Use of suprasegmental information in the perception of Spanish lexical stress by Spanish heritage speakers of different generations

Ji Young Kim

Department of Spanish, Italian, and Portuguese
University of Illinois at Urbana-Champaign, USA
jkim315@illinois.edu

Abstract

The present study examines the perception of Spanish lexical stress by Spanish heritage speakers of different generations and compares their performance to that of Spanish native controls and English second language (L2) learners of Spanish. Previous studies have shown that English L2 learners experience great difficulty in perceiving Spanish lexical stress. Such difficulty is argued to be derived from English listeners using different strategies from Spanish listeners in the perception of stress. Given that Spanish heritage speakers share the same dominant language with English L2 learners (English), but differ from them with regard to the first language (Spanish), the present study intends to seek whether heritage speakers show similar or different patterns when compared with L2 learners. The present study also intends to account for the heterogeneity among heritage speakers by comparing heritage speakers of different generations. Using a forced-choice identification task with stressed minimal pairs of paroxytone and oxytone verbs, results showed that while 1st generation US-born heritage speakers pattern like Spanish native controls by paying more attention to the acoustic cues of the stimuli, 1.5/2nd generation US-born heritage speakers pattern like English L2 learners by showing bias towards paroxytone verbs.

Index Terms: speech perception, lexical stress, heritage language phonology

1. Introduction

Broadly speaking, heritage speakers are people who grow up exposed to both a majority language and an ethnic minority language. Despite their heterogeneity, heritage speakers share some characteristics in common. Generally speaking, heritage speakers, more specifically heritage speakers in the US, are early bilingual speakers of English (majority language) and a non-English home language (minority language) who have lived most or all of their lives in the US. The parents of these speakers are native speakers of the home language (heritage language), thus, heritage speakers acquire this language as their first language (L1) and, for many of them, systematic exposure to English, which is their second language (L2), does not occur until they enter institutional settings such as kindergarten and elementary school. Since English is the majority language and the heritage language is a minority language, heritage speakers' use of English increases as they grow up, whereas their use of the heritage language becomes limited to familial settings. This subsequently results in a gradual shift of language dominance from the heritage language (L1) to English (L2). Therefore, while heritage speakers have very strong command of English, their command of the heritage language is usually short of the native speaker level of their parents or peers raised in their home countries [9]. Heritage speakers are similar to L2

learners in that they are more dominant in English than the heritage language, but are different from them in that while heritage speakers learn the heritage language as the L1, L2 learners learn it as a L2. Studies on L2 phonology have shown that English L2 learners of Spanish experience great difficulty perceiving Spanish lexical stress [11]. Thus, the present study intends to see whether Spanish heritage speakers show similar patterns to English L2 learners when they perceive Spanish lexical stress.

2. Lexical stress in Spanish and English

Stress may be defined as prominence in a syllable of a word resulting from extra muscular energy that is acoustically manifested in higher pitch, longer duration, and higher intensity, in some cases with segmental effects [7]. Although researchers agree that these three universal parameters are important indices of stress, languages differ from one another with regard to what parameter functions as the primary correlate of stress. Spanish and English are typologically similar in that stress is phonologically contrastive in both languages, but they differ in the realization of stress [5]. With regard to the acoustic correlates of stress in the two languages, recent studies, have shown that when pitch accent is controlled, the primary correlate of stress is duration in Spanish [10], while it is vowel quality in English [1]. Unlike English, in which unstressed vowels undergo vowel reduction, resulting in a schwa [ə], in Spanish vowels maintain their vowel quality regardless of whether they are stressed or unstressed [5]. Thus, Spanish listeners would not be sensitive to changes in vowel quality, but rather to changes in duration and other suprasegmental cues to identify stressed vowels.

