

Laryngeal elevation in native and non-native lexical tone production

Fengyue Lisa Zhao¹, Sam Tilsen¹, and Mark Tiede²

¹Cornell University (USA), ²Yale University (USA)

Prior acoustic studies have found differences between native and non-native speakers in precision of lexical tone production, but corresponding differences in articulatory mechanisms remain underexplored. This study presents a novel comparison of laryngeal elevation in native and non-native speakers during tone production using coronal-orientation laryngeal ultrasound. Both laryngeal elevation and cricothyroid contraction can contribute to vocal fold tension and thus f_0 modulation [1, 2], yet their relative roles have been difficult to assess, mostly due to methodological challenges. We hypothesized that native and non-native speakers differ in their reliance on extrinsic (laryngeal elevation) versus intrinsic (cricothyroid) mechanisms of f_0 control to produce tone contrasts, with native speakers relying more on intrinsic control and non-native speakers more on extrinsic control. This predicts weaker correlations between laryngeal elevation and f_0 in native speakers than in non-native speakers. Our results support this prediction and suggest a broader acquisition pattern: learning to produce tones may involve reweighting articulatory strategies, shifting from general extrinsic mechanisms toward intrinsic control.

Method. Four participants were tested: two speakers of languages without lexical tone and two native Mandarin speakers. The non-native speakers were phonetically trained and produced Mandarin tones on the basis of standard numeral representations. The ultrasound probe was held by the participant at the right thyroid lamina in a coronal orientation to capture vertical laryngeal movement, following the protocol in [1]. Synchronized audio and laryngeal ultrasound (with a standard clinical convex probe, 90-100 fps) were recorded simultaneously. All participants completed blocks of high-to-low (HL) and low-to-high (LH) pitch sweeps along with repetitions of four Mandarin tones in the syllable *ma*. Two approaches were used to infer laryngeal elevation from ultrasound images: (i) image pixel intensity centroid, and (ii) weighted sum of pixel intensity sampled on a semipolar image grid, with weights from averages of lasso regressions predicting f_0 . Figure 1 shows representative frames from an HL sweep, along with f_0 and intensity centroid, confirming that changes in intensity centroid values reflect f_0 movement in sweeps. Figure 2 shows example f_0 trajectories and laryngeal-elevation traces from the lasso-based method for one native and one non-native male speaker. The lasso-based method was found to effectively identify image regions most predictive of f_0 , reducing noise from irrelevant portions of the ultrasound field.

Results. Preliminary analyses with both measures replicate the general pattern reported in [1]: larynx elevation is positively correlated with f_0 across tasks. Figure 3a presents correlation coefficients between the centroid-based laryngeal elevation measure and f_0 for each participant. The two female speakers showed higher correlations than the two male speakers, a pattern opposite to that reported in [1]. Crucially, Figure 3b compares correlations between sweeps and tones. Native speakers show stronger correlations between laryngeal elevation and f_0 in sweeps than in tones, whereas non-native speakers show either weaker or the opposite effect. Under the assumption that the elevation- f_0 correlation for sweeps provides a reference value for utilization of laryngeal elevation, the greater reduction in correlation from sweeps to tones observed in native vs. non-native speakers indicates that in the course of learning native speakers of tone languages may come to rely less on extrinsic laryngeal elevation and more on intrinsic muscle activity (e.g., cricothyroid manipulation) to produce tones than their non-native counterparts. Overall, these findings support a role for laryngeal elevation in tone production, while also motivating a broader hypothesis: speech learning may involve a shift in synergistic organization across articulatory mechanisms. This hypothesis can be tested in future work with larger samples.

