EXAMINING THE EFFECTS OF STRESS ON VOWEL PRODUCTION IN HERITAGE VS. MONOLINGUAL SPANISH SCHOOL-AGED CHILD SPEAKERS

Gemma Repiso-Puigdelliura[1], Daniel Pape[1]

[1]McMaster University
repisopg@mcmaster.ca, paped@mcmaster.ca

ABSTRACT

While Spanish adult heritage speakers (i.e., HSs) show vocalic reduction in unstressed position, it is still not known whether vowel reduction is already present as HSs develop their language speech-specific patterns. We examined the effect of lexical stress on 24 primary school-aged child HSs (mean age = 8;11 ± 2) and compared their non-high vowel productions to those of 12 Spanish monolingual speakers (mean age = 8;12 ±1;6). Our results showed that non-high vowels in disyllabic words are produced higher in unstressed positions when compared to stressed positions. In addition, *stress* interacted with *group* (HS vs monolingual) for the /e/ vowel. Along the front-back dimension, our findings only show a cross-over interaction between *stress* and *group* for /a/. For all other vowels, no significant differences were found between HSs and monolinguals. The lack of vowel reduction during childhood may be due to overall late development of vowel reduction patterns.

Keywords: lexical stress, child heritage speakers, vowel reduction

1. INTRODUCTION

Heritage speakers (HSs) are early bilinguals that are exposed to their family language (i.e., heritage language) simultaneously or sequentially with the majority language of the society [1]. While speech production in HSs presents more target-like characteristics than that of L2ers [2], HSs' productions have been shown to diverge from canonical descriptions of the language [3]. It is still unclear, however, whether such divergences arise during speech development due to early grammar interaction or whether HSs undergo attrition as their input to the heritage language is reduced [4]. To investigate this question, it is important to examine HSs' speech production during their primary school years as HSs gain systematic exposure to the majority language and may undergo a shift in their dominance from the heritage to the majority language [5].

Early accounts of the Spanish vowel system (i.e., /a e i o u/) describe it as stable and relatively unaffected by lexical stress [6]–[8]. It is worth mentioning, however, that studies on dialectal variation have found that in dialects such as Andean Spanish, Mexican Spanish, or Peninsular Spanish unstressed vowels can vary from stressed vowels with respect to duration and vowel quality [9]–[11].

Unlike in Spanish, in English vowel reduction is a phonological phenomenon, and vowel quality is, therefore, more consistently affected by stress [12]. Unstressed English vowels are more central than stressed vowels and are produced with less effort [6], [12], [13]. When in contact with English, Spanish HSs show vowel centralization in unstressed positions [14], [15]. Ronquest [14] found that unstressed vowels are shorter than stressed vowels and demonstrate a more reduced vocalic space. This reduction, however, is not in the direction of the centralized schwa. The vowels /e a o/ in the above-mentioned study had lower F1 values and thus were produced higher in unstressed position. All unstressed vowels except for /a/ showed centralization in the F2 dimension. Similarly, Elias et al. [15] demonstrated that unstressed /a/, /e/, and /o/ have lower F1 values than their stressed counterparts. Along the F2 dimension, their results show an interaction between lexical stress and vowel phoneme, indicating that stressed /e/ and /i/ are more fronted than unstressed /e/ and /i/, whereas stressed /o/ and /u/ are more posterior than their unstressed counterparts. Moreover, vowels produced in code-switched mode are, overall, more centralized than vowels produced in monolingual mode. To summarize, adult Spanish HSs present a more reduced vowel space in unstressed position, which is characterized by lower F1 values overall and more compressed F2 values, with respect to the stressed position. To our knowledge, however, no studies have investigated whether a similar reduction is present earlier in childhood during the development of the HSs' two phonologies.

In this study, in order to examine whether the reduction patterns found in previous studies on adult HSs would appear already during the primary school years, we explore the production of /a/, /o/, and /e/ in the semi-spontaneous speech of a group of child HSs and compare the results to that of a group of monolingually-raised Spanish speakers.

