis

Linear mixed-effects regression [2] was applied in R [26]. Separate models were fit for the gestural overlap and the voiceless stop VOT data. Model outputs are reported using the lmerTest [20] package *anova* function, which provides estimated *F*-test equivalents of *lmer* model fits. The overlap model included fixed effects of C1 Voicing (voiced vs voiceless), C1 Place of articulation (labial vs velar), and Speaker Group (native vs non-native), and all interactions, along with random intercepts for Participant and random Participant-wise slopes for Voicing. The (voiceless stop) VOT model included fixed effects of C1 Place (labial vs velar), Onset Cluster Complexity (C1 vs C1C2), and Speaker Group (native vs non-native), and all interactions, along with random intercepts for Participant and random Participant-wise slopes for Voicing. We applied parsimony in our random effects structure since the data did not support maximal random effects [1].

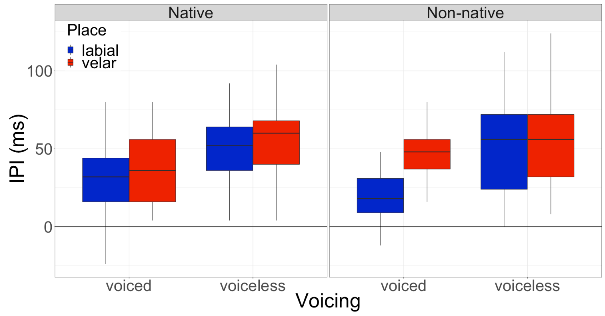
3. RESULTS

3.1. Gestural Overlap

The gestural overlap model revealed significant effects of Voicing ($F(1,8.02)=16.21$, $p<.01$) and Place ($F(1,1009.4)=45.73$ $p<.01$) but no effect of Speaker Group. Overall, then, the two non-native speakers' gestural coordination patterns approximated those of the native speakers well, supporting our first hypothesis. The boxplots for each cell of the design are presented in in Fig. 1. The non-native speakers, like the native speakers, exhibited significantly greater overlap (lower IPIs) when C1 was voiced (left pair of boxplots in the two panels of Fig. 1) than when it was voiceless (right pair of boxplots in each panel) in labial-initial words ($t(9.74)=3.56$, $p<.01$; though not in velar-initial words). Likewise, greater overlap was observed (lower IPI values) for labial (blue boxplots) than velar (red boxplots) among both native and non-native groups.

Nonetheless, the Speaker Groups' cell-specific boxplots in Fig. 1 are not identical. A significant three-way interaction of Voicing x Place x Speaker Group was observed ($F(1,1009.98)=16.92$, $p<.01$). The interaction is driven by Speaker Group—the non-native speakers showing an interaction of Voicing and Place ($F(1,1009.98)=14.22$, $p<.01$) not present for the native speakers. Among the non-native speakers, the effect of Voicing on overlap was greater among /p, b/-initial words (blue boxplots in the right panel of Fig. 1) than among /k, g/-initial words (red boxplots in the right panel of Fig. 1). Stated conversely, the effect of place on overlap in the non-native speakers' voiced stop-initial /b, g/ words (left pair of boxplots in the right panel of Fig. 1) was greater than in their voiceless stop-initial /p, k/ words (right pair of boxplots in the right panel of Fig. 1) and greater than among the native speakers. In fact, model predictions for overlap in the non-native speakers' voiceless-initial onsets were not significantly different from one another.

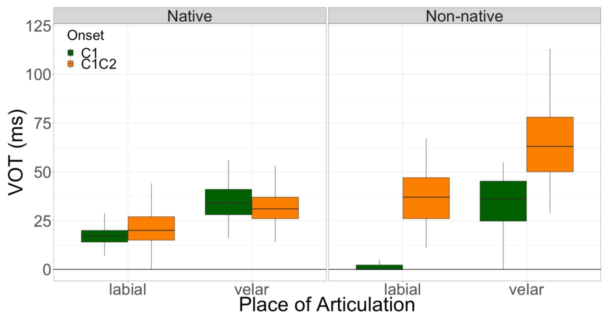
Figure 1: Boxplots of Native (*n*=6) and Non-native (*n*=2) Spanish speakers' Inter-Plateau Intervals by Voicing and Place of Articulation.



3.2. VOT

As in the gestural overlap model, the effect of Speaker Group was not significant in the VOT model, consistent with our second hypothesis that the L2 speakers would reasonably accurately approximate the L1 speakers' oral-laryngeal coordination patterns. The model did, however, reveal significant effects of C1 Place ($F(1,885.71)=501.10$, $p<.01$) and Onset Cluster Complexity ($F(1,5.77)=86.20$, $p<.01$). VOT was shorter among bilabials than velars in both Speaker Groups at both levels of Cluster Complexity. Cellwise boxplots are presented in Fig. 2.

Figure 2: Boxplots of Native (*n*=6) and Non-native (*n*=2) Spanish speakers' VOTs by Place of Articulation and Onset Cluster Complexity.



