Discrete constriction locations describe a comprehensive range of vocal tract shapes in the Maeda model

JL Gaines[1], KS Kim[2], B Parrell[3], V Ramanarayanan[2,4], SS Nagarajan[2], JF Houde[2]

[1]UC Berkeley-UCSF Graduate Program in Bioengineering
[2]Department of Otolaryngology - Head and Neck Surgery, University of California, San Francisco
[3] Department of Communication Sciences and Disorders, University of Wisconsin–Madison
[4]Educational Testing Service R&D, San Francisco

Introduction: Distinct vowel sounds are produced when the shape of the vocal tract changes, shifting the formant frequencies of the sound. Understanding which features of vocal tract shape contribute to changes in formant frequency is relevant to characterizing the speech motor control task. In this exploration, we use the Maeda model of the vocal tract [1] to generate a comprehensive range of vocal tract shapes that produce ecologically valid formant frequencies in an unsupervised manner. We note that the vocal tract shapes in this set produce a complete range of valid frequencies across the first two formants (F1 and F2); however, the vocal tract shapes in this set have a limited subset of constriction locations in the tongue body region. These results extend previous findings based on X-ray data [2] to show that this discrete parameterization of constriction location is a fundamental characteristic of the human vocal tract and is not limited to only those configurations used for linguistic vowel contrasts [3].

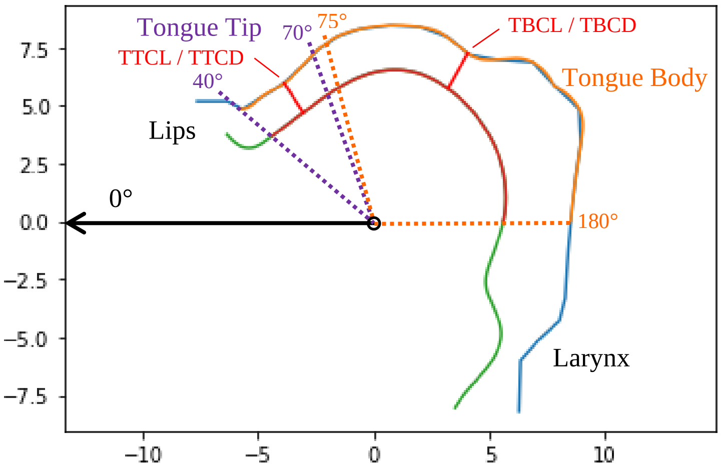

Figure 1. Calculation of task parameters tongue tip constriction location (TTCL), tongue tip constriction degree (TTCD), tongue body constriction location (TBCL), and tongue body constriction degree (TBCD).

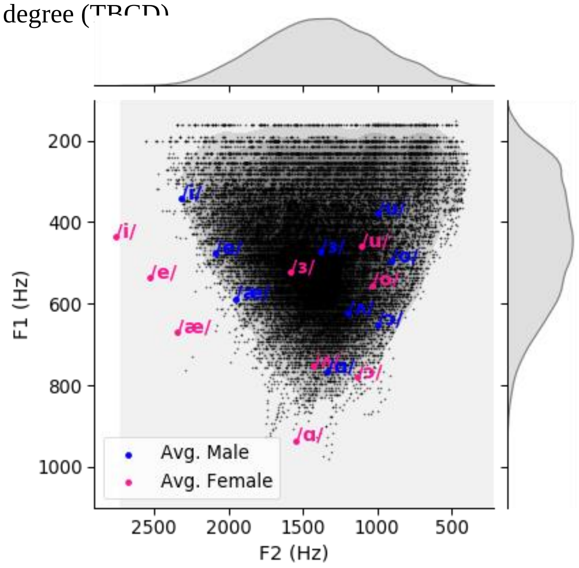

Figure 2. The formants produced by the set of valid vocal tract shapes covers the F1-F2 space with bell-shaped density centered at the median value of each formant. For reference, average formant frequencies for male and female speakers are included [4].