English L2 learners of Spanish are shown to have great difficulty in identifying the position of stress in Spanish even after explicit instructions [11], most likely due to influence from their native language (English). Although lexical activation is sensitive to all the information that is available in speech signals, whether it is segmental or suprasegmental, it is likely that listening strategies that are applied in the native language influence the way people listen to foreign language input [2]. English is more segmentally-based [12] and generally suprasegmental cues only provide redundant information [3]. Thus, English listeners are not accustomed to paying attention to suprasegmental information when this is the only cue available in the speech signal, unlike Spanish listeners, who are sensitive to even small differences in suprasegmental cues [3]. Therefore, the goal of the present study is to examine whether Spanish heritage speakers who are dominant in English also show difficulty in perceiving stress contrasts in Spanish due to influence from English, and whether differences are found among heritage speakers, who are found to be highly heterogeneous. Among various factors, the present study will focus on the effect of heritage speakers' generation on their perception of Spanish lexical stress.

3. Experiment

3.1. Participants

In total, 89 subjects participated in the present study: 25 Spanish native speakers (NS) (18F, 7M), 17 Spanish heritage speakers (HS) (13F, 4M), and 47 English L2 learners of Spanish (L2) (36F, 11M). The NSs were recruited in a North-Central region of Mexico. They reported that, although they have learned languages other than Spanish, this did not happen until later in life and they use only Spanish most of the time (avg. 93%). Thus, even though these speakers are strictly speaking not monolinguals, the present study will consider them as such, because they do not use the other languages functionally. The L2s are beginner to intermediate-level Spanish major or minor students recruited from a Spanish grammar course at a university in the Midwest, US. The L2 learners reported that they did not learn Spanish before the age of 9 (avg. 11.62 years) and their daily use of Spanish is less than 10% (avg. 7.34%). The HSs were born and raised in the US, mostly in the Chicago area, to Mexican families. Based on their (1) age of acquisition, (2) language use, and (3) language proficiency, it was determined that the heritage speakers were all English-dominant. That is, the heritage speakers (1) are early bilinguals in that they acquired both Spanish and English before the age of 5, which is before the period when foreign accent starts to appear if a language is not learned by then [4], (2) currently use English more frequently than Spanish, and (3) self-rated their proficiency in Spanish lower than English. The HSs were sub-divided into two groups based on whether they are 1st generation US-born Mexican-Americans (both parents are from Mexico) (HS1: 8F & 3M) (HS1) or 1.5 generation (only one parent is 1st generation US-born) or 2nd generation US-born Mexican-Americans (both parents are 1st generation US-born) (HS2: 5F & 1M) (HS2). 1.5 and 2nd generation heritage speakers were grouped together due to small number of participants (three speakers each) and based on the assumption that these speakers would behave differently from HS1s, given that they are exposed to Spanish less frequently than HS1s and they consider Spanish as a L2, unlike HS1s who consider it as a native language.

3.2. Test materials

32 stress minimal pairs that differ only in the location of lexical stress were used in the present study. The stress pairs consisted of disyllabic Spanish regular –ar verbs in the first person singular of the present indicative (e.g., *PAso* 'I pass') and the same verbs in the third person singular of the (simple) past perfective tense (e.g., *paSO* 'he/she/you(formal) passed'). The former case always has stress on the penultimate syllable (paroxytone) and the latter case always has it on the last syllable (oxytone). Apart from the target items, 68 fillers with verbs of both present and past tense were included. Each item was inserted in the second-to-last position of a meaningful sentence with narrow focus on the last word (subject). Whereas English has relatively fixed word order and greater flexibility in assigning the nuclear stress to words in different locations, Spanish has greater syntactic freedom and new information items are generally moved to the end of the sentence, where they receive nuclear stress [5, 6]. In the present case, by locating the subject in the last position, *Por la plaza paso yo* 'I pass through the plaza', the subject carries the narrow focus. Thus, by using such sentence structure, it was possible to separate lexical stress from pitch accent, given that

the word that carries the nuclear accent is the subject (*yo*), not the target word (*PAso*). All the stimuli were produced by a male native speaker of Mexican Spanish and were recorded in a sound-attenuating booth using an AKG C520 head-mounted microphone and a Marantz PMD570 solid state recorder. The last word of each sentence was removed, leaving the sentences incomplete (*Por la plaza paso…*). Since transitions always occurred from /o/ (back vowel) to either /e/ or /j/ (front vowel/glide), using the spectrogram display in *Praat*, the cut-off point was determined as the moment when the F2 value started to make a noticeable increase. After removing the last words, all the tokens were played to make sure no transition information was included in the speech signals.