The effect of Cluster Complexity participated in two-way interactions with Place of Articulation ($F(1,885.66)=4.49$, $p<.05$) and with Speaker Group ($F(1,885.71)=56.38$, $p<.01$). Among the native speakers, Cluster Complexity only yielded a significant difference in VOT between labial-initial words ($t(6.37)=2.43$, $p<.05$), velars showing no significant difference. However, among the non-native speakers, the interaction of Cluster Complexity and Place yielded significant effects of Cluster Complexity for both labials ($t(9.55)=9.46$, $p<.01$) and velars ($t(12.35)=8.12$, $p<.01$).

Among pairwise comparisons between Speaker Groups, the only condition not to show a significant difference in VOT was the velar-initial singleton onset. The non-native speakers' labial-initial singleton VOTs were significantly shorter than those of the native speakers ($t(6.48)=2.87$, $p<.05$), while the non-native speakers' C1C2 clusters exhibited systematically longer VOT than those of the native speakers at both the labial ($t(6.50)=3.55$, $p=.01$) and velar ($t(6.42)=7.45$, $p<.01$) Places of Articulation.

4. DISCUSSION

We assessed gestural coordination in the onset clusters of L1-English, L2 speakers of Spanish, comparing them with those of native speakers. We predicted that the L2 speakers would approximate the coordination patterns of the native speakers reasonably well, based on the advanced level of proficiency demonstrated by the participants and on the more restricted onset phonotactics of Spanish. We also expected that language-specific differences in segmental inventories and typical coordination patterns, along with the pervasive influence of the L1, could limit the precision with which the L2 speakers approximated L1 speech.

Our findings supported these hypotheses. The absence of a main effect of native language both on articulatory overlap and on VOT is fully consistent with the claim that the L2 speakers' gestural coordination patterns approximate those of the L1 with a reasonable degree of accuracy, reflecting their advanced level of proficiency. Whether learning of an L2 that has more complex onset phonotactics than a speaker's L1 would permit this degree of accuracy remains to be tested (though see work by Davidson and colleagues [10,12]). The presence of differences beyond these main effects demonstrates a limitation on the precision of this approximation.

Among these differences, some are attributable to L1 coordination patterns. First, L2 speakers showed a more substantial difference than L1 speakers in overlap between voiced labials and velars (see Fig. 1, left boxplot pairs in both panels). This difference is partially due to greater average overlap (lower IPIs) in the L2 speakers' voiced labial-initial clusters compared with those of the native speakers. In turn, this greater overlap is consistent with C-centre gestural organization—a characteristic of English. Second, the greater effects and interactions of Place of Articulation among the non-native speakers compared with the native speakers is consistent with articulator effects reported elsewhere in the literature on intra-oral and oral-laryngeal gestural coordination in English [23, 22]. Thus, the non-native speakers' place-modulated variability in overlap (Fig. 1, right panel) and VOT (Fig. 2, right panel) can be attributed to transfer from their L1.

In addition to these L1 transfer effects in the L2 coordination patterns, we see that subtle distinctions in overlap distinguishing the native speakers' voiceless labial and velar segments are absent for the non-native speakers. We can be sure that this level of articulatory detail is well below that which receives any attention in the language-learning classroom. However, our L1 data show that these intricately subtle distinctions form part of the systematic coordination of native speakers. From the present data, it is not possible to determine whether this distinction is perceptible at all, whether it is perceptible without these particular non-native speakers perceiving it, or whether they perceive it without the pattern having transferred to production.

Finally, and in distinction to the absence of native-like effects among non-native speakers, we also observe the presence of effects in non-native speakers that are absent for native speakers. The difference between the VOTs of the native and non-native speakers' labial singleton onsets is particularly striking in this regard. One might expect that native English speakers would produce longer voiceless stop VOTs in L2 Spanish due to transfer from the L1, when, in fact, we see the reverse. We attribute this difference to exaggeration of a L1-L2 difference with which the L2 speakers are particularly familiar (given that aspiration is one of the aspects of consonantal articulation that receives attention in the L2 Spanish language classroom). Thus, while the range of the native speakers' VOTs for labial singletons extends relatively high, those of the non-native speakers remain, exaggeratedly close to zero. Note, in comparison, the substantially higher non-native VOTs in clusters, when additional oral gestures are appended to the segmental sequence (from singleton C1V, to cluster C1C2V).

Although pairwise timing patterns (e.g., oral-laryngeal coordination in VOT) may be learned to some level of expertise, adding gestures (from C1V to C1C2V) seems to induce a shift towards familiar L1 patterns, even among advanced L2-speakers.

Moving forward, we plan to broaden our sample of L2-Spanish speakers to assess the generalizability of the presently reported findings, and to assess a greater range of proficiency, thus charting the trajectory from novice to expert speaker. Another way to augment this research is to investigate how precisely the English coordination patterns of a parallel set of native Spanish speakers' approximate those of native English speakers.

