Supplemental Material S1. Supplementary information on methodology.

Measurable vowels

The number of measurable vowel replicates varied by speaker, averaging 60 (range: 44–69). On average, each speaker produced 11 measurable replicates of /i/ (range: 7–12), 25 measurable replicates of /æ/ (range: 23–26), 14 replicates of /ɑ/ (range: 6–18), and 10 replicates of /u/ (range: 5–12). In some instances, there were small reductions in the number of measurable kinematic replicates, relative to acoustic replicates, due to instrumental tracking errors. Speaker and vowel specific differences in the number of contributing vowel replicates reflected issues related to speech disability (e.g., poor vocal support) as well as idiosyncrasies in the phonetic details of phrasing and coarticulation.

Reliability on manual measurements

The only manual aspects that pertain to measurements was the segmentation of vowels and formant track correction. All other measurements in this study were algorithmic. The vowels for segmentation were all coded and discussed by the two authors so that the same segmentation approach was adopted between the two research labs that conducted the measurement tasks. For reference, for 30 randomly sampled vowel segments measured at James Madison University, the Pearson correlations for the original segmentation time points and inter-experimenter resegmentation are $r = 1$ for the vowel segment onset and $r = 0.999$ for the offset. The averaged differences between onset and offset time points for the original and repeated segmentation are -0.56ms and 4.98ms (absolute differences: 15.23ms and 24.53ms, respectively). For the 30 randomly selected vowel segments, there are a total of 628 acoustic measurement time points. The repeated-measure F1 and original F1 values yield $r = 0.88$, and for F2, $r = 0.95$. The averaged absolute differences between the original and repeated F1 and F2 values are 37.29 and 45.71, respectively. These are consistent with reports in the literature.