3.3. Procedures

In order to avoid priming effect, two lists were used, each containing only one member of each of the minimal stress pairs, with paroxytones and oxytones randomly distributed between the two lists. Half of the participants completed one list and half of them completed the other. The participants sat in front of a computer and listened to the incomplete sentences through headphones. In Mexico, the stimuli were presented through a Dell Vostro 230 desktop computer with KOSS UR-20 headphones and, in the US, they were presented through a Samsung SENS R410 laptop computer with Sony MDR 7506 headphones. While listening to the stimuli, the participants saw two subject words (e.g., *yo* "I" vs. *él* "he") on each side of the computer screen. Then, they decided which of the two options the following word had to be by clicking on either the left or the right key, which were indicated with colored stickers on the keyboard. The incomplete sentences were presented in pseudo-randomized order and the order of the subject words on the screen was counter-balanced. The presentation of the stimuli was done using *E-Prime*.

4. Results and Discussion

Participants' accuracy and response time (RT) were automatically collected through *E-Prime*. With regard to accuracy, all correct responses were coded as "1" and all incorrect responses were coded as "0". The effect of Group (NS/HS1/HS2/L2), Stress Pattern (paroxytone/oxytone), and the interaction of the two factors on participants' Accuracy (correct/incorrect) was statistically analyzed using logit mixed effects modeling with subject and item as random factors. The *glmer* function in the *lme4* package in R was used for the analysis. The two fixed factors (Group and Stress Pattern) were centered using contrast-coding. The best fitting model according to backwards selection included random intercepts for subject and item with no slope terms. Figure 1 shows the accuracy rate for paroxytones, oxytones, and fillers of each group. All four groups performed close to ceiling when perceiving the fillers, which confirms that the errors that occurred could not be due to problems regarding Spanish verb conjugation, which could be considered as a possible confound due to the experimental design of the present study (i.e., identify the subject of the verb). Results showed that NSs (the baseline group) performed significantly better than HS2s (β = -2.1618, SE = 0.4121, z = -5.245, p < 0.001) and L2s (β = -2.8936, SE = 0.2812, z = -10.291, p < 0.001), while the difference between NSs and HS1s was only marginally significant (β = -0.6444, SE = 0.3744, z = -1.721, p = 0.0852). A significant main effect was found in Stress Pattern (β = 2.1644, SE = 0.6195, z = 3.494, p < 0.001), indicating that

overall participants were better at perceiving paroxytones. However, a significant interaction was also found between Stress Pattern and HS2 (β = -3.2335, SE = 1.1367, z = -2.845, p < 0.01) and between Stress Placement and L2 (β = -4.3969, SE = 0.7048, z = -6.239, p < 0.001). This suggests that NSs' accuracy difference between paroxytones and oxytones was significantly different from that of HS2s and L2s. As shown in Figure 1, while NSs performed slightly better when perceiving oxytones, HS2s and L2s performed worse in this stress pattern. Such difference in direction explains why significant interactions have occurred. With regard to HS1s, no significant interaction was found between Stress Pattern and HS1, suggesting that the accuracy difference between the two stress patterns was not different between NSs and HS1s. In fact, HS1s behaved similarly to NSs, with the accuracy rate of oxytones slightly higher than paroxytones.
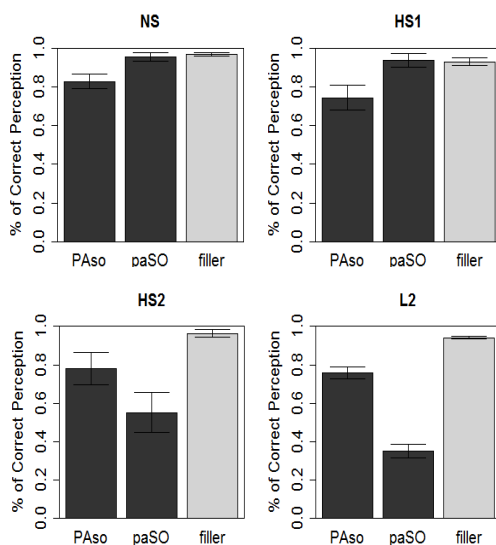