5. ACKNOWLEDGEMENT

This work has been supported both by the European Research Council (AdG 249440, https://erc.europa.eu/) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project number 317633480—SFB 1287, Project C04. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

6. REFERENCES

[1] Bates, D., Kliegl, R., Vasishth, S., Baayen, R. 2015. Parsimonious mixed models. Available from *arXiv*:1506.04967.

[2] Bates, B., Mächler, M., Bolker, B., Walker, S. C. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.

[3] Bombien, L., Hoole, P. 2013. Articulatory overlap as a function of voicing in French and German consonant clusters, *JASA* 134(1), 539–550.

[4] Bradley, T. 2006. Spanish complex onsets and the phonetics–phonology interface. In Optimality-Theoretic Studies in Spanish Phonology, F. Martinez Gil & S. Colina (Eds.), 15–38. Amsterdam: John Benjamins.

[5] Browman, C. P., Goldstein, L. 1986. Towards an articulatory phonology. *Phon. Yearbook* 3, 219–252.

[6] Browman, C. P., Goldstein, L. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5, 25–34.

[7] Byrd, D., Krivokapic, J., Lee, S. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *JASA* 120, 1589–1599.

[8] Byrd, D., Tobin, S., Bresch, E., Narayanan, S. 2009. Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *JPhon* 37(1), 87–110.

[9] Council of Europe. 2011. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Council of Europe.

[10] Davidson, L. 2005. Addressing phonological questions with ultrasound. *Clin. Ling. & Phon.* 19(6/7), 619–633.

[11] Davidson, L., Roon, K. 2008. Durational correlates for differentiating consonant sequences in Russian. *JIPA* 38(2), 137–165.

[12] Davidson, L., Stone, M. 2003. Epenthesis versus gestural mistiming in consonant cluster production. *Proc. WCCFL* 22, 165–178.

[13] Flege, J. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *JPhon* 15(1), 47–65.

[14] Gafos, A. (2002). A grammar of gestural coordination. *NLLT* 20(2), 269–337.

[15] Gafos, A., Hoole, P., Roon, K. & Zeroual, C. 2010. Variation in timing and phonological grammar in Moroccan Arabic clusters. *LabPhon* 10: Berlin/New York: Mouton de Gruyter.

[16] Gibson, M. Sotiropoulou, S. Tobin, S., Gafos, A. submitted. Intergestural and oral-laryngeal timing in word initial Spanish onset clusters.

[17] Gibson, M. Sotiropoulou, S. Tobin, S., Gafos, A. 2017. On some temporal properties of Spanish consonant-liquid and consonant-rhotic clusters. *Proc. 13th Conf. P&P in German-speaking countries*, 73–76.

[18] Goldstein, L., Byrd, D., & Saltzman, E. 2006. The role of vocal tract gestural action in units understanding the evolution of phonology. In M. Arbib (ed), *Action to language via the mirror neuron system*. Cambridge: Cambridge University Press, 215–249.

[19] Hermes, A., Mücke, D., Grice, M. 2013. Gestural coordination of Italian word-initial clusters: the case of 'impure s'. *Phonology* 30(1), 1–25.

[20] Kuznetsova, A., Brockhoff, P., Christensen, R. 2015. *lmerTest: Tests in linear mixed effects models.* Pkg vers. 2.0–20, Vienna: R Found. for Stat. Comp.

[21] Lenz, R. 1892. Chilenische Studien. *Phonetische Studien* 5, 272–293.

[22] Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20(3), 384–422.

[23] Marin, S. & Pouplier, M. 2010. Temporal organization of complex onsets and codas in American English: testing the predictions of a Gestural Coupling Model. *Motor Control* 14, 380–407.

[24] Navarro Tomás, T. 1918. Diferencias de duración entre las consonantes españolas. *Revista de Filología Española* 5, 367–393.

[25] Quilis, A. 1970. El elemento esvarabático en los grupos [pr, br, tr...]. In *Phonétique et linguistique romanes: Mélanges offerts à M. G. Straka*, 99–104. Lyon-Strasbourg: Société de Linguistique Romane.

[26] R Core Team (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Found. for Stat. Computing. URL: http://www.R-project.org/.

[27] Rao, R. 2019. *Key Issues in the Teaching of Spanish Pronunciation*. London: Routledge.

[28] Shaw, J., Gafos, A., Hoole, P., Zeroual, C. 2011. Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology* 28, 455–490.

[29] Sotiropoulou, S., Gibson, M., Tobin, S. Gafos, A. 2018. Temporal stability patterns of stop-liquid and stop-rhotic clusters in Spanish. *Proc. 16th Conf. on Lab. Phon*, 372.

[30] Tiede, M. (2005). Mview: Software for visualization and analysis of concurrently recorded movement data. New Haven, CT: Haskins Laboratories.

[31] Tobin, S. *Phonetic Accommodation in Spanish-English and Korean-English Bilinguals: A Dynamical Account.* (Unpubl. doctoral diss.). Storrs, CT: UConn.