DEMYSTIFYING HERITAGE ACCENT: HERITAGE ACCENT PERCEIVED BY HERITAGE SPEAKERS

Joo Kyeong Kim, Ji Young Kim, John Carter, & Jenny Eonsuh Choi

University of California, Los Angeles
anitajookyeong@g.ucla.edu, jiyoungkim@g.ucla.edu, jackcarter1868@g.ucla.edu, choijen@g.ucla.edu

ABSTRACT

Heritage speakers (HSs) generally sound more native-like than L2 learners (L2s), despite having an accent that is different from non-heritage native speakers (NHNS) or a "heritage accent." However, there is little understanding on the relative impact of segments and prosody on heritage accent. Moreover, how the heritage accent is perceived by their peer HSs is not well understood. The present study examines the relative contribution of segments and prosody in the perception of heritage accent by two groups of raters, HSs and NHNSs of Mexican Spanish. Semi-spontaneous speech samples were collected from three speaker groups (i.e., NHNS, HS, and L2), which were used as auditory stimuli for an accent rating task. HSs and NHNSs evaluated the nativeness of the speech samples presented in three different conditions (i.e., original, segment-only, and prosody-only). The results suggest that for both rater groups, prosody plays a bigger role in heritage accent perception, while L2 foreign accent is mainly perceived in the segments.

Keywords: heritage accent, accent rating, resynthesized speech, heritage speaker raters

1. INTRODUCTION

Heritage languages are minority languages acquired naturalistically in a bilingual or multilingual environment, such as diasporic languages spoken by immigrants and their children [14, 19]. The speakers of heritage languages, or heritage speakers, are generally considered to sound more native-like than second language learners, despite having an accent that diverges from that of non-heritage native speakers [1, 9, 12, 15]. However, the phonetic features that are responsible for the so-called "heritage accent" [16] are not well understood. While some studies have shown that deviation in prosody from the monolingual norm is associated with heritage accent [8, 20, 22], it is unclear whether prosody contributes more to heritage accent than segments, given that the two are inevitably intertwined. Moreover, research on heritage accent has largely been based on non-heritage native speakers' impression [1, 9, 12, 15], while there is little understanding on how heritage speakers evaluate other heritage speakers' speech.

In order to fill these gaps in the literature, we conducted an online accent rating task using resynthesized speech with both heritage and non-heritage native listeners as raters. Specifically, we seek to elucidate (1) which of the two aspects of speech, segments or prosody, contribute more to perceived heritage accent and (2) whether heritage speakers consider other heritage speakers' speech native-sounding.

2. METHODS

2.1. Participants

2.1.1. Speakers

Semi-spontaneous speech samples were recorded from 18 adult speakers of Spanish in their twenties. The speakers consisted of six heritage speakers (HSs), six second language learners (L2s), and six non-heritage native speakers (NHNSs). The L2s and the NHNSs were included as control groups that represent two extreme points of the native-non-native accent continuum [9]. All three groups were balanced by gender. The HSs were second-generation Mexican Americans, whose parents were from Mexico and moved to the US as adults. All the HSs were born and raised in Southern California. Three L2s were native speakers of American English and the other three L2s were native speakers of Korean. Lastly, the NHNSs were born and raised in Mexico.

2.1.2. Raters

Fifty-four raters participated in the accent rating task. Half of them were NHNSs (mean age: 23.0 years; 8 female, 19 male), who were all born and raised in Mexico. The other half were HSs residing in Southwestern US (24 California, 2 Texas, and 1 Nevada) and at least one of their parents was from a Spanish-speaking country. For 21 out of 27 HS raters, either one or both parents were from Mexico. For the remaining 6 raters, the parents were from other Spanish-speaking countries: Argentina (1), Colombia (1), Ecuador (1), Guatemala (1), and El Salvador (2). Since dialect may have an effect on accent ratings, in

this study, we only report data of the 21 HS raters with Mexican Spanish input (mean age: 25.6 years; 12 female, 9 male), consistent with other participants.

All the NHNS raters and 13 HS raters were recruited via Prolific (https://www.prolific.co) which is an online research participant recruitment platform. The other 8 HSs were recruited at the University of California, Los Angeles (UCLA).

2.2. Stimuli

To elicit semi-spontaneous speech, we asked the speakers to describe and narrate a wordless picture book, "Frog, Where Are You?" [13]. Two ten-second samples were extracted from each speaker, resulting in a total of 36 speech samples. All the samples were free of morphosyntactic errors. Pauses longer than 0.3 seconds were manually trimmed and speech rate was controlled across the groups using Praat [3]. Then, the samples were resynthesized to test two additional conditions (i.e., segments-only and prosody-only) other than the original condition.