Figure 1: *Accuracy rate of paroxytones (left), oxytones (middle), and fillers (right) in each group*

It is interesting to note that while HS2s' and L2s' perception of paroxytones was target-like, their accuracy rates for oxytones were either close to chance level or lower, implying that HS2s and L2s have a bias towards selecting paroxytones. Participants' sensitivity and response bias towards paroxytones were analyzed by calculating their d-prime scores and response criterion (C scores), respectively. HIT was considered as the case in which participants selected a paroxytone when the stimulus was a paroxytone and FALSE ALARM was considered as the case in which participants selected a paroxytone when the stimulus was an oxytone. A one-way ANOVA with Group as independent variable was conducted on participants' d-prime scores and C scores. A main effect was found for both d-prime scores (F(3,85) = 54.13, p < 0.001) and C scores (F(3,85) = 19.31, p < 0.001). TukeyHSD post-hoc analyses showed that for both scores, the L2s differed from both the NSs and H1s (p < 0.001), while no significant difference was found between the L2s and H2s. Likewise, no significant difference was found between NSs and HS1s. Figure 2 shows that HS2s and L2s performed differently from NSs and HS1s in that their d-prime scores were close to zero, indicating that they had very low sensitivity in distinguishing the two stress patterns. The

negative C scores confirm that this low sensitivity is due to bias towards selecting paroxytones. It is important to note that the perception pattern shown in the HS2 group was not consistent for all speakers in this group. Among the 6 HS2s that participated, half of them showed a strong bias toward paroxytones (paroxytone: avg. 83.33% accuracy; oxytone: avg. 20.83% accuracy). Among the remaining three speakers, two showed similar accuracy rates in the two stress patterns (paroxytone: avg. 81.25%; oxytone: avg. 84.38%) and one showed 100% accuracy rate in paroxytones, while in oxytones the accuracy rate was only chance level (56.25%). According to the HS2s' language profile, the three participants that showed a strong bias toward paroxytones had not been exposed to Spanish before avg. 7.33 years of age, while the other three acquired Spanish from birth. This finding may explain why different patterns were found among HS2s.
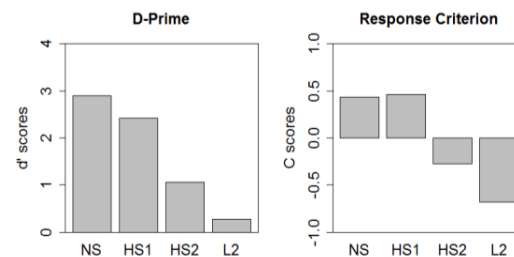

Figure 2: *D-prime (sensitivity) and response criterion (response bias) scores in each group*

Given that all the participants listened to the same stimuli, the different patterns found between NSs and HS1, on the one hand, and (half of) HS2s and L2s, on the other, can be explained through the way participants attended to the acoustic cues of the stimuli. Further analyses were conducted on the acoustic properties of the stimuli used in the experiment. Regarding the vowel quality of stressed and unstressed vowels, which is considered to be the critical acoustic cue in English, but not in Spanish, raw F1 and F2 values of the stressed and unstressed vowels were extracted and normalized using Lobanov normalization. The effect of Stress (stressed/unstressed) on the normalized F1 and F2 values was analyzed using linear mixed effects modeling with item as random factor. Tokens that were produced with creaky voice were removed from the analysis. Results showed that there was no significant effect of Stress in any of the three vowels used in the study (/a/, /i/, and /u/). Thus, vowel quality could not have had an effect on participants' response.

