Supplemental Material S5. Supplemental methods.

Manual Acoustic Remeasurement

The automated formant extraction values were screened for plausibility and the spectrograms associated with these tokens were visually inspected if more than five flagged tokens appeared in a given session. Remeasurement, if necessary, was completed using a wideband spectrogram and the fast Fourier transform spectral slice. Firstly, all measured tokens were screened for audio distortion due to signal clipping, and those with over 15% clipped samples within the target /ɹ/ interval were manually verified to ensure that the automated formant estimation algorithm was not impacted by audio distortion. A total of 118 tokens were flagged during this screening process (3.2% of all extracted tokens) for visual inspection. Manual remeasurement was then competed for files for which audio distortion precluded accurate automated formant extraction (0.89% of all extracted tokens).

Additionally, all audio files were screened for inaccurate automated formant estimation of F2 and F3 using two criteria: by screening for implausible F2 values and screening for F3-F2 distances that did not align with the perceptual rating of the token. Implausible F2 values were defined as automatically extracted rhotic F2 values that were lower than two standard deviations below the subject’s own average baseline F2 in /ɔ/ (which, as a back vowel, has F2 values near the low extreme). Automatically extracted F2 values that were greater than two standard deviations above the age-and-gender matched average F2 from the Lee et al. (1999) norms were also flagged for visual inspection. Manual formant remeasurement was completed for tokens where poor formant tracking precluded accurate automated formant extraction.

Screening for F3 values needing remeasurement was completed relative to the age-and-gender matched F3-F2 normative distance from Lee et al. (1999) and the perceptual rating of the production. For tokens that were rated as “correct,” those that had an automatically extracted F3-F2 difference more than 2 standard deviations above the normative F3-F2 mean were flagged for visual inspection (4.5% of all extracted tokens). For tokens that were rated as “incorrect,” those that had an automatically extracted F3-F2 difference less than 2 standard deviations above the normative F3-F2 mean were flagged for visual inspection. Manual formant remeasurement was completed for tokens where poor formant tracking precluded accurate automated formant extraction. Overall, 617 (16.7%) extracted tokens were flagged for visual inspection and 503 (13.6%) were manually remeasured. A total of 12 tokens were identified as unusable, resulting in F2 and F3 measurements for 3,685 tokens across all subjects.

Reference

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children’s speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. https://doi.org/10.1121/1.426686