2. METHODS

2.1. Participants

24 child HSs (12 M, 12 F, mean age = 8;11 ± 2, age range = 5;3 to 11;11) raised in California with two caregivers from Mexico (except for 2 participants with one caregiver from Mexico and one English-speaking caregiver [Australia and Los Angeles], and 3 participants with one caregiver from Mexico and one caregiver from El Salvador) and 12 monolingually-raised child Spanish speakers from Mexico (i.e., SpanMonoSs, 8 M, 4 F, mean age = 8;12 ±1;6, age range = 6;1 to 11;1) participated in this study. A parental background questionnaire showed that the child HSs and the SpanMonoSs significantly differed in the amount of Spanish relative to English to which they were exposed at home and at school ($t(30.87) = -4.93$, $p < 0.001$), and in the amount of Spanish relative to English that they produced at home ($t(26.53) = -7.22$, $p < 0.001$). In particular, relative to English, SpanMonoSs were exposed to more Spanish at home and at school (M = 85.9%, SD = 14.6%) than child HSs (M = 55.65%, SD = 21.8%), and produced more Spanish at home (M = 96.9%, SD = 5.7%) than child HSs (M = 53.0%, SD = 28.7%).

2.2. Task

Semi-spontaneous speech was elicited using the wordless book *Frog, where are you?* [16]. Participants were asked to narrate the story as naturally and detailed as possible. The session was conducted on Zoom by the first author. During the session, caregivers were asked to record the participants' speech using a smartphone positioned 4 inches from the children's mouths. Recordings were done using the ShurePlusMotiv, an app that records with a sampling rate of 48.1 kHz and a sample size of 16 bits in an uncompressed audio format.

2.3. Formant extraction

Disyllabic content words[1] were extracted from the data and the formant frequencies F1 and F2 were measured at the acoustic midpoint of the vowels using PRAAT's formant extraction algorithm ("To Formant... burg") with adapted parameter settings for children's vocal tracts (max. formant: 8000 Hz; max. number of formants: 5; window length: 25 ms; pre-emphasis from 50 Hz) as a first approximation of the correct formant value. The automatically extracted spectral maxima (i.e., formant candidate) of the described Praat algorithm were then carefully manually corrected as needed based on the full LPC spectra (as main source), but also guided by FFT broadband spectra and the FFT broadband spectrograms. Thus, for each timepoint, the automatic measurement value (obtained by Praat's (adapted) automatic formant extraction algorithm) was manually checked and, if necessary, corrected, based on (1) the temporal formant movement information visible in the spectrogram, (2) the spectral information visible in the broadband FFT spectral slice (5ms window length, identical to the spectrogram settings) and (3) the spectral information visible in a generated LPC spectral slice (Praat LPC object with LPC order = 16, sampling rate of the underlying audio file = 16000 Hz). This hybrid approach of measuring formants for notoriously difficult child speech data based on the combination of the three methods, but mainly relying on the LPC spectra and maxima, ensures that eventual automatic formant tracking errors are noticed and thus manually corrected. Vowels were then normalized[2] using the Lobanov method of vowel normalization using the R package phonR [17].

4. RESULTS

A total of 3226 non-high vowels[3] were extracted from the semi-spontaneous speech. Six linear mixed effects models for the vowels /a/ /e/ /o/ in the Lobanov transformed F1 and F2 dimensions were run using the lme4 package [18] and p-values were obtained with the lmerTest package [19]. The variables *stress* (i.e., stressed vs. unstressed) and *type of speaker* (child HSs vs. SpanMonoSs) were entered as fixed effects, and the intercepts for participant and word were allowed to vary.

4.1. Results for /a/

With regard to /a/ (See Fig. 1), the model for F1 showed that unstressed vowels were produced with lower values of F1 (i.e., higher) than stressed vowels ($\beta = -0.32$, $SE = 0.12$, $t = -2.59$, $p < 0.001$, CI 95% [-0.57, -0.08]). No main effect for *type of speaker*, or interaction between *type of speaker* and *stress* was found. The model for F2 showed a significant effect for *stress* ($\beta = -0.43$, $SE = 0.13$, $t = -3.34$, $p < 0.001$, CI 95% [-0.68, -0.18]) and a significant cross-over interaction was found between *stress* and *type of speaker* ($\beta = 0.68$, $SE = 0.14$, $t = 4.71$, $p < 0.001$, CI 95% [0.09, 0.34]), indicating that the direction of the effect of *stress* on F2 depends on the variable *type of speaker*: In the child SpanMonoSs, /a/ is more fronted in the stressed condition than in the unstressed condition, whereas in the child HSs /a/ is more fronted in the unstressed condition than in the stressed condition.

