Acoustic and linguistic influences on f0 imitation

Kuniko Nielsen[1], Rebecca Scarborough[2]

[1]*Oakland University (USA),* [2]*University of Colorado Boulder (USA)*
nielsen@oakland.edu, rebecca.scarborough@colorado.edu

Studies in phonetic imitation have shown that speakers imitate some phonetic patterns to which they are exposed [1, 2]. However, it is still 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 online pitch imitation experiments: one in which participants were exposed to a linguistically-unmarked overall pitch difference (in this case, low f0) and one in which participants were exposed to a linguistically-salient manipulated pitch accent realization (in this case, extra high H in L+H*). If imitation targets the specific acoustic value of a model talker's naturally low pitch, we expect participants to converge acoustically toward the talker's low f0 in both experiments. On the other hand, if imitation targets linguistic patterns, participants should imitate the linguistically meaningful pattern, i.e., the high contrastive focus pitch accent, even if it results in acoustic divergence from the model talker's generally low f0.

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*, eighteen participants (9M, 9F, data collection is on-going) produced 80 utterances of the form "The word is [X]," where the target word could be a color, animal, or shape pictured on the screen. In *Experiment 2,* fourteen participants (7M, 7F, also on-going) were shown a 3x3 grid composed of different shapes in different colors and were asked to describe the location of an animal on the grid. The animal moved from trial to trial to elicit 60 contrastive sentences, e.g., "Now the mouse is on the *red* square." In the exposure and shadowing tasks, listeners heard versions of the model talker's speech acoustically manipulated so that L+H* contrastive peaks were 1.2 times their natural peak height (unedited mean=179Hz; edited mean=215Hz). The degree of manipulation was chosen to ensure that the peak was saliently enhanced but would still be lower than female participants' f0 peaks. In Expt. 1, participants' mean f0 was measured as the average f0 across each utterance. In Expt. 2, f0 was measured at the hand-labeled f0 peak in each target word; a relative peak height was calculated as the height of the f0 peak divided by the utterance average f0. Participants were recruited and paid through the online recruitment platform Prolific, and the online experiment was set up and administered using Gorilla [3].

Results showed that female participants in Expt. 1 lowered their utterance average f0 in shadowing (168Hz) and post-test (170Hz), relative to baseline (174Hz), converging toward the lower f0 of the model talker; male participants showed no change (baseline:101Hz, shadow:101Hz, post:101Hz). In Expt. 2, on the other hand, male participants (but not female participants) increased their peak f0 (realized in a contrastive pitch accent) in both shadowing (132Hz) and post-test (132Hz), relative to baseline (126Hz).

Separate mixed effects linear regressions on f0 for each experiment (utterance mean f0 for Expt. 1 and relative f0 peak for Expt. 2) included block and gender and their interaction as fixed factors, and random intercepts by-participant. The Expt. 1 model confirmed that both shadowing [$t=-3.47$, $p<.001$] and post-test [$t=-2.92$, $p<.01$] f0 were significantly lower than baseline. As expected, men were lower in pitch than women [$t=-9.95$, $p<.0001$]. There were significant gender by block interactions as well: female participants lowered their f0 more than male participants in both shadowing [$t=2.470$, $p>.05$] and post-test [$t=2.053$, $p<.05$].

The Expt. 2 model of relative peak f0 showed no main effect of block or gender, but there were significant interactions between block and gender in both tasks [shadowing: $t=2.683$,

p<.001; post-test: t=3.134, p<.001], showing that only male participants increased the relative peak f0 in these tasks. A second model looking at raw peak f0 (not relative to utterance f0) showed that the shadowing peaks were lower than the peaks at baseline [t= -2.27, p<.05] (reflecting a lowering of f0 overall, as in Expt. 1), but post-test peaks were not [t=-0.46, p>0.1]. There was also a significant interaction between shadowing and gender [t=-2.526, p<.05], due to the lowering of shadowing peaks for women, while men's peaks were raised. This pattern seems to reflect the overall f0 lowering of females in shadowing also seen in Expt. 1.