Apart from vowel quality, the differences in pitch, duration, and intensity between stressed and unstressed vowels were measured by subtracting the pitch, duration, and intensity values of the unstressed vowel from those of the stressed vowel in each item. As Figure 3 shows, the pitch differences were lower than zero in paroxytones, whereas in oxytones they were higher than zero, indicating that regardless of stress pattern the pitch was always higher in final syllables than in penultimate syllables. With regard to intensity differences, stress vowels had higher intensity than unstressed vowels in both paroxytones and oxytones (i.e., the values were higher than zero) and such tendency occurred to a similar degree in the two stress patterns. However, the information obtained from pitch and intensity differences does not explain why NSs and HS1s perceived oxytones with higher accuracy than paroxytones, because higher pitch in final syllables in

paroxytones would have misled their perception to oxytones and similar intensity differences in the two stress patterns would have led to similar accuracy rates. Rather, it is more likely that NSs and HS1s were sensitive to duration, which is argued to be the critical parameter for stress in Spanish [10]. The duration differences of the stimuli were slightly higher than zero in paroxytones, while in oxytones they were much higher than zero. This indicates that in oxytones the stressed vowels were much longer than the unstressed vowels, while in paroxytones, they were only slightly longer, although this difference was statistically significant (β = 60.747, SE = 13.051, t = 5.464, p < 0.001). This is a possible explanation of why NSs and HS1s had higher accuracy rates in oxytones. However, despite larger duration differences in this stress pattern, (half of) HS2s and L2s had lower accuracy rate, suggesting that HS2s and L2s may not pay much attention to the acoustic cues in order to understand the message, supporting [2]. Rather, it is likely that they have a bias toward selecting paroxytones, supposedly because these verbs are the first form that is learned in the classroom (present tense), hence the "easier" form to which they are more accustomed.
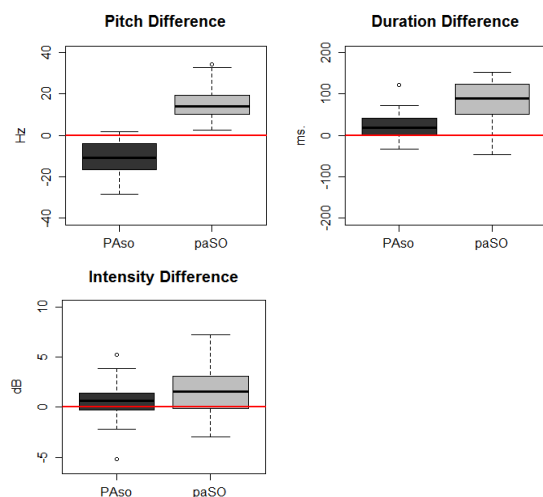

Figure 3: *Difference in pitch, duration, and intensity between stressed and unstressed vowels of stimuli*

RT was calculated as the amount of time (ms) elapsed after the onset of the target word. Only RTs of correct responses were considered in the analysis. The effect of Group, Stress Pattern, and the interaction of the two factors on participants' RT was statistically analyzed using linear mixed effects modeling with subject and item as random factors. Results showed that NSs responded significantly faster than HS2s (β = 829.71, SE = 260.76, t = 3.182) and L2s (β = 620, SE = 141.28, t = 4.388). There was no significant difference between NSs and HS1s. Regarding Stress Pattern, no main effect was found and no significant interaction was found between Stress Pattern and any of the three (non-baseline) groups, suggesting that paroxytones and oxytones were perceived with similar RTs across all four groups. That is, although the accuracy rate for oxytones was very low, HS2s and L2s did not take a particularly longer processing time when perceiving these words. This suggests that the low accuracy rate found in oxytones in (half of) HS2s and L2s is not derived from processing difficulty. Rather it is more likely to be due to bias toward paroxytones.