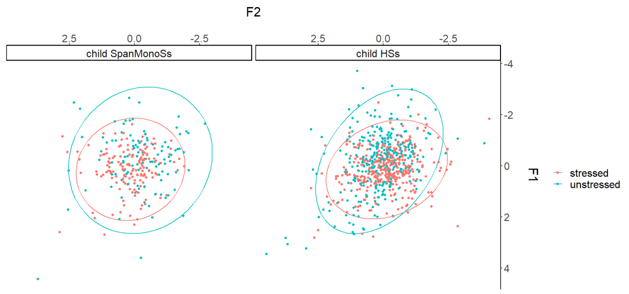

Figure 1: Lobanov-normalized F1 and F2 values for /a/ by variables *type of speaker* and *stress*. The data ellipses expect to contain 95% of the population distribution.

4.2. Results for /e/

As for /e/ (See Fig. 2), the model for F1 showed that unstressed vowels were produced with lower F1 values (i.e., higher) than stressed vowels (β = -0.78, *SE* = 0.21, *t* = -3.56, *p* < 0.001, CI 95% [-1.21, -0.39]), and that *stress* interacted with *type of speaker* (β = -0.47, *SE* = 0.22, *t* = 2.15, *p* = 0.03, CI 95% [0.02, 0.87]), but no significant effect of *type of speaker* was found. Post-hoc analyses indicated that while *stress* significantly shifted F1 on the child SpanMonoSs (β = 0.78, *SE* = 0.22, *t* = 3.48, *p* < 0.001), it did not do so for the child HSs (β = 0.31, *SE* = 0.19, *t* = 1.61, p = 0.10). No main effect of *type of speaker* was found. The model for F2 did not show any significant effects for any of the variables or their interaction.

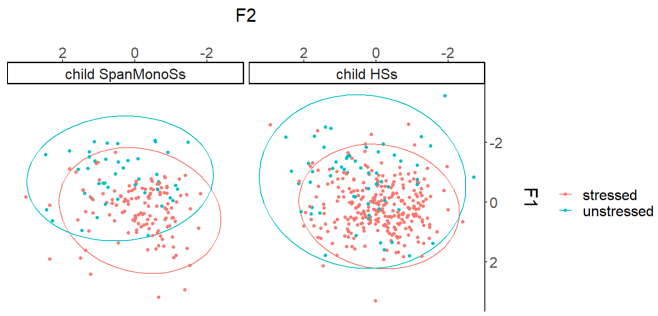

Figure 2: Lobanov-normalized F1 and F2 values for /e/ by *type of speaker* and *stress*.

4.3. Results for /o/

Concerning the results for /o/ (See Fig. 3), the model for F1 demonstrated that unstressed vowels were produced with lower F1 values (i.e., higher) than stressed vowels (β = -0.28, *SE* = 0.09, *t* = -2.88, *p* = 0.004, CI 95% [-0.49, -0.07]). No effect of *type of speaker* nor interaction between *type of speaker* and *stress* was found. No significant differences were found along the F2 dimension.

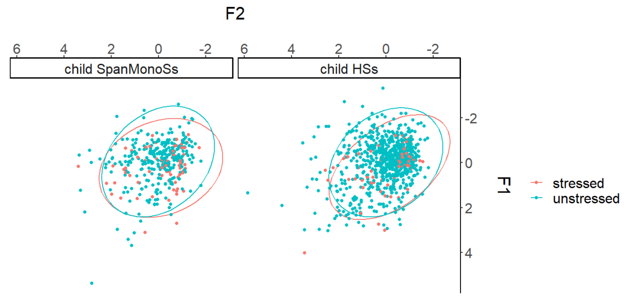

Figure 3: Lobanov-normalized F1 and F2 values for /o/ by *type of speaker* and *stress*.