Across the experiments, then, participants did imitate f0, but their behavior indicates two different patterns of imitation. In contrastively unmarked utterances (Expt. 1), female participants converged toward the low f0 of the model talker in both shadowing and post-test blocks, replicating previous studies on f0 imitation in which pitch (with no phonologically relevant distinctions) was imitated [4, 5]. Male participants in our study, whose f0 was already similar to the model talker, showed no lowering. When producing contrastively focused target items (Expt. 2), female participants again shifted downward, producing lower pitch accent peak f0 (but with no change in relative peak) during shadowing. In other words, they shifted toward the model talker's actual f0, but opposite his pattern of raised pitch accents. But in Expt. 2, this lowering was not carried over to the post-test block as in Expt. 1. Male participants did exhibit imitation (i.e., increased relative peaks), but we cannot determine from the current data whether they were imitating a specific acoustic value or a linguistic target (i.e., a high contrastive pitch accent), since both the actual f0 target and the pattern of raised pitch accent were higher than the male participants' baseline f0 peaks.

We argue nonetheless that speakers take into account linguistic factors when imitating. Absent linguistically meaningful structures, speakers may imitate acoustic targets directly, as seen in overall f0 lowering for female talkers in Expt. 1. When encountering and interpreting a linguistically meaningful structure (as in the pitch rise of a contrastive focus pitch accent in Expt. 2), the patterns of imitation are altered, especially if the acoustic and linguistic target are incongruent. Our results suggest that speakers' speech behavior reflects sensitivity to both a specific acoustic value and a linguistically-interpreted target. Additional data from female participants is being collected in order to further investigate the extent to which the presence (or salience) of incongruent linguistic and acoustic targets affects imitation.

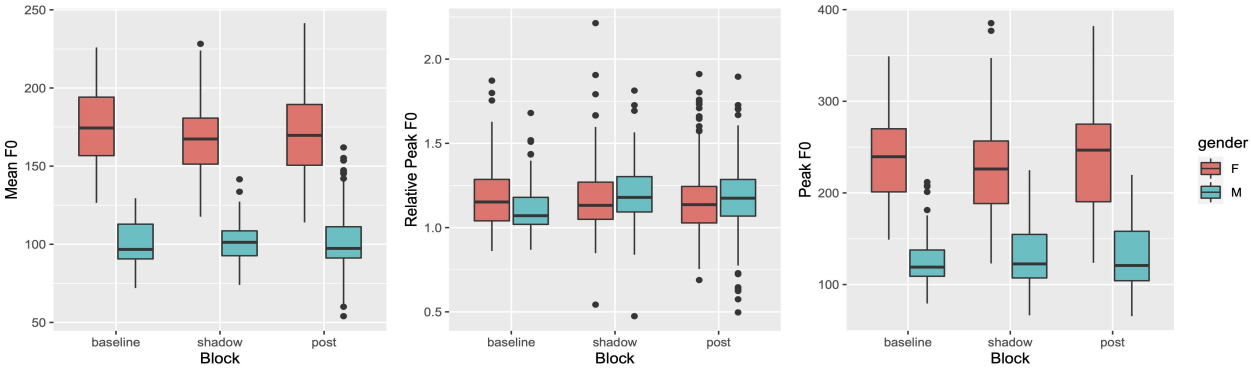

Fig 1. *Mean f0 (left) from Expt. 1 and relative peak f0 (middle) and peak f0 (right) from Expt. 2, by block. (Note that the model talker had a mean f0 of 101Hz, a mean relative peak of 1.99, and a mean peak f0 of 215Hz.)*

References

[1] Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013) Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *JML*, 69(3), 183-195.

[2] Nielsen, K. (2011) Specificity and abstractness of VOT imitation. *J Phon*, 39(2), 132-142.

[3] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020) Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, 52(1), 388-407.

[4] Dilley, L. C. (2010) Pitch Range Variation in English Tonal Contrasts: Continuous or Categorical? *Phonetica*, 67, 63–81.

[5] Babel, M., & Bulatov, D. (2012) The role of fundamental frequency in phonetic accommodation. *Lang & Sp*, 55, 231-248.