5. Conclusions

In the present study, the perception of Spanish lexical stress by Spanish heritage speakers and English L2 learners of Spanish was examined. Heritage speakers are similar to L2 learners in that English is the dominant language for both groups, but they are also different from each other in that while heritage speakers are exposed to Spanish at an early age, L2 learners do not learn Spanish until later. Thus, it is considered that heritage speakers have better command of Spanish phonology than L2 learners, yet not to the point that they are comparable to Spanish monolinguals, due to influence from English. The present study sub-divided the heritage speakers based on generation (1st generation US-born: HS1 vs. 1.5/2nd generation US-born: HS2). The findings indicate that while HS1s performed similarly to Spanish native speakers, HS2s were more similar to L2 learners. Given that HS2s receive less input in the heritage language than HS1s, such findings may suggest that the amount of input has an effect on heritage language phonology. However, when further analyzing the perception patterns in this group, it was found that regardless of whether a person is 1.5 or 2nd generation, those who acquired Spanish from birth showed a different pattern from those who have not been exposed to Spanish until later. This finding implies a strong effect of age of acquisition on heritage speakers' prosody, which is a linguistic aspect that is argued to be acquired from infancy [8]. However, in order to generalize such argument, it is necessary to further analyze such effect by examining a larger number of heritage speakers while taking into account factors such as language use and generation.

6. References

[1] Campbell, N. and Beckman, M. E., "Stress, prominence and spectral tilt", in A. Botinis, G. Kouroupetroglou, and G. Carayiannis [Eds], Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications, 67-70, Athens: ESCA and University of Athens, 1997.
[2] Cutler, A., "Native listening: Language experience and the recognition of spoken words", Cambridge, MA:MIT Press, 2012.
[3] Delattre, P., "A comparison of syllable length conditioning among languages", International Review of Applied Linguistics in Language Teaching, 4(14): 183–198, 1966.
[4] Flege, J. E., "The intelligibility of English vowels spoken by British and Dutch talkers", in R. D. Kent [Ed], Intelligibility in Speech Disorders: Theory, Measurement, and Management, Studies in Speech Pathology and Clinical Linguistics, 157-232, Amsterdam, The Netherlands: John Benjamins Publishing, 1992.
[5] Hualde, J. I., "The sounds of Spanish", Cambridge: Cambridge University Press, 2005.
[6] Ladd, D. R., "Intonational phonology (2nd edition)", Cambridge, UK: Cambridge University Press, 2008.
[7] Ladefoged, P., "A Course in Phonetics (4th ed.)", Fort Worth, TX: Harcourt College Publishers, 2001.
[8] Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C., "A precursor of language acquisition in young infants", Cognition, 29: 143–178, 1988.
[9] Montrul, S., "Incomplete acquisition in bilingualism: Re-examining the age factor", Amsterdam: John Benjamins, 2008.
[10] Ortega-Llebaria, M., "Phonetic cues to stress and accent in Spanish", in M. Díaz-Campos [Ed], Selected Proceedings of the 2nd conference of Laboratory approaches to Spanish Phonology, 104-118, Somerville, MA: Cascadilla Press, 2006.
[11] Saalfeld, A. K., "Stress in the beginning Spanish classroom: an instructional study", Doctoral dissertation, University of Illinois at Urbana-Champaign, 2009.
[12] Soto-Faraco, S., Sebastián-Gallés,N., and Cutler, A., "Segmental and suprasegmental mismatch in lexical access", Journal of Memory and Language, 45: 412 – 432, 2001.