

The effect of linguistic information on f0 imitation

Rebecca Scarborough¹, Alyssa Strickler¹, Kuniko Nielsen²

¹University of Colorado, Boulder, ²Oakland University

Studies in phonetic imitation have shown that speakers imitate some phonetic patterns to which they are exposed (e.g., Pardo et al., 2013; Nielsen, 2011). However, it is unclear what aspects of the speech signal speakers are responding to when they change their speech behavior: a specific acoustic value or a linguistically-interpreted target. To address this issue, we conducted two pitch imitation experiments, one in which participants were exposed to a linguistically-unmarked overall pitch difference and one in which participants were exposed to a linguistically-salient manipulated pitch accent realization. If imitation targets the specific acoustic value of a model talker's pitch, we expect participants to converge acoustically toward that f0 in all cases; if imitation targets linguistic patterns, we expect participants to imitate the linguistically meaningful pattern, e.g., a contrastive focus pitch accent, even if it results in acoustic divergence.

Both experiments included 4 blocks: 1) baseline, in which American English speaking participants produced sentences based on information presented on-screen; 2) exposure, in which participants heard stimulus sentences presented auditorily; 3) shadowing, in which the participants repeated sentences presented auditorily; and 4) post-test, which was like the baseline task. The model talker whose speech was presented in the exposure and shadowing blocks was a male with a naturally low f0 (mean=101Hz in carrier phrases). In *Experiment 1*, 7 participants (2M 5F, data collection is on-going) produced 49 utterances of the form "The word is [X]," where the target word could be a color, animal, or shape. In *Experiment 2*, 10 participants (4M 6F, also on-going) were shown a 4x4 grid composed of different shapes in different colors and were asked to describe the location of an animal on the grid. The animal was moved from trial to trial to elicit 52 contrastive sentences, e.g., "Now the mouse is on the *red* square." In the shadowing task, each of the model talker's L+H* contrastive peaks was acoustically manipulated to be 1.5 times the peak in his normal production (unedited mean=209Hz). In Expt. 1, participants' mean f0 was measured from the carrier phrase ("The word is") of each utterance. In Expt. 2, f0 was hand measured at the f0 peak in each target word.

Results showed that participants in Expt. 1 lowered their carrier-phrase f0 in shadowing (175Hz) and post-test (176Hz), relative to baseline (190Hz), converging toward the lower f0 of the model talker. In Expt. 2, female participants *increased* their peak f0 (realized in a contrastive pitch accent) in both shadowing (221Hz) and post-test (231Hz), relative to baseline (217Hz). Separate mixed effects linear regressions on f0 for each experiment included Block and Gender and their interaction as fixed factors, and random intercepts by-participant. The Expt. 1 model confirmed that both shadowing [$t=-4.53$, $p<.0001$] and post-test [$t=-3.91$, $p<.0001$] f0 were significantly lower than baseline; as expected, men were lower than women [$t=10.7$, $p>.0001$]. The Expt. 2 model supports that post-test peaks (but not shadowing peaks) were higher than baseline overall [$t=1.83$, $p=.068$]. Men's peaks were lower than women's [$t=-5.35$, $p=.0001$].

Across the experiments participants did imitate f0, but their behavior indicated two different patterns of imitation. In non-focused carrier phrases, they converged toward the low f0 of the model talker, replicating previous studies on f0 imitation (e.g., Babel & Bulatov, 2012); however, in contrastively focused target items, their productions shifted toward the linguistic pattern displayed by the model speaker, namely a high L+H* peak (and for the female

participants, even exceeded the specific f0 target of the model talker). We argue that speakers take into account linguistic relevance when imitating. Absent linguistically meaningful structures (as in the overall f0 of a speaker), speakers may imitate acoustic targets directly. (Alternately, they may be imitating a socially-relevant pattern expressed by low f0.) When encountering and interpreting a linguistically meaningful structure (as in the pitch rise of a contrastive focus pitch accent), however, they imitate the linguistically-relevant pattern produced by the speaker.

References

- Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and speech*, 55(2), 231-248.
- Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132-142.
- Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, 69(3), 183-195.