

Supplemental Material S4. Identification performance on untrained-novel talker.

Referring back to our method section, in the identification tests, we included an untrained female talker (F2) in addition to the trained talker (F1). It is our intention to examine whether the current training paradigm is effective in generalizing trained level tone categorization to stimuli produced by an untrained novel talker, and whether the effect of talker interacts with the consolidation process discussed in the main text.

First, participants were able to generalize trained categorization such that their performance was significantly different from the chance level of response accuracy (i.e., 0.33) on the trained stimuli (mean accuracy: 0.71), $t(157) = 35.98$, $p < .001$, and untrained stimuli (mean accuracy: 0.65), $t(157) = 30.63$, $p < .001$, in the identification tests.

Second, in the main text, the aptitude model on identification revealed a significant interaction between pitch threshold and test at HV but not LV group. To test whether this interaction can be further explained by the effect of talker (2 levels: F1 Trained vs. F2 Untrained; deviation coding: -0.5 , 0.05), we ran a mixed effects model with talker, tested nested within group, and pitch threshold as fixed effects, and by-participant and by-tone intercepts as random effects. Results showed a main effect of talker ($\beta = -0.44$, $SE = 0.03$, $z = -13.55$, $p < .001$), indicating better tone categorization in stimuli produced by the trained talker (F1) than the untrained talker (F2). There was significant interaction between talker and pitch threshold ($\beta = 0.10$, $SE = 0.03$, $z = 3.02$, $p = .003$). Post-hoc analyses revealed that although different in magnitude, for both trained and untrained talker, pitch threshold predicted overall identification accuracy (Trained: $\beta = -0.33$, $SE = 0.07$, $z = -4.97$, $p < .001$; Untrained: $\beta = -0.20$, $SE = 0.05$, $z = -3.99$, $p < .001$).

The results showed that training variability and pitch aptitude influenced tone consolidation (and training progress), however, neither of them contributed to the talker generalization of tone categorization (i.e., with identification accuracy of the stimuli produced by the untrained talker being significantly above chance). Training variability has been shown to promote generalization to new talkers in phonetic training studies on segmental contrasts (Bradlow et al., 1999; Lively et al., 1994; Logan et al., 1991). For tone learning, Perrachione et al. (2011) showed that, regardless of pitch aptitude, learners had better generalization abilities (measured by a ratio of posttest performance with untrained talkers and training performance with trained talkers; see note 1 of Dong et al., 2019) following the HV training than the LV training. However, Dong et al. (2019) found no evidence for better generalization following the HV training than the LV training when comparing the accuracy of stimuli produced by the trained and untrained talkers in the ID posttests. Thus, the finding is not conclusive regarding the effect of training variability, which is further modulated by other factors such as training length and talker presentation, on talker generalization in tone learning (for a systematic review, see Zhang et al., 2021). On the other hand, the issue of talker generalization became complicated for the current design including an overnight sleep, which potentially has promoted both the consolidation of trained materials as well as the generalization to novel talkers (Qin & Zhang, 2019). Thus, the overnight consolidation might have obscured the effect of talker generalization, resulting in a lack of interaction between talker and other factors (e.g., training variability). Since learners need to efficiently map variable acoustic input to tonal categories produced by different talkers for successful learning, this research calls

for future studies to test the effect of training variability (and pitch aptitude) on tone consolidation and generalization.

References

- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics*, 61(5), 977–985. <https://doi.org/10.3758/BF03206911>
- Lively, S. E., Pisoni, D. B., Yamada, R. A., Yoh'ichi, T., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. iii. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, 96(4), 2076–2087. <https://doi.org/10.1121/1.410149>
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, 89(2), 874–886. <https://doi.org/10.1121/1.1894649>
- Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, 130(1), 461–472. <https://doi.org/10.1121/1.3593366>
- Qin, Z., & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PLoS ONE*, 14(12), e0221498. <https://doi.org/10.1371/journal.pone.0221498>
- Zhang, X., Cheng, B., & Zhang, Y. (2021). The role of talker variability in nonnative phonetic learning: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, 64(12), 4802–4825. https://doi.org/10.1044/2021_jslhr-21-00181