5. DISCUSSION

5.1. General effects of stress on F1 and F2

Similar to previous studies on HSs' vowel production [14], [15], our results demonstrate that, overall, non-high vowels (i.e., /a e o/) in unstressed position are produced with lower F1 values (i.e., higher) when compared to those in stressed position. Unlike Ronquest [14] and Elias et al., [15], however, the front-back dimension (i.e., F2) was not significantly affected by *stress* in our data.

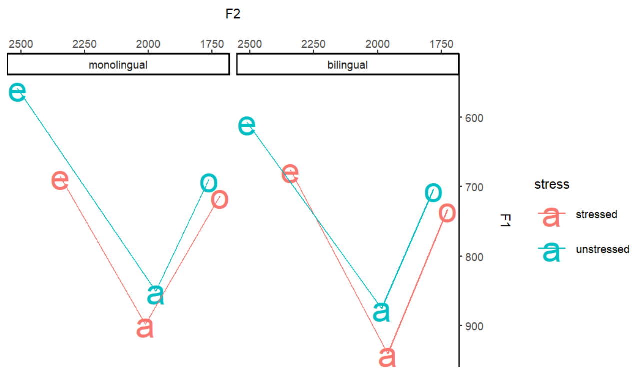

Figure 4: Mean Lobanov-normalized F1 and F2 values by vowel, *stress*, and *type of speaker*

Despite our unstressed vowels shifting towards a more reduced acoustic space (see Fig. 4), the lower F1-values for unstressed /e/ and /o/ and the lack of significant movement along the F2 dimension do not suggest that Spanish-speaking children centralize non-high vowels to /ə/. Instead, it is possible that F1 effects are a result of prominence-related universal phonetic principles. Prominent vowels are associated with a lower jaw position [20] and more extreme articulatory movements [21], [22], which could lead to more open vowel productions in general and more peripheral F1 values [21], [23]. Two strategies support an enhanced production of accented vowels. First, the 'sonority expansion strategy' states that speakers adopt a more open vocal tract to create a larger oral passage which results in louder and thus more open vowels [22]. The 'localized hyperarticulation', in turn, posits that prominence affects tongue movement, by which vocal targets become more distinct and contrasts between vowels are strengthened [24]. In sum, we propose that the higher F1 values in stressed vowels compared to unstressed vowels, seen for both monolingual and HSs groups, could be a result of increased intensity in prominent vowels attained through larger jaw opening and lowering of the tongue dorsum.

5.2. Child heritage speakers compared to Spanish monolinguals

Contrary to our predictions, our findings did not show that child HSs produced unstressed-stressed vowels with a greater difference than that of child SpanMonoSs. Our model for /a/ F2 showed an interaction between the type of speaker and stress, suggesting that, while unstressed vowels produced by child HSs move forward on the F2 dimension, unstressed vowels produced by SpanMonoSs move backward. It is possible that the Spanish /a/ from the child HSs is affected by the English vocalic space, in that /ɑ/ (i.e., stressed) is more backward than the Spanish /a/ [25], and /ə/ is more forward than its stressed counterpart. In addition, the fronted English /æ/ could further restrict the available acoustic space that child HSs have in Spanish. Our model for /e/ F1 showed that only the child SpanMonoSs, but not HSs, produced lower stressed /e/ vowels compared to unstressed ones, with no significant effect between the two populations. It is hypothesized that an expected prominence effect is constrained in the child HSs due to the existence of /ɛ/ in their English acoustic vowel space. In principle, we should see the same interaction for the /o/ vowel[4]. However, the COT-CAUGHT merger in Californian English [26], [27] and the limited number of stressed /o/ tokens in our data set (N = 147) could explain the lack of a significant interaction. Our results, hence, do not provide evidence for early transfer of vowel reduction from English to Spanish in HSs. To explain why our results diverge from those of adult HSs, we could posit that English reduction patterns are still in development among the child HSs in this study (mean age = 8;11 ± 2), and thus, cannot be transferred to the HL at this point in their development. In fact, patterns of reduction are characteristic of expert motor control [28], [29] and some of them can extend until the primary school years. For example, Goffman [30] found that function words are less reduced in children up to the age of 7 than in adults. Future research should include an examination of the English vowels of the HSs to determine whether the children at the time of testing have acquired English-like vowel reduction patterns.