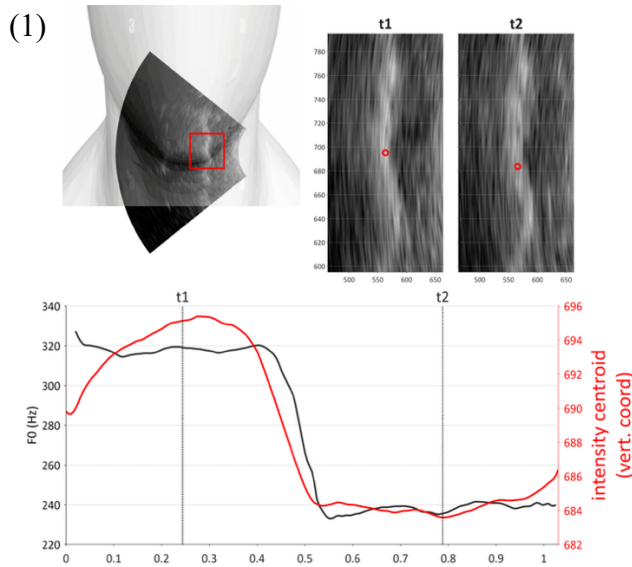
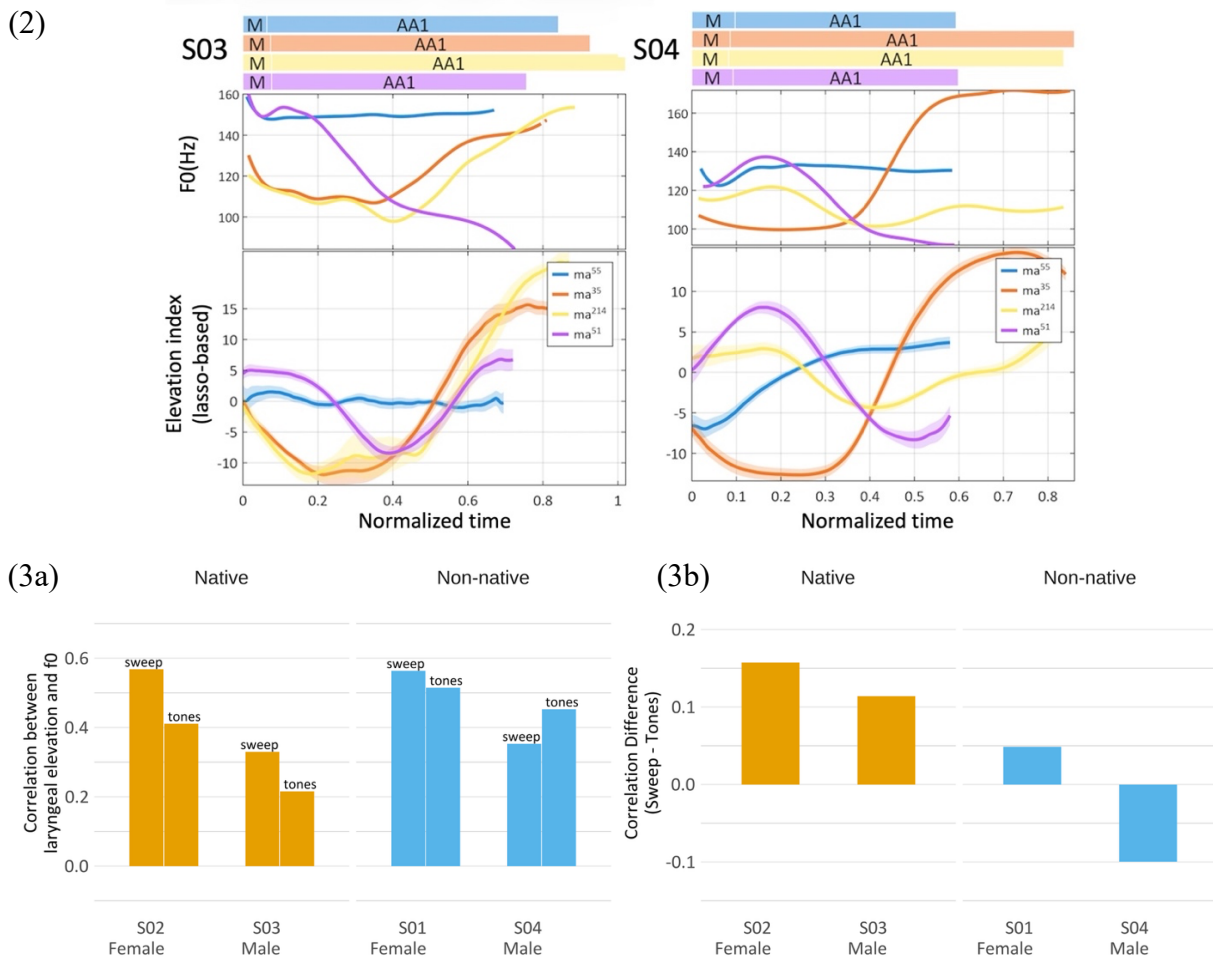


Fig 1. Two laryngeal ultrasound frames an HL pitch sweep. Changes in intensity centroid values (red) reflect f0 movement (black).

Fig 2. Example f0 trajectories and elevation index values from lasso-based method for a native (S03) and a non-native (S04) speaker. Horizontal bars at the top show durational differences between tones.

Fig 3. (a) Correlation between laryngeal elevation (centroid-based) and f0 for each participant across sweeps and tones.

(b) Difference in correlations between sweeps and tones; positive values indicate a greater correlation in weeps than in tones.



[1] Moisik, S. R., Lin, H., & Esling, J. H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association*. [2] Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of Vertical Larynx Movement and Cervical Lordosis in F0 Control. *Language and Speech*.