Phonological Categories in perception and production: the link and individual variability

Kuniko Nielsen

Oakland University (USA)
nielsen@oakland.edu

Fine phonetic details are used by listeners in processing speech and affect listeners' subsequent speech productions [1, 2] suggesting a close link between speech perception and production. However, the link between phonological categories in speech perception and production is still largely unknown. While previous studies found correlations between speech perception and production [e.g., 3-6], many studies also failed to find such correlations [e.g., 7-11], revealing the complex nature of the perception-production link.

The current study explores the link between perception and production by examining the correlation between individual speakers' categorical boundaries in perception and production through VOT. To capture a holistic picture of phonological categories, different variables that reflect production categories (e.g., mean, minimum, and maximum VOT for /p/ and /b/, and the midpoint of the gap between the two categories) are examined in both isolated and connected speech. Individual variability of perceptual categorical boundaries is also examined: while VOT production is shown to be highly variable across speakers yet structured within speakers [e.g., 12], individual variability of VOT in perception, such as the categorical boundary for voicing contrast, is largely unknown. Previous studies showed that perceptual cues vary across listeners [13, 14], so we would expect to find individual variability in categorical boundaries as well.

Thirty native speakers of American English participated in an online experiment which examines the correlation between perceptual categorical boundaries for the /p/-/b/ contrast and production variables of isolated and connected speech. A native speaker recorded the tokens bear and pear, and a 9-step *bear-pear* continuum was created (on a VOT scale from 12ms to 52ms) as identification task stimuli. Each token was presented twice in random order. Production stimuli for the isolated speech and connected speech were 36 monosyllabic words with initial stops and a short passage from the book "Peppa Pig: Family Trip" (116 words, 20 bilabial stops), respectively. During the production task, participants were randomly presented with the test words on the computer screen (each token was presented twice), and were asked to read them aloud at a comfortable pace. The short passage was presented subsequently, and the participants were asked to read it at a comfortable pace.

The results showed that perceptual boundaries in VOT vary widely across speakers (19.5ms - 47ms), and so did the slope. As expected, the mean production values in VOT also showed a wide variability (/p/: 45ms - 124ms, /b/:-80ms - 29ms). Mixed-effects modeling and additional linear regression analysis showed no significant correlation between /b/-/p/ perceptual categorical boundary and any of the production variables in isolated speech ($p>0.1$) (Fig. 1), while a significant correlation was found between the categorical boundary and mean /p/ VOT in connected speech (adjusted $R^2 = 0.41$, $p<0.001$) (Fig. 2). There was no significant correlation between isolated and connected speech nor between categorical boundary and mean /b/ VOT.

Our results showed a significant correlation between (perceptual) categorical boundary and mean /p/ VOT in connected speech, which suggests that the representation of phonemic categories is likely to include fine phonetic details. Our results also showed that the type of production task matters in examining the perception-production link: isolated speech was more hyperarticulated and variable in our data, potentially obscuring the link. One limitation of the current data is that the perception data comes from only one continuum (i.e., *bear-pear*). To address the lack of fine-grained perception data, additional data is currently being collected for a second experiment which includes more stimuli for the categorical perception task. The data will be analyzed using a Bayesian mixed effects regression model.

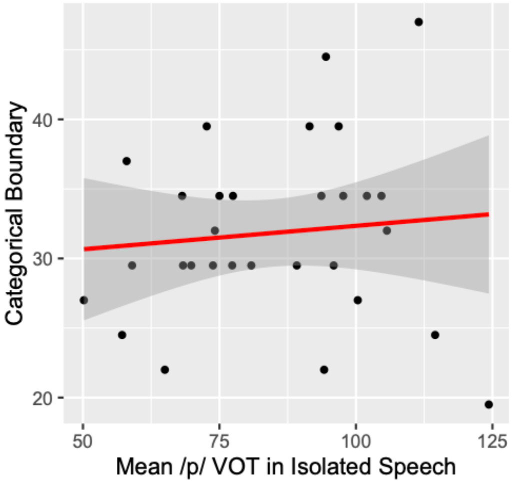

Fig.1 Mean /p/ VOT values and /b/-/p/ categorical boundary for each participant in isolated speech

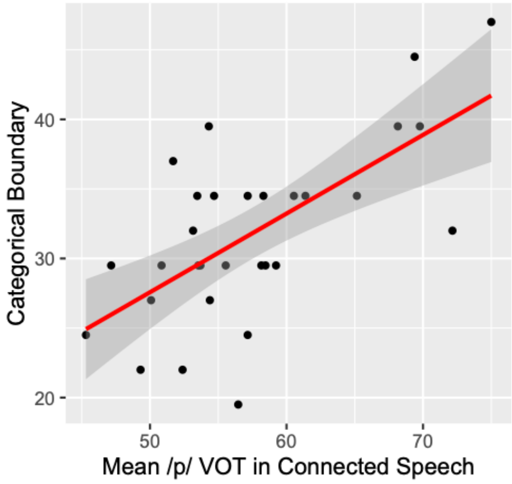

Fig.2 Mean /p/ VOT values and /b/-/p/ categorical boundary for each participant in connected speech

[1] McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. Journal of memory and language, 60(1), 65-91.

[2] Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. Psychological review, 105(2), 251.

[3] Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. The Journal of the Acoustical Society of America, 116(4), 2338-2344.

[4] Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. Language, 94(4), 931-968.

[5] Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. Language, Cognition and Neuroscience, 34(6), 769-786.

[6] Pinget, A. F., Kager, R., & Van de Velde, H. (2020). Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. Language and Speech, 63(3), 660-685.

[7] Bailey, P. J., & Haggard, M. P. (1973). Perception and production: Some correlations on voicing of an initial stop. Language and speech, 16(3), 189-195.

[8] Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. The Journal of the Acoustical Society of America, 113(5), 2850-2860.

[9] Kraljic, T., Samuel, A.G., and Brennan, S.E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. Psychological Science. 19:332–338.

[10] Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. The Journal of the Acoustical Society of America, 132(2), EL95-EL101.

[11] Cheng, H. S., Niziolek, C. A., Buchwald, A., & McAllister, T. (2021). Examining the Relationship Between Speech Perception, Production Distinctness, and Production Variability. Frontiers in Human Neuroscience, 15.

[12] Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. Journal of Phonetics, 61, 30-47.

[13] Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. Journal of phonetics, 28(4), 377-396.

[14] Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. Journal of Phonetics, 59, 40-57.