For the segments-only condition, we first controlled for the rhythm so that all the samples represented the rhythm of a non-heritage native speaker. For this purpose, we had a non-heritage native speaker of Mexican Spanish, who did not participate in this study as a speaker, read aloud the transcripts of the 36 speech samples as naturally as possible. The duration of each segment in the native speaker samples was measured and then superimposed onto the test stimuli using Praat [3]. Then, we monotonized the samples using the 'Monotonize' function in Praat Vocal Toolkit [4]. This command monotonizes the pitch contour to the pitch median of the selected audio file.

For the prosody-only condition, we used the Hann Band filter in Praat [3] to remove formant values below a certain cut-off point. The frequency range was set from 0 to 320 Hz for male speakers and from 0 to 400 Hz for female speakers, and smoothing was set to 50 Hz [7, 20, 21]. These low-pass filtered samples solely retained prosodic properties. After the resynthesis, we were left with a total of 108 stimuli (= 36 samples * 3 conditions).

2.3. Procedures

The accent rating task was administered on SurveyMonkey (https://www.surveymonkey.com). The experiment was conducted in Spanish for the NHNS raters and in English for the HS raters. The experiment consisted of a background questionnaire and three blocks of accent rating tasks (i.e., one block for each stimulus type). The background questionnaire was used to ensure that participants fit the criteria for each group. For the NHNS raters, we only included individuals born and raised in Mexico. For the HSs, we only included individuals born and raised in the US or those that were born in Mexico and moved to the US before receiving formal education in Spanish (i.e., at or before the age of 6).

The participants were presented first with the original stimuli, followed by the segments-only stimuli and the prosody-only stimuli. For each item, the raters were asked to evaluate the nativeness of the speaker on a scale of 1 to 6 (= completely native). There was a training period with two speech samples prior to all three blocks. These two samples were not part of the experimental stimuli. One sample was from a NHNS and the other was from an L2.

For the segments-only stimuli, the raters were instructed that they would be hearing speech sounding like unsophisticated robots or androids. For the prosody-only stimuli, they were informed that the speech would sound as if someone was talking in an adjacent room, but with a closed door. They were also told that they would not be able to recognize the words. Instead, a transcript corresponding to the speech sample was presented on the screen and the raters were instructed to read it aloud before listening to the audio file. The entire study took around 45 minutes to complete.

3. RESULTS

A total of 5,112 responses from 48 raters were collected in the accent rating task. As for one HS rater, only the ratings of the original stimuli were included in the analysis for giving consistent ratings (1 = completely non-native) for the other two conditions with resynthesized stimuli. Table 1 demonstrates the descriptive statistics of accent ratings by stimulus type, by speaker group, and by rater group.

Stimulus Type	Mean	Median	SD
Original	4.19	5	1.77
Segments-only	3.91	4	1.63
Prosody-only	4.08	4	1.49
Speaker Group	**Mean**	**Median**	**SD**
HS	4.13	4	1.40
L2	2.95	3	1.62
NHNS	5.11	5	1.07
Rater Group	**Mean**	**Median**	**SD**
HS	4.18	5	1.56
NHNS	3.97	4	1.69

Table 1: Descriptive statistics of accent ratings (1 = completely non-native, 6 = completely native). HS: heritage speakers, L2: L2 learners, NHNS: non-heritage native speakers.

We performed a mixed effects linear regression analysis using the *lmer*() function in the *lme4* package [2] in R [17] to examine whether stimulus type, speaker group, and rater group have an effect on accent ratings. As fixed effects, we entered stimulus type (reference level = original), speaker group (reference level = HS), and rater group (reference level = HS), which were contrast-coded using simple coding. Participant was included as a random effect. The best fitting model selected through backward elimination included an intercept for participant with by-participant random slope for speaker group. Statistical significance of the fixed effects was analysed through likelihood ratio tests, which were performed using the *anova*() function in the *car* package [5].

Adding the rater group did not improve the model fit (p = 0.08). Thus, in the full model, we only included stimulus type, speaker group, and the interaction between the two fixed effects. We found a significant main effect of stimulus type ($\chi^2(2)$ = 41.828, $p < 0.001$) and a significant main effect of speaker group ($\chi^2(2)$ = 108.9, $p < 0.001$). Moreover, there was a significant interaction between the two fixed effects ($\chi^2(4)$ = 333.47, $p < 0.001$).