Alternatively, our findings could be a result of the articulatory costs of vowel reduction. The late-developing nature of reduction patterns may render vowel centralization taxing from a processing and articulatory perspective during early speech production. In turn, recent accounts on the structure of heritage language grammars suggest that processing costs may determine whether an area of the HL is vulnerable to language transfer or grammar restructuring [31]. Given the potentially higher costs for reducing vowels, a possible pressure of the majority language grammar will be less likely to reorganize the Spanish phonologies into adopting vowel reduction from the HSs' English phonologies. Future research should investigate vowel reduction using a longitudinal design to better understand the period at which HSs start showing the patterns of vowel reduction found in adult HSs.

6. REFERENCES

[1] G. Valdés, "Heritage language students: Profiles and possibilities," in *Handbook of Heritage, Community, and Native American Languages in the United States: Research, Policy, and Educational Practice*, T. G. Wiley, J. K. Peyton, S. C. K. Moore, D. Christian, and N. Liu, Eds., Routledge, 2014, pp. 41–49. doi: 10.4324/9780203122419.

[2] L. M. Knightly, S.-A. Jun, J. S. Oh, and T.-K. Au, "Production benefits of childhood overhearing," *J Acoust Soc Am*, vol. 114, no. 1, pp. 465–474, 2003, doi: 10.1121/1.1577560.

[3] R. Ronquest and R. Rao, "Heritage Spanish phonetics and phonology," in *The Routledge Handbook of Spanish as a Heritage Language*, K. Potowski, Ed., London/New York: Routledge, 2018, pp. 164–177.

[4] M. Polinsky, "Bilingual children and adult heritage speakers: The range of comparison," *International Journal of Bilingualism*, vol. 22, no. 5, pp. 547–563, 2018, doi: 10.1177/1367006916656048.

[5] S. Montrul, "Heritage language development: Connecting the dots," *International Journal of Bilingualism*, vol. 22, no. 5, pp. 530–546, Oct. 2018, doi: 10.1177/1367006916654368.

[6] P. Delattre, "An acoustic and articulatory study of vowel reduction in four languages," *IRAL - International Review of Applied Linguistics in Language Teaching*, vol. 7, no. 4, 1969, doi: 10.1515/iral.1969.7.4.295.

[7] T. Navarro Tomás, *Manual de pronunciación española*, 12th ed. Madrid: CSIC, 1918.

[8] A. Quilis and M. Esgueva, "Realización de los fonemas vocálicos españoles en posición fonética normal," in *Estudios de fonética*, M. Esgueva and M. Cantarero, Eds., Madrid: Consejo Superior de Investigaciones Científicas, 1983, pp. 159–251.

