Intonation meets emotion: Evidence from production for robust phonological contrasts in F0

Daniel R. Turner[1], Jennifer S. Cole[1]
[1]*Northwestern University*
dturner@u.northwestern.edu, jennifer.cole1@northwestern.edu

Intonation is notoriously variable in the speech signal, making empirical validation of phonological theories like the Autosegmental-Metrical (AM) model (Ladd, 2008) for Mainstream American English (MAE) a deep challenge. We examine emotion as a source of variation in intonation production with a study that crosses AM's phonologically contrastive phrase-final (nuclear) tunes with psychometrically contrastive emotions (Fontaine et al. 2007). **Methods.** An imitation experiment was conducted following Cole et al. (2023/C23), extended by (a) presenting intonational tunes in rich pragmatic context, (b) eliciting an emotional portrayal, and (c) recruiting trained voice actors (N=12) along with university students (N=19) as participants. We tested a subset of 8 nuclear tunes comprised of sequences of Low/High pitch accent, phrase accent and boundary tone: LLL, LLH… HHH. Crossing 8 tunes with 5 emotions (love, pride, anger, shame, neutral) gives 40 tune-emotion combinations, which participants produced over 3-syllable names (e.g. *Marilyn*). The final dataset included 4,764 tune-emotion productions. C23 found production and perception evidence for at most 5 of the 8 predicted tune categories, when tunes were presented without context. We hypothesized that with a specified pragmatic context, our participants would successfully imitate all 8 tunes. In an exploratory analysis, we also looked for interactions of emotion and tune that might enhance or reduce the tune distinctions produced in emotionally neutral or unspecified pragmatic contexts. Pitch contours from imitated productions across emotions were analyzed with k-means clustering to test the correspondence between emergent clusters and phonologically specified tune categories. Effects of emotion and tune-emotion interaction on contour variation were modeled using GAMM regression, with tune, emotion and their interaction as predictors, and by-speaker random effects. **Results.** Clustering results showed 6 robust distinctions among the emotional tune productions, one more than in C23, which is still short of the 8-way distinction predicted by the AM model. GAMM results showed main effects of emotion and distinct contour shapes for each tune in at least one emotion condition, yet no single emotion condition exhibited the full 8-way contrast. Tune shape is preserved under emotion, with emotion effects seen only in interactions with specific tunes, which suggests an interaction between pragmatic and emotional contexts in tune production.

References: • Cole et al. (2023). *Laboratory Phonology:* 1-51. • Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). *Psychological Science*, 18 (12), 1050-1057. • Ladd, D. R. (2008). Intonational phonology (2nd ed). Cambridge Univ. Press