We obtained the summary of the results of the full model using the *summary*() function in the base package of R [17] and performed post-hoc pairwise multiple comparisons between levels using the *emmeans*() function in the *emmeans* package [11]. Results of the effect of stimulus type showed that the original stimuli were rated significantly higher than the segments-only stimuli (β = 0.287, $p < 0.001$) and the prosody-only stimuli (β = 0.118, $p < 0.05$). Post-hoc pairwise comparisons revealed that the prosody-only stimuli were also rated significantly higher than the segments-only stimuli (β = –0.168, $p < 0.001$). That is, removing the segments from the original stimuli did not affect the overall accent ratings as much as removing the prosody did.

As for the effect of speaker group, results demonstrated that HSs' speech was rated significantly higher than that of the L2s (β = 1.172, $p < 0.001$) and significantly lower than that of the NHNSs (β = −0.975, $p < 0.001$). Post-hoc pairwise comparisons confirmed that L2s' speech was rated significantly lower than that of the NHNSs (β = −2.147, $p < 0.001$). This suggests that, to both the HS and the NHNS raters, the HSs sounded more native-like than the L2s and less native-like than the NHNSs.

Now we turn to the interaction between stimulus type and speaker group. Figure 1 demonstrates the effect plot of the interaction between stimulus type and speaker group, which was created using the *allEffects*() function in the *psych* package [18].

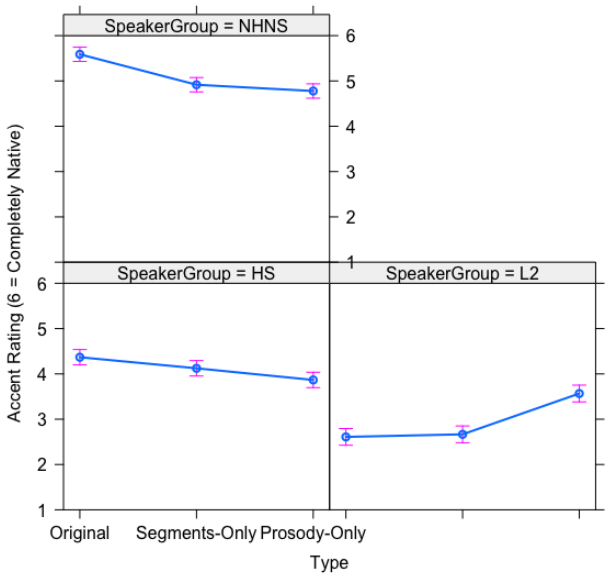

Figure 1: Effect plot of the interaction between stimulus type and speaker group.

Post-hoc pairwise comparisons showed that the above-mentioned pattern of HSs' speech being rated significantly higher than L2s' speech and significantly lower than NHNSs' speech was consistent with the original stimuli and with the segments-only stimuli (ps < 0.001) but not with the prosody-only stimuli. NHNSs' prosody-only stimuli received significantly higher ratings than those of the HSs and the L2s, while the difference between HSs' and L2s' ratings in this condition did not reach statistical significance (p = 0.089) (NHNS > HS = L2).

With regard to the role of segments and prosody in accent ratings, we found different patterns across speaker groups. HSs' segments-only stimuli were rated significantly lower than the original stimuli (p < 0.05), and the ratings of prosody-only stimuli were even lower than those of the segments-only stimuli (p < 0.05). Regarding L2s' speech, the accent ratings of the segments-only stimuli did not significantly differ from those of the original stimuli (p = 1), but both stimulus types were rated significantly lower than the prosody-only stimuli (ps < 0.001). Lastly, the accent ratings of NHNSs' speech demonstrated that the segments-only stimuli and the prosody-only stimuli were rated lower than the original stimuli (ps < 0.001). The accent ratings of the two resynthesized stimulus types did not significantly differ from each other (p = 0.645). These findings are summarized in Table 2.

Speaker Group	Accent Rating
HS	O > S > P
L2	O = S < P
NHNS	O > S = P

Table 2: Summary of the findings (higher/lower accent rating = higher/lower perceived nativeness). O = original, S = segments-only, P = prosody-only.

4. DISCUSSION

Our data demonstrated that HSs sound more native-like than L2s, but less native-like than NHNSs. This finding is consistent with previous research on perceived nativeness of heritage speakers with NHNS as raters [1, 9, 12, 15]. It is noteworthy that these patterns were also found in HS raters. HS raters' average accent rating of HSs' unaltered speech was 4.21 (SD = 1.31) and their average accent rating of NHNSs' unaltered speech was 5.11 (SD = 0.98). That is, while HSs may consider their linguistic variety (i.e., heritage Spanish) to be native, they evaluate it as sounding less native-like than non-heritage Mexican Spanish, which is likely to be associated with the standard language ideology that excludes and devalues heritage language varieties [6, 10]. Future research should examine whether HSs indeed recognize other speakers of heritage Spanish and, if so, tap into their explicit and implicit attitudes toward this linguistic variety.