[9] A.-M. Delforge, "Unstressed vowel reduction in Andean Spanish," in *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, L. Colantoni and J. Steele, Eds., Sommerville, MA: Cascadilla Proceedings Project, 2008, pp. 107–124.
[10] R. Marín Gálvez, "La duración vocálica del español," *E.L.U.A.*, vol. 10, pp. 213–226, 1994.
[11] J. M. Lope Blanch, "En torno a las vocales caedizas del español mexicano," *Nueva Revista de Filología Hispánica (NRFH)*, vol. 17, no. 1/2, 1964, doi: 10.24201/nrfh.v17i1/2.1507.
[12] M. Fourakis, "Tempo, stress, and vowel reduction in American English," *Journal of the Acoustical Society of America*, vol. 90, no. 4, 1991, doi: 10.1121/1.401662.
[13] W. R. Tiffany, "Nonrandom sources of variation in vowel quality.," *J Speech Hear Res*, vol. 2, 1959, doi: 10.1044/jshr.0204.305.
[14] R. Ronquest, "An Acoustic Examination of Unstressed Vowel Reduction in Heritage Spanish," in *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, C. Howe, S. Blackwell, and M. Lubbers Quesada, Eds., Cascadilla Proceedings Project, 2013, pp. 151–171.
[15] V. Elias, S. McKinnon, and Á. Milla-Muñoz, "The effects of code-switching and lexical stress on vowel quality and duration of heritage speakers of Spanish," *Languages*, vol. 2, no. 4, p. 29, 2017, doi: 10.3390/languages2040029.
[16] M. Mayer, *Frog Where Are You?* New York: Dial Press, 1969.
[17] D. R. Mccloy, "Vowel normalization and plotting with the phonR package," *Technical Reports of the UW Linguistic Phonetics Laboratory*, vol. 3, 2012.
[18] D. Bates, D. Sarkar, M. D. Bates, and L. Matrix, "The lme4 Package," *October*, vol. 2, no. 1, 2007.
[19] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *J Stat Softw*, vol. 82, no. 13, pp. 1–26, 2017, doi: 10.18637/JSS.V082.I13.
[20] M. E, Beckman, and J. Edwards, "Articulatory evidence for differentiating stress categories," in *Phonological Structure and Phonetic Form*, 2010. doi: 10.1017/cbo9780511659461.002.
[21] D. Erickson, "Articulation of Extreme Formant Patterns for Emphasized Vowels," 2002.
[22] K. de Jong, M. E. Beckman Ohio, and J. Edwards, "The interplay between prosodic structure and coarticulation," 1993.
[23] P. A. Keating, "Comments on the jaw and syllable structure," *J Phon*, vol. 11, no. 4, 1983, doi: 10.1016/s0095-4470(19)30839-3.
[24] K. J. Jong, "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *Journal of the Acoustical Society of America*, vol. 97, no. 1, 1995, doi: 10.1121/1.412275.
[25] A. R. Bradlow, "A comparative acoustic study of English and Spanish vowels," *Journal of the Acoustical Society of America*, vol. 97, no. 3, pp. 1916–1924, 1995, doi: 10.1121/1.412064.
[26] P. Eckert, "Where do ethnolects stop?," in *International Journal of Bilingualism*, 2008. doi: 10.1177/13670069080120010301.
[27] A. A. Metcalf, "Directions of Change in Southern California English," *J Eng Linguist*, vol. 6, no. 1, 1972, doi: 10.1177/007542427200600104.
[28] M. A. Redford, "The Acquisition of Temporal Patterns," in *The Handbook of Speech Production*, 2015. doi: 10.1002/9781118584156.ch17.
[29] V. Kapatsinski, S. Easterday, and J. Bybee, "Vowel reduction: A usage-based perspective," *Italian Journal of Linguistics*, vol. 32, no. 1, 2020, doi: 10.26346/1120-2726-146.
[30] L. Goffman, "Kinematic differentiation of prosodic categories in normal and disordered language development," *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 5, 2004, doi: 10.1044/1092-4388(2004/081).
[31] M. Polinsky and G. Scontras, "Understanding heritage languages," *Bilingualism: Language and Cognition*, vol. 23, no. 1, pp. 4–20, Jan. 2020, doi: 10.1017/S1366728919000245.
[32] J. I. Hualde, "Unstressed words in Spanish," *Language Sciences*, vol. 31, no. 2–3, pp. 199–212, Mar. 2009, doi: 10.1016/j.langsci.2008.12.003.

[1] In Spanish, some function words are considered to be unstressed (e.g., *menos amigos* 'except for friends') [32]. Thus, to avoid confounds, we decided to exclude function words from our data. Similarly, Elias et al. [15] excluded high-frequency words (definite articles, possessives, prepositions, conjunctions).

[2] In order to determine if vowel normalization would give different results compared to non-normalized children formant data, we run all analyses of this paper also on non-normalized data. The results were all identical. Reasoning to either include or not include vowel normalization was that at the mean age of nine years no relevant differences between vocal tract sizes (i.e., male/female but also across children of the same gender) were assumed to be relevant and thus it was not clear whether speaker normalization would be necessary for this data. Since there were no differences, we report the normalized data here.

[3] Only non-high vowels were included in the analysis given the small number of tokens for /i/ (N of unstressed = 23, N of stressed = 515) and /u/ vowels (N of stressed =26, N of unstressed = 29).

[4] That is, /o/ compared to /ɔ/ for the English grammars of the HSs. Despite the fact that [26] reports a rising in the merger, it is still possible that a raised merger does not exert the same influence on the Spanish acoustic space as /ɛ/ could.