Imitation of the acoustic realization of Spanish stress: production and perception

Bethany MacLeod & Sabrina Di Lonardo Burr, Carleton University

Phonetic imitation is the process in which a talker's pronunciation comes to sound more similar to another person's after being exposed to their speech. Much of the research on phonetic imitation to date has focused on imitation of specific acoustic dimensions of individual sounds, such as vowel formants (e.g. [1], [4]). However, we know less about imitation of relative measures, such as the acoustic realization of lexical stress. Previous work suggests that there are three acoustic correlates of stress in Spanish: fundamental frequency (F0 - most robust cue), duration, and intensity (least robust cue). In words in isolation, stressed vowels tend to be higher pitched, longer, and louder than unstressed ([7]). Our study contributes to developing our understanding of imitation in relative measures by exploring the imitation of the acoustic realization of Spanish stress. We do this via two experiments: a shadowing task to explore how talkers produce imitation and a perceptual task to see how listeners perceive imitation. We have 3 research questions:

1. Do shadowers imitate the model talkers' acoustic realization of lexical stress in terms of F0, duration, and intensity?
2. Can listeners perceive that shadowers have imitated and does imitation of the three acoustic correlates of stress contribute to the listeners' perception that the shadowers have imitated?
3. If so, do listeners use imitation of the acoustic correlates to different extents depending on the strength of those correlates as cues to stress (i.e. F0 > duration > intensity)?

In Experiment 1, 48 female native Mexican Spanish speakers participated in the shadowing task. The participants read aloud 40 Spanish disyllabic words controlled for stress three times (baseline phase), then listened to one of four pre-recorded female native Mexican Spanish-speaking model talkers producing the same words and immediately repeated them (shadowing phase). Vowel duration, mean F0, and mean intensity were measured for both vowels in all words in the model talker, baseline, and shadowed recordings using Praat ([2]). The difference between the values of the first and second vowels for each of F0, duration, and intensity was calculated for all recordings, generating variables we call differentials for each acoustic correlate of stress ([5], [9]).

In Experiment 2, the recordings from the 48 shadower + model talker pairs comprised the stimuli in a 4IAX perceptual experiment ([11]) which 87 Spanish-speaking listeners completed. Listeners heard two pairs of words (XA XB), where X was always the model talker's token and A and B were either the baseline or shadowed token (counterbalanced) produced by a shadower. The listeners' task was to decide which of X and A or X and B were more similar to each other. The proportion of trials in which the listeners chose the pair containing the shadowed token is taken to reflect the proportion of trials in which the shadowers imitated. Bayesian mixed-effects models fit using the *brms* package ([3]) in R ([12]) were used to analyse the data, with linear models for the acoustic data and logistic models for the perceptual data.

Results for Question #1: The shadowers imitated the model talkers on all three differentials, shifting the most on the duration differential, followed by F0, and least on intensity differential.

Results for Questions #2 and #3: The listeners perceived imitation in 53.6% of trials, a proportion significantly higher than 50% ($\beta_{intercept}$ = 0.14 [0.11, 0.17]) that falls in line with previous subtle results ([10], [13]). Imitation on all three differentials was used to make the judgements; however, the extent to which they were used did not align with the differentials' strength as cues to stress. Instead, the listeners used imitation of the differentials in relation to how much the shadowers had imitated them. Furthering our understanding of phonetic imitation is

important for developing accounts of second-dialect acquisition and sound change ([8]) and may also have implications for models of teaching and learning second language pronunciation ([6]).

[1] Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 178–189.

[2] Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer Version 6.1.08, retrieved 9 December 2019 from http://www.praat.org/ (D. Weenink, Ed.).

[3] Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1).

[4] Clopper, C. G., & Dossey, E. (2020). Phonetic convergence to Southern American English: Acoustics and perception. *The Journal of the Acoustical Society of America*, 147(1), 671–683.

[5] Kim, J. Y. (2020). Discrepancy between heritage speakers' use of suprasegmental cues in the perception and production of Spanish lexical stress. *Bilingualism: Language and Cognition*, 23(2), 233–250.

[6] Lewandowski, N., & Jilka, M. (2019). Phonetic convergence, language talent, personality and attention. *Frontiers in Communication*, 4(18), 1–19.

[7] Llisterri, J., Machuca, M., de la Mota, C., Riera, M., & Ríos, A. (2003). The perception of lexical stress in Spanish. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2023–2026). Barcelona: Causal Productions.

[8] Niedzielski, N., & Giles, H. (1996). Linguistic accommodation. In H. Goebl, P. Nelde, Z. Starý, & W. Wölck (Eds.), Kontaktlinguistik – Ein internationales Handbuch zeitgen¨ossischer Forschung (pp. 332–342). Berlin/New York: Mouton de Gruyter.

[9] Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in central Catalan and Castilian Spanish. *Language and Speech*, 54(1), 73–97.

[10] Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, and Psychophysics*, 79(2), 637–659.

[11] Pisoni, D. B., & House Lazarus, J. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 55(2), 328–333.

[12] R Development Core Team (2012). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org/.

[13] Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66(3), 422–429.