In this study, we also found that the NHNS > HS > L2 pattern was observed in the original and the segments-only stimuli but not in the prosody-only stimuli where HSs' ratings did not significantly differ from L2s' ratings (NHNS > HS = L2). Moreover, among the three stimulus types, the HSs received the lowest accent ratings in the prosody-only stimuli, whereas the L2s received the highest ratings in this condition. In other words, divergence from non-heritage native accent in HSs' and L2s' speech is likely to derive from prosody and segments, respectively. These findings suggest that heritage accent is not only quantitatively different from L2 foreign accent (i.e., heritage accent sounds more native-like than L2 foreign accent), but also qualitatively different from it; heritage accent is a type of native accent that is different from non-heritage native accent mainly in the prosody, while L2 foreign accent is a non-native accent that diverges from heritage/non-heritage native accent mainly in the segments.

Future research should examine which prosodic feature in HSs' speech and which segmental feature in L2s' speech play the strongest role in perceived heritage and L2 foreign accents.

5. REFERENCES

[1] Au, T. K., Oh, J. S., Knightly, L. M., Jun, S.-A., Romo, L. F. 2008. Salvaging a childhood language. *Journal of Memory and Language* 58(4), 998–1011.

[2] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1–48.

[3] Boersma, P., Weenink, D. 2021. *Praat: Doing phonetics by computer* (6.1.38) [Computer software]. http://www.praat.org/

[4] Corretge, R. 2020. *Praat Vocal Toolkit*. http://www.praatvocaltoolkit.com

[5] Fox, J., Weisberg, S. 2019. *An {R} Companion to Applied Regression*, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

[6] Gasca Jiménez, L., Adrada-Rafael, S. 2021. Understanding heritage language learners' critical language awareness (CLA) in mixed language programs. *Languages* 6, 37.

[7] Jilka, M. 2000. *The contribution of intonation to the perception of foreign accent* [Unpublished doctoral dissertation]. Stockholm University.

[8] Kan, R. T. Y. (2020). Suprasegmental and prosodic features contributing to perceived accent in heritage Cantonese. *Proceedings of the 10th International Conference on Speech Prosody*, 101–105.

[9] Kupisch, T., Barton, D., Klaschik, E., Lein, T., Stangen, I., van de Weijer, J. 2014. Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal* 11(2), 123–150.

[10] Leeman, J. 2012. Investigating language ideologies in Spanish as a heritage language. In: Beaudrie, S., Fairclough, M. (eds), *Spanish as a Heritage Language in the United States: The State of the Field*. Georgetown University Press, 43–59.

[11] Lenth, R. V. 2022. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.7.2. https://CRAN.R-project.org/package=emmeans

[12] Lloyd-Smith, L., Einfeldt, M., Kupisch, T. 2020. Italian-German bilinguals: The effects of heritage language use on accent in early-acquired languages. *International Journal of Bilingualism* 24(2), 289–304.

[13] Mayer, M. 1969. *Frog, where are you?* Dial Press.

[14] Montrul, S., Polinsky, M. 2021. Introduction: Heritage languages, heritage speakers, heritage linguistics. In: Montrul, S., Polinsky, M. (eds), *The Cambridge Handbook of Heritage Languages and Linguistics*. Cambridge University Press, 1–10.

[15] Oh, J. S., Jun, S.-A., Knightly, L. M., Au, T. K. 2003. Holding on to childhood language memory. *Cognition* 86, B53–B64.

[16] Polinsky, M., Kagan, O. 2007. Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass* 1, 368–395.

[17] R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

[18] Revelle, W. 2022. *psych: Procedures for Personality and Psychological Research*, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 2.2.5.

[19] Rothman, J. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2), 155–163.

[20] Shin, E. 2005. The perception of foreign accents in spoken Korean by prosody: Comparison of heritage and non-heritage speakers. *The Korean Language in America* 10, 103–118.

[21] Trofimovich, P., Baker, W. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28(1), 1–30.

[22] Wrembel, M., Marecka, M., Szewczyk, J., Otwinowska, A. 2019. The predictors of foreign-accentedness in the home language of Polish-English bilingual children. *Bilingualism: Language and Cognition* 22(2), 383–400.