Methods: The Maeda model [1] is a data-driven model that maps features of vocal tract shape to the formant frequencies of the resulting speech sounds. The seven Maeda features, which are designed to be maximally independent from one another, correspond strongly with jaw position, tongue dorsal position, the arched or flattened shape of the tongue, tongue apex position, larynx length, lip height, and lip protrusion. Variation in each parameter is normalized to the number of standard deviations above or below the mean.

To generate a dataset spanning all possible vocal tract shapes, each Maeda feature was assigned each of the following values: [-3, -1.5, 0 1.5, 3] standard deviations from the mean. All permutations of these trial values were input to the Maeda model, giving a total of 5^7 = 78,125 vocal tract shapes. Of these, vocal tract shapes resulting in a valid set of formant frequencies (five formants greater than 200 Hz and less than 10 kHz) were retained and the rest were removed, leaving 23,009 valid shapes.

For each valid vocal tract shape, constriction points were identified in the tongue body and tongue tip, and constriction task parameters describing these points were calculated. Positions along the vocal tract were defined on a polar coordinate system as seen in Figure 1. The region between 40° and 70° from the horizontal axis was defined as the tongue tip. The position of narrowest constriction in this region was calculated as the tongue tip constriction location (TTCL), and the distance between the tongue and the palate at the TTCL was calculated as the tongue tip constriction degree (TTCD). Similarly, the region between 75° and 180° from the horizontal axis was defined as the tongue body, and the location and distance of the narrowest constriction in this region were calculated as the tongue body constriction location (TBCL) and tongue body constriction degree (TBCD).

Results: As seen in Figure 2, the formants produced by the set of valid vocal tract shapes cover the F1-F2 space in the range of 200 – 1000 Hz for F1 and 400 – 2500 Hz for F2. The distribution is approximately bell-shaped with highest density near the median values for both formants.

Although the F1-F2 space can be characterized by a two-dimensional bell-curve density, the corresponding set of vocal tract shapes have an irregular distribution across constriction task parameter values, as seen in Figure 3A. Tongue body constriction location exhibits a trimodal distribution, showing the constriction is restricted to locations around 95°, 120°, or 180° from the horizontal axis. For reference, the subsets of vocal tract shapes characterized by each of these constriction locations are plotted in Figure 3B. Tongue tip constriction location is less restricted, but it shows greatest density around 40-45° and 55-65° from the horizontal axis. There are also notable groups at the boundary of the defined tongue tip and tongue body ranges. These points indicate vocal tract shapes with no clear constriction in the given region, causing the boundary (anterior palatal region) to be the narrowest point. The degree of constriction for both components appears to be unrestricted, spanning the entire range from 0 to 65 mm, although the tongue body is more likely to have constriction degree less than 35 mm.

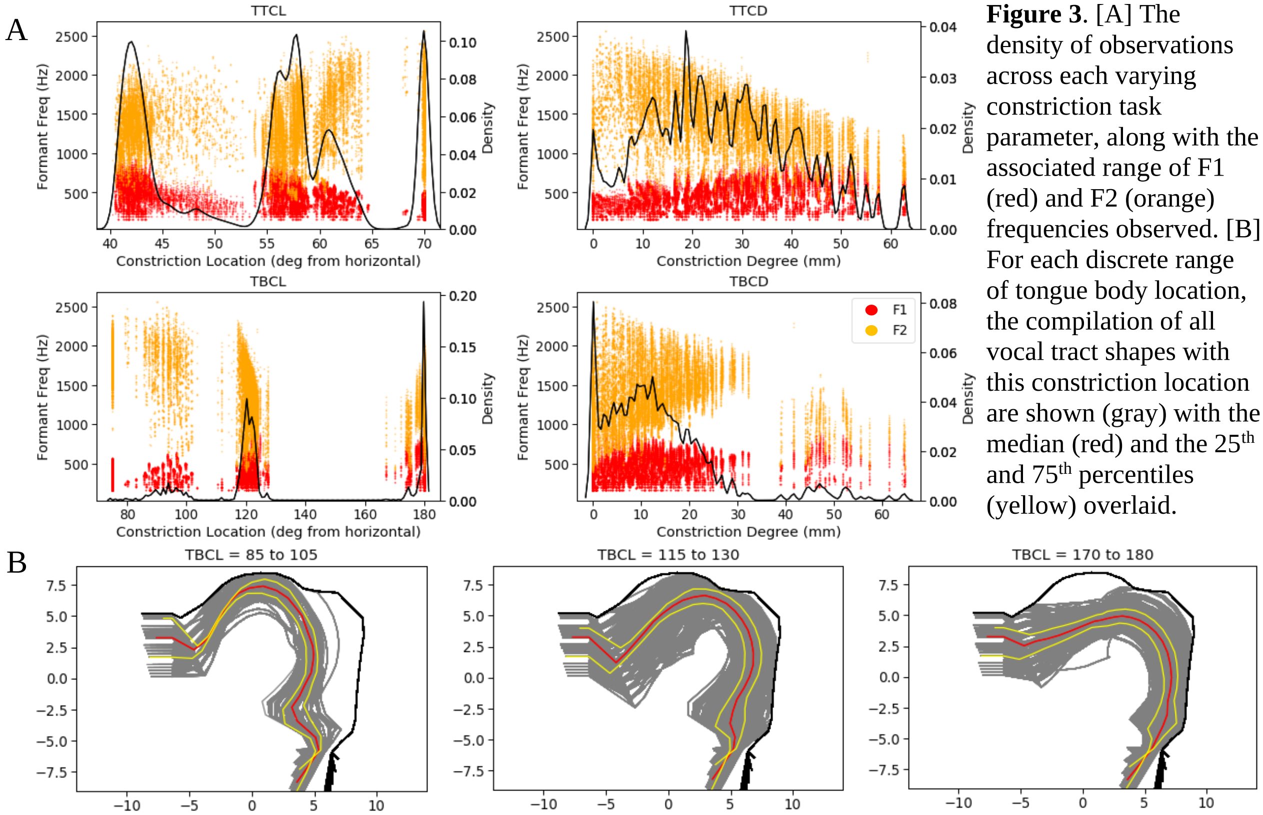

Figure 3. [A] The density of observations across each varying constriction task parameter, along with the associated range of F1 (red) and F2 (orange) frequencies observed. [B] For each discrete range of tongue body location, the compilation of all vocal tract shapes with this constriction location are shown (gray) with the median (red) and the 25th and 75th percentiles (yellow) overlaid.

Conclusion: The set of valid vocal tract shapes widely spanned a triangular vowel space defined by the first two formants. However, the tongue body constriction locations corresponding with these shapes were found only in discrete ranges of values. This implies that the large range of vocal tract shapes generated by varying the seven Maeda features can be reduced to three discrete constriction locations in the tongue body, showing this effect is not limited to linguistic vowel categories [3].

Citations: **[1]** S.Maeda. “Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model,” in *Speech production and speech modeling*, W. Hardcastle & A. Marchal, Eds. The Netherlands: Kluwer Academic Publishers, 1990, pp. 131-149. **[2]** S.Wood. “A radiographic analysis of constriction locations for vowels.” *J Phon.* vol. 7, no. 1, pp. 25-43, 1979. https://doi.org/10.1016/S0095-4470(19)31031-9. **[3]** L.-J.Boe. “The geometric vocal tract variables controlled for vowel production: proposals for constraining acoustic-to-articulatory inversion.” *J Phon.* vol. 20, no. 1, pp. 27-38, 1992. https://doi.org/10.1016/S0095-4470(19)30251-7. **[4]** J.Hillenbrand, et al. “Acoustic characteristics of American English vowels.” *J Acoust Soc Am.* vol. 97, no. 5, pp. 3099-3111, 1995. doi:10.